NASA/CP-1998-206853

p 10

Image Registration Workshop

NASA Goddard Space Flight Center
Greenbelt, Maryland, USA
November 20 - 21,1997
http:llcesdis.gsfc.nasa.gov/IRW

Proceedings

Edited by Dr. Jacqueline Le Moigne

Sponsored by: USRA/CESDIS, NASA Goddard Space Flight Center's Applied Information Sciences Branch and Earth & Space Data Computing Divison; and the WASH/NOVA Chapter of the IEEE Geoscience & Remote Sensing Society

Message from the General Chair

It is with great pleasure that I welcome you to the First NASA/Goddard Image Registration Workshop (IRW'97). Automatic image registration has often been considered as a preliminary step for higher-level processing, such as object recognition or data fusion. But with the unprecedented amounts of data which are being and will continue to be generated by newly developed sensors, the very topic of automatic image registration has become an important research topic. This Workshop is one of the first to concentrate on the issue of automatic image registration, and we were overwhelmed by the answer to our Call for Papers. We received submissions from all over the world, dealing with theoretical aspects of image registration as well as various applications such as satellite or medical imagery

This Workshop presents a collection of very high-quality work which has been grouped in four main areas:

- theoretical aspects of image registration
- applications to satellite imagery
- applications to medical imagery
- image registration for computer vision research.

We feel that the interaction of researchers coming from these four different viewpoints will be very fruitful and will lead to the development of improved image registration techniques. As a catalyst to this interaction, and under the sponsorship of Knowledge Transfer Technology (KTT), Global Science Technology (GST), and Hughes-STX (HSTX), a roundtable discussion has been organized during the Workshop.

IRW'97 would not have happened without the interest of the NASA/Goddard Space Flight Center's Applied Information Sciences Branch in automatic image registration or without their continuing support. We would also like to *thank* the following for their support: USRA/CESDIS, the NASA/Goddard Space Flight Center's Earth and Space Data Computing Division, and the Washington/Northern Virginia Chapter of the IEEE Geoscience and Remote Sensing Society

I am very grateful to the Technical Program Committee members for reviewing the papers and especially to Harold Stone (NEC) for his help in producing the proceedings. Most of all, I would like to acknowledge Georgia Flanagan (CESDIS) for her tremendous amount of work at all stages of the organization of the Workshop. I sincerely think that without her, this Workshop would still be a dream.

In conclusion, I would like to thank all of the authors and the participants of IRW'97 and I hope to see all of you next year at IRW98!

Jacqueline Le Moigne
USRA/CESDIS

IRW General Chair

Jacqueline Le Moigne, USWCESDIS
NASNGSFC - Code 930.5
Greenbelt, MD 20771
301-286-8723
301-286-1777 (fax)
lemoigne@cesdis.usra.edu

Technical Program Committee

Rama Chellappa, University of Maryland at College Park
Samir Chettri, GST and NASA Goddard Space Flight Center
Robert Cromp, NASA Goddard Space Flight Center
Tarek El-Ghazawi, George Washington University
Nazmi El-Saleous, University of Maryland at College Park
Emre Kaymaz, KTT
Bao-Ting Lerner, KTT
B. S. Manjunath, University of California, Santa Barbara
Manohar Mareboyana, Bowie State University
David Mount, University of Maryland at College Park
Nathan Netanyahu, University of Maryland at College Park
John Pierce, KTT
Srinivasan Raghavan, Hughes STX
Aya Soffer, University of Maryland at Baltimore County
Harold Stone, NEC Research Institute
James Tilton, NASA Goddard Space Flight Center
Eric Vermote, University of Maryland at College Park
Wei Xia-Serafino, Hughes STX

IRW Coordinator

Georgia Flanagan, USRA/CESDIS
NASNGSFC - Code 930.5
Greenbelt, MD 20771
301-286-2080
301-286-1777 (fax)
georgia@cesdis.usra.edu

TABLE OF CONTENTS

Session VI - General Methods with Application to Medical Imagery
Chairs: *Rama Chellappa and WeiXia*

Session VII - Theory and General Methods
Chairs: *B. S. Manjunath and John Pierce*

Session VIII - Computer Vision
Cham: *Srini Ragahavan and Emre Kaymaz*

POSTER SESSION

Session I

General Methods

Image Registration by nonlinear wavelet compression and singular value decomposition

Jorge E. Pinzón[1], John F Pierce[2],
Susan L. Ustin[3] and Claudia M Castañeda[3]
'Department of Mathematics
University of California at Davis
Davis, CA 95616
[2]KT-Tech, Inc.,
9801 Greenbelt Road, Suite 314,
Lanham MD 20706
[3]Department of Land, Air and Water Resources
University of California at Davis
Davis, CA 95616

Abstract

Current research in image registration focuses on automatic, fast, accurate, and reliable techniques whose application is feasible to the large amount of remote sensing data being generated by Earth and space missions. Image registration by point matching involves two main steps: the identification of significant features, commonly referred to as control points, and the matching of the features to identify the mapping function that brings one image into alignment with the master (reference) image. Automatically identifying control points is difficult, because (1)points significant in an image ("hot points") need not match with corresponding points in a reference image, and (2) even when matched pairs of points are obtained, there must be a significant number of them suitably distributed spatially to compute accurately a mapping function for the registration.
In this paper, we present a three step algorithm with feedback which automatically registers images differing by a translation, a rotation, and/or a homogeneous scaling (homotheity). First, we identify potential control points as hot points by highly compressing the reference and input images using non-linear wavelet techniques. Second, we use singular decomposition (SVD) methods to compute singular values and principal directions for these ensembles of hot points. Third, we use these data from each ensemble to determine a registering transformation Feedback between the first and second stage allow the SVD computations to optimize the probability that the ensembles of hot points manifest control points. Feedback among the three stages designed to optimize a correlation function refines the accuracy of the registering transformation.

1 Introduction

We present an algorithm for registering images by global, two-dimensional, rigid transformations (translation, rotation, uniform scaling). The algorithm registers images by extracting from each an ensemble of points, a significant number of which match under the transformation (control points) [1]. It then generates the registering transformation from these ensembles using tools from the theory of singular value decomposition (SVD) for matrices.

The algorithm consists of three stages with feedback. The first stage of the algorithm compresses the reference and input images, extracting hot pixels which manifest features significant to each image. The second stage carries out the SVD computations on each of the compressed images to extract singular values and principal directions for each ensemble. The third stage uses the centroid, singular values, and principal directions for each ensemble to compute the registering transformation.

A stability feedback loop between the second and fist stages allows the SVD computations to refine the compression and to maximize the probability that the ensemble of hot pixels in both images manifest viable control points, a significant number of which will match. A feedback loop among the third, second, and first stages refines the accuracy of the registration computed by the third stage.

We designed the algorithm to be computationally simple, fast accurate, reliable and automatic, to the extent that these requirements axe compatible. For example, given a choice between speed and accuracy, we designed the algorithm for speed; given a choice between speed and computational simplicity, we designed the algorithm for simplicity

In Section 2, we present the logical foundation for the algorithm We introduce the elements of the SVD and how they can be used with the ensembles of control points from the reference and input images to assess the registering transformation. We then indicate how the addition of random points which do not match to either ensemble affects the conclusions for the computed registration. It is this analysis which establishes the viability of using this technique computationally for registration purposes.

We then present the grounds by which we use nonlinear compression to extract "hot pixels" which have a high probability of manifesting features significant in each image, and thereby have a significant probability of representing control points whch match under the registration. It is this analysis which serves as the basis for the stability feedback loop, whereby the SVD computations optimize the compression.

In Section 3, we lay out the elements of the algorithm. In Section 4, we present the results of the validation and performance tests. In Section 5, we present conclusions about the shortcomings of the current version of the algorithm and set forth ways by which we could improve it in future versions.

2 Control Point Extraction and Registration Tools

In this section we present the basis for applying SVD tools for registration purposes. We present also the basis for using wavelet compression and the SVD computations to extract from each image a significant number of candidate control points, sufficiently distributed spatially, so as to produce an accurate registration

2.1 SVD and Registration

We take a ***global, rigid transformation*** acting on any subset in a fixed 2-D Euclidean space to be a transformation of the form

$$OQ = s\mathcal{R}_\theta OP + t, \tag{1}$$

where O is a fixed, but arbitrary point in the space, P is a point in the domain subset, Q is the image point of P under the transformation, OP and **OQ** are the position vectors from O to ***P*** and Q, respectively, s is a (non- negative) scalar representing a homogeneous scaling (homothetity), $\mathcal{R}_\theta$ denotes a rotation about the *z-axis* centered at O, counter-clockwise through the angle θ, and ***t*** denotes a two-dimensional vector characterizing a translation.

Take as given a reference and an input image, which we view as subsets of the Euclidean space. We say the images ***register*** under the transformation Equation (1), or the transformation ***regzsters*** the images if the transformation takes the reference image onto the input image.

Consider two ensembles of points P and $\mathcal{Q}$ in the reference and input images, respectively. We say the ensembles ***match*** under the transformation if points of the first are taken onto the second in a one-to-one manner by the transformation. When this occurs, we can show (1) that the centroids of the ensembles register under the transformation, (2) that we can isolate the rotation and scaling parameters for the transformation by examining the positions of points in each ensemble relative to their respective centroid, (3) that we can deduce the rotation and scaling parameters from the SVD computations for two matrices built from the components of the position vectors for the points in each ensemble relative to their respective centroids, and (4) once these parameters are computed, we can deduce the translation parameters from the position of the centroid in the input image and the rotated and scaled position of the centroid in the reference image.

Specifically, let P_0 and Q_0 denote the centroids of the respective ensembles. If the two ensembles match under the transformation, then Equation (1), implies that

$$OQ_0 = s\mathcal{R}_\theta OP_0 + t, \tag{2}$$

or that the centroids match under the transformation, 'Equations (1) and (2) imply that points ***P*** and Q in the reference, and input images register under transformation Equation (1), provided

$$Q_0Q = s\mathcal{R}_\theta P_0P \tag{3}$$

Equation (3) asserts that we may isolate the scaling and rotation parameters for the transformation by considering the position of points in each image, relative to the centroid of the respective ensembles.

We now demonstrate that if we arbitrarily order each ensemble, form (with respect to a **fixed** spatial basis) the $m \times 2$ matrix of coordinates for each point in each (ordered) ensemble, relative to its respective centroid, and compute the singular value decomposition of each matrix: (1) the ratio of the corresponding right singular values for the two matrices, ordered by magnitude, determines the scaling factor s, and (2) the difference in the directions of the corresponding right principal singular vectors determine the rotation angle θ for the transformation

Specifically, let P and $\mathcal{Q}$ be two ensembles of points in the reference and input images, respectively, which match under the transformation Equation (1) Order each ensemble arbitrarily. If m denotes the cardinality of each ensemble, Equation (3) implies, for $1 \leq i \leq m$,

$$Q_0Q_{\pi(i)} = s\mathcal{R}_\theta P_0P_i, \tag{4}$$

where π denotes the permutation on the index set $\{1, \ldots, m\}$ which specifies which points in the two ensembles match pairwise under the transformation and the two orderings.

We construct a matrix equation representing the system of Equations (4). Let **X** denote the $m \times 2$ matrix whose rows are the components (with respect to a fixed spatial basis) of the position vectors $X_i = P_0P_i$ to the ordered points in the ensemble in the reference image, relative to its centroid Let Y denote the $m \times 2$ matrix which is the counterpart built (with respect to the same fixed spatial basis) from the ordered ensemble of points in the input image, relative to its centroid. A straightforward computation shows that the matrix equation counterpart to Equation (4) is

$$\mathbf{Y} = s\mathbf{P}\mathbf{X}\mathbf{R}_\theta^T, \tag{5}$$

where

$$\mathbf{P}^T\mathbf{P} = \mathbf{I}. \tag{6}$$

Here, P is an $m \times m$ matrix which represents the action on the matrix **X** induced by the permutation π acting on the $\{X_i\}$ Equation (6) asserts that it is an orthogonal matrix. R_θ is the (orthogonal) 2×2 matrix representing the rotation $\mathcal{R}_\theta$ with respect to the fixed spatial basis.

We now demonstrate we can assess, in principle, the rotation and ***scaling*** parameters for the transformation from the singular values and principal directions for the matrices associated with the ensembles of points in the reference and input images. For any $m \times n$ real matrix X, a ***right singular value*** for X is a real, positive number σ satisfying

$$X^TXu = \sigma^2u, \tag{7}$$

for some $u \in R^n$, $\| u \| = 1$ Term u satisfying (7) a ***right principal singular direction*** associated with σ

A computation similar to that leading to Equation (6) shows that the singular values and principal singular directions for a matrix are independent of the ordering of the rows of the matrix. Indeed, for the case where $\mathbf{X}$ is the $m \times 2$ matrix associated with the positions of the ensemble of points in the reference image, relative to the centroid, the principal singular directions specify the lines of the semi–major and semi–minor axes of the smallest ellipse centered at the centroid enveloping the ensemble, and the ratio of the singular values specify the eccentricity of the ellipse [3, p. 26]. In other words, the singular values and principal directions characterize ***geometric*** qualities of the ensemble which are independent of the ordering of the points.

We can now see how the singular values and principal directions for the matrices associated with the ensembles of points in the reference and input images characterize the rotation and scaling parameters for the transformation.

Theorem 2.1. Let P and $\mathcal{Q}$ denote ensembles of points in a reference and input image, respectively, matched under a transformation of the form Equation (1) which registers the images. Order the two ensembles arbitrarily. Let $\mathbf{X}$ and $\mathbf{Y}$ denote the $m \times 2$ matrices whose rows are the components (with respect to a fixed spatial basis) of the position vectors to the ordered points in the reference and input ensembles, relative to their respective centroids.

If

$$\mathbf{X}^T\mathbf{X}\mathbf{u} = \sigma^2\mathbf{u}, \tag{8}$$

then

$$\mathbf{Y}^T\mathbf{Y}\mathbf{v} = \tau^2\mathbf{v}, \tag{9}$$

for

$$\mathbf{v} = \mathbf{R}_\theta\mathbf{u}, \quad \text{and} \quad \tau = s\sigma, \tag{10}$$

and conversely.

Proof. X and $\mathbf{Y}$ satisfy Equation(5), for some orthogonal matrix $\mathbf{P}$ representing the action of a suitably chosen permutation on m elements. Let $\sim$, $\mathbf{u}$ satisfy Equation (8). Then

$$\mathbf{Y}^T\mathbf{Y}(\mathbf{R}\mathbf{u}) = s^2\mathbf{R}\mathbf{X}^T\mathbf{P}^T\mathbf{P}\mathbf{X}\mathbf{R}_\theta^T(\mathbf{R}_\theta\mathbf{u}). \tag{11}$$

Equations (9) and (10) follow from Equation (11), and the orthogonal nature of $\mathbf{P}$ and $\mathbf{R}_\theta$

□

Equations (9) and (10) assert that the difference in the orientations of the principal directions for the ellipses enveloping the ensembles of points in the reference and input images determines the angle parameter for the registering transformation, and the ratio of the corresponding singular values, ordered by magnitude determines the scaling parameter

When the ensembles contain points whch do not match under registration, there is impact upon the conclusions of Theorem 2.1 However, if the points in the ensembles which do not match under registration are randomly distributed, the SVD computations of the orientation of the principal singular directions and of the singular values are, in the mean, unaffected Consequently, the impact upon the computations of the rotation and scaling parameters will be, in the mean, minor. Figure 1 demonstrates this assertion empirically. In this figure, we show average variation of the principal direction after 100 computations of ensembles augmented by 20% of randomly generated set of points. In all cases, the variation is less than 1 degree.

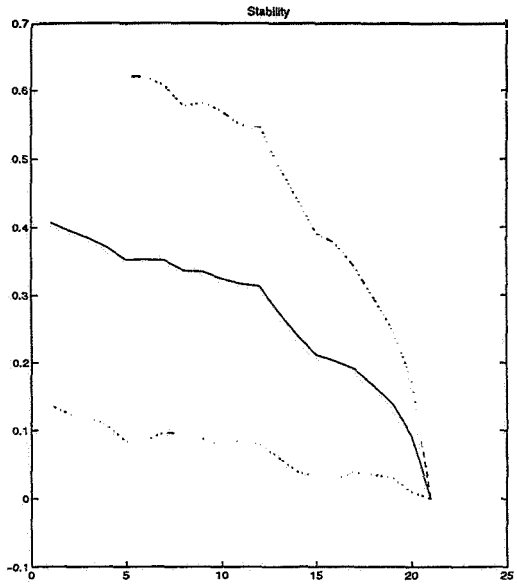


Figure 1 Stability of principal directions in an randomly augmented ensemble.

Table 1 summarizes one of many computations of scale factor, rotation, and translation parameters for an ensemble of points transformed by the the values in parenthesis then augmented 20% by a randomly generated set of points. For the rotation and scale parameters, the results concur with the assertion in Figure 1

However, the computation of the translation parameters, which is based upon centroid computations is more sensitive to the presence of points in one ensemble not matching under registration to points in the other, as Table 1 shows. The problem becomes acute when the ensembles significantly differ, because of cropping of the input image, for example. We address this concern in the conclusion, Section 5.

Stats	scale (2)	$\alpha(60°)$	$T_x(40)$	$T_y(-35)$
mean	1.82	59.92	37 24	-45.31
std	0.02	0.49	2.79	1.87
mse	0.03	0.24	15.32	109.77

2.2 Compression and Control Points

We use a nonlinear wavelet compression strategy to extract from the reference and input images "hot pixels" We assume that among them are points which manifest features inherent to each image, and thereby are candidates for control points which would match under the registering transformation We transform the images using a Fast Wavelet Transform, relative to an orthogonal wavelet basis, in this case, Coiflet. From a histogram of the magnitude of the

wavelet coefficients, we establish a filter which maintains a significant interval of those nonzero coefficients which exceed an adaptive thresholding percentage of the maximum magnitude of the coefficients, and nullifies all other coefficients. Reconstructing spatial images from the "softly" thresholded coefficients produces compressed, but "cloudy" versions of the reference and input images. Incrementing the filter interval can reconfigure the "clouds" in these compressed images.

A hard thresholding of these spatial images, using a filter built from the histogram of the pixel gray level values of these images, allows us to "removepenumbra", and produce compressed, hot pixel versions of the reference and input images. Incrementing this filter allows us to add or remove hot pixels from the compressed images.

By itself, this process will not automatically extract ensembles of points which manifest features inherent to the images, and thereby present themselves as candidate control points. However, results of the SVD computations can control the incrementation of the filters in the compression and spatial thresholding steps, and thereby optimize the probability that a significant number of hot pixels in each image do in fact manifest inherent features of the image. This control is the foundation for a stability feedback loop.

We base the assertion on the following observation if you add to an ensemble of points a collection of randomly distributed points, the orientation of principal singular direction associated with the original ensemble will be, in the mean, unaffected. This observation is the basis of how we designed the program to automatically control the compression and thresholding filters.

We assume that hot pixels which manifest features inherent in an image are not randomly distributed, and that hot pixels which manifest artifacts in an image, such as incidental lighting, are. For a given compression, we design the program to increment the thresholding filter, to produce or remove hot pixels from each compressed image. The program computes the rate of change in the orientation of the principal singular direction, with respect to the addition of hot pixels. If the rate of change is sufficiently nonzero, points manifesting significant features are more likely being added, and the incrementing process should continue. If the rate of change approaches zero, points being added are more likely random, and the compression/thresholding process should terminate.

3 The Algorithm

The flowchart for the wavelet and SVD registration algorithm is displayed in Figure 2.

As the flowchart shows, the reference and input images are read in, and decomposed by Fast Wavelet Transform [4]. These decomposition then enter the hot pixel extraction stage, which is controlled by the stability feedback loop. For each image compressed, hot pixel renderings are extracted, in the manner described in Section 2.2. We then compute the singular values and principal directions for each compressed image, in the manner described in Section 2.1 If the ratio of the singular values in an image is not sufficiently bounded or does not deviate sufficiently from unity, the stability loop increments the filter governing the soft thresholding parameter for the wavelet coefficients in the compression stage, producing a new configuration of hot pixels. If this ratio is sufficiently bounded and bounded away from unity, the stability loop then increments the ater governing the hard thresholding parameter for the pixel values in the hot point extraction stage, thereby varying the number of hot points in the compressed image. The loop then computes the rate of change of the orientation of the principal singular vector for the image with respect to the number of hot points, in the manner described in Section 2.2. The loop continues to add or remove hot points in this manner, until this rate of change becomes sufficient small. When this occurs, the stability loop outputs the final selection of compressed, hot pixel renderings of the reference and input images to the registration module.

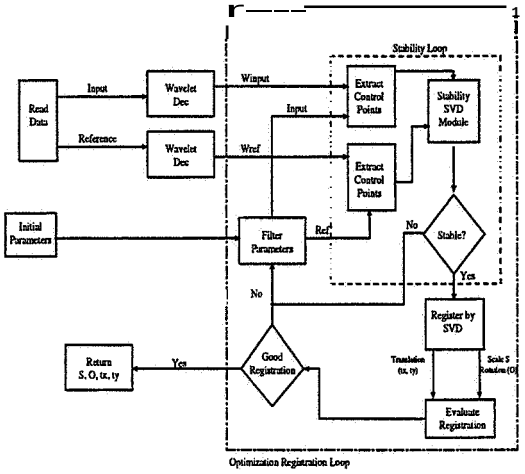


Figure 2. Registration System Diagram

In the manner described in Section 2.1, the registration module computes the singular value decompositions of the $m \times 2$ matrices derived from the compressed, hot pixel renderings of the reference and input images, and from the decompositions computes a first approximation to the scaling, rotational, and translational parameters for the registering transformation. The module then transforms the compressed reference image in the manner dictated by the parameters, and compares the resulting image to the compressed input image, using a distance correlation measure. If the measure is not sufficiently small, the registration module initiates a loop to refine the registration computations. It increments the hard thresholding parameter to refine the compressed, hot pixel reference images, and iterates the registration step to compute a refinement for the translation parameters. When the distance correlation measure is rendered sufficiently small by the iteration, the module outputs the refined values for the scaling, rotational, and translational parameters.

4 Validation and Performance Results

In Table 2 we summarize some of the results of validation and performance tests of the algorithm on three types of images: a portrait of a girl, a Landsat TM image, and and several AVHRR images. The images are available at the ftp

site ftp://cesdis.gsfc.nasa.gov/pub/people/lemoigne/at outgoing/TOOLBOX_TESTING.

In the case of the girl and the TM image, the input images were prepared as specific translations and and rotations of the reference images. We present these values in the fist column of Table 2 as a *six* digit number, *two* digits each for the angle, horizontal and vertical offsets.

The reference image for the AVHRR scenes is the file avhrr_map, available at the ftp site. The scenes listed in the table are images taken by the satellite at different times of the same day, as indicated in the file name. In each case, the input image differs from the reference image by a pure translation

For each input image, we present the angle, horizontal and vertical offsets from the reference image in columns two through four In each case, the scale parameter automatically rounded up to one, as expected.

In columns five through seven we present our measurement of the performance of the algorithm as times required to input/output images to and from the algorithm, to perform the wavelet compression, and to perform the registration. This computations were executed in a workstation alpha-dec-osf3.2.

	Input		Results			ime (secs)	
	$\theta_0 T_x T_y$	θ	T_x	T_y	Input	Comp.	Reg.
G	550000	49.96	-4.83	112	0.77	32.10	52.02
I	050000	5.25	2.80	-4.75	0.80	26.30	82.93
R	050604	8.98	1.61	10.57	0.85	29.75	88.38
L	052060	17.84	24.49	21.47	0.89	29.87	55.55
	002060	0.93	3.65	39.71	0.87	26.58	55.60
	180000	26.96	9.16	15.28	1.01	27.30	84.66
T	180050	-8.33	-8.19	24.10	1.26	28.32	52.85
M	045000	113	41.51	-8.49	0.95	29.35	66.61
	040502	5.05	9.44	0.04	0.96	27.13	54.41
	005000	-5.35	40.91	-16.54	1.06	29.61	33.37
A	sa1244	-4.89	-28.08	-6.67	2.00	32.49	59.47
V	sa1250	-3.98	5.65	14.69	3.17	32.55	66.22
H	sa1300	-3.34	-0.98	-2.43	1.86	25.38	51.91
R	sa1408	-4.64	21.59	18.27	1.83	31.42	33.53
R	sa1411	-35.35	8.17	-11.41	173	25.90	53.71

Although the registration of the GIRL and TM images is good, significant deviations were noticed in the AVHRR images, particularly for sa1411 We attribute this deviation to extensive cloud coverage in the image. We discuss this situation in the conclusions, Section 5

5 Conclusions and Generalizations

In summary we present a three stage algorithm with feedback to automatically register images differing by a global rigid transformation The significant features of the algorithm are the use of nonlinear wavelet compression to extract control points in each image and the use of singular value decomposition tools to register these points. In contrast to linear wavelet compression, the use of nonlinear wavelet compression greatly filters the number of extraneous or redundant control points. The SVD tools are significant, because they allow us to register ensembles of control points without having to know specifically which points match up pairwise.

The algorithm has two major shortcomings. First, because wavelet decompositions are neither translation nor rotation invariant, selecting the compression and thresholding filters is intricate. Small variations in the registration parameters do not necessarily result in small changes to the compression and thresholding filters. The stability loop based upon the SVD computations allow us to minimize the production of artifacts at the compression stage and to lessen the impact on the computation of the rotation parameter, however, the algorithm remains vulnerable in the computation of the translation parameter We are examining how preprocessing the images with a Hough transform or edge detector might mitigate these conditions for this version of the algorithm

Second, the images which the algorithm ingests must inherently have within them a sufficient number of control points, sufficiently distributed spatially, in order to function successfully In particular, the algorithm is noticeably unsuccessful on images present scenes obscured significantly by clouds. This was seen in the validation tests of Section 4, where the algorithm gives reasonable results for AVHRR images depicting basically cloudless scenes, but unreasonable results for the AVHRR image depicting a cloud filled scene.

In the next version of the algorithm we anticipate correcting this second shortcoming by preprocessing images with cloud filtering algorithms, as they become available. In particular, such algorithms based upon fractal cloud models and multi–spectral feature extraction methods appear particularly promising.

We anticipate correcting the first shortcoming in the next version of the algorithm by refining the compression stage through the introduction of a steerable pyramid [2] A steerable pyramid is an overcomplete wavelet transform which produces a representation that is almost translation invariant (subbands are aliasing–free) and rotation invariant (subbands are steerable). We anticipate this tool will stabilize the filter selection processes and result in a more accurate computation of the translation parameters.

References

[1] DeVore, R. A., et. al. *Using nonlinear wavelet compression to enhance image regzstration.* Wavelet Applications IV, H Szu, Chair/Editor, AeroSense97, Orlando FA, April, 1997 SPIE Proceedings, V 3078, 1997, pp. 539–551

[2] Simoncelli, E. P and Freeman, W T *The steerable pyramad. a flexible architecture for multiscale derivative computatzon.* IEEE Second Int'l Conf. on Image Processing, Washington D.C., 1995.

[3] Trefethen, L. N and Bau, D. *Numerical Linear Algebra.* SIAM, Philadelphia, PA, 1997

[4] Wickerhauser, M. V *Adapted Wavelet Analysis* from *Theoy to Software.* A. K Peters, Ltd., Wellesley, MA, 1995.

An Eigenspace Approach to Multiple Image Registration*

Haim Schweitzer (haim@utdallas.edu)
The University of Texas at Dallas
P O Box 830688
Richardson, Texas 75083

Abstract

Multiple images can be indexed by their projections on a few "eigenfeatures" These eigenfeatures are the eigenvectors of a large covariance matrix, constructed from the images. It is known that registration is essential for computing "useful" eigenfeatures, and a preliminary step of putting the images in register is common practice We propose to evaluate multiple image registration by the quality of their eigenfeatures. An algorithm that simultaneously registers the images and computes the eigenfeatures is proposed. The key idea is to iterate the following two steps: 1 Eigenfeatures are computed from the images. 2. New images are computed by registering the images on the eigenfeatures subspace. In the next iteration Step 1 is applied to the set of images that were most recently computed in Step 2 The implementation of this simple procedure requires a novel definition of registration on an image subspace. It is demonstrated that the algorithm registers multiple images, and produces improved eigenfeatures.

Keywords: Image-registration, Image-indexing, Eigenfeatures, Motion-compensation

1 Introduction

Let $x_1, \ldots, x_m$ be m images, viewed as n dimensional vectors of pixel values. In principal component analysis one determines a set of k "eigenfeatures" (also called primary images) $v_1, \ldots, v_k$ so that each one of the original m images can be approximated as:

$$x_i \approx \sum_{j=1}^{k} f_{ij} v_j \quad \text{for } i = 1, \ldots, m \tag{1}$$

The scalars f_{ij} are typically used as content based indexes to the image set It is known [Fuk90] that the optimal choice of eigenfeatures (in the sense of minimizing the mean squared error) is the k eigenvectors associated with the k largest eigenvalues of the images matrix of second moments. A related approximation allows for a constant term in addition to the linear terms in (1):

$$x_i \approx v_0 + \sum_{j=1}^{k} f_{ij} v_j \quad \text{for } i = 1, \ldots, m \tag{2}$$

In this case it is known [Fuk90] that the optimal choice for v_0 is the images mean, and $v_1, \ldots, v_k$ are the k eigenvectors associated with the k largest eigenvalues of the images covariance matrix (The approximation in (2) is reduced to the approximation in (1) if the mean is subtracted from each image before the matrix of second moments is computed.)

Eigenfeature analysis of images has received a lot of attention recently For applications to object recognition see [TP91, MN95]. Applications to image indexing and digital libraries are discussed in [SW96, PPS94, Sch98b, Sch98a, Schss] Applications to learning and scene analysis can be found in [MN95, MN94].

The approximations (1), (2) are linear in pixel values but keep pixel locations fixed Therefore, they may not give good estimates in situations where images are related by a simple coordinate transformation such as translation This limitation of eigenspace techniques is usually handled by a preliminary registration step, where translation and possibly other transformations are applied to put the images in register The registration is sometimes carried out by hand (e.g., [RB95, SW96]), and in other cases it involves automated preprocessing that may require segmentation (e.g., [MN95, CTCG95]) In some recent work [BP96, OVVSss, VP96] registration is computed using optical flow

The preliminary step of registering multiple images appears more challenging than the registration of one image on another, since it is not always clear on what to register the images. In previous work the most common choice was to attempt registration on one of the images (e.g., [Sch95]), or, in the case of face recognition, on a generic face (e.g., [RB95]). An interesting procedure for computing an "average" face as the target for registration is described in [OWSss].

It is clear that in many situations a single template may not be sufficient, and registration should be attempted on one of several different templates. For example, in work on face recognition one may choose generic frontal and side views as the templates. Another situation where not all images can be registered on a single template is the analysis of a complex image motion sequence. A common approach is to attempt extracting "motion layers" (e.g., [SA96]). This involves registering areas in an image that appear to be mov-

*Research supported in part by NSF grant IRI-9309135

ing uniformly, segmenting them out, and repeating the process to compute additional components.

1.1 The main contribution

We use the observation that the registration of multiple images enables the computation of useful eigenfeatures to measure the quality of multiple image registration. Thus, for a given set of images we attempt global registration that would minimize the well defined error of approximating the registered images in terms of a few eigenvectors. An immediate observation is that in order to minimize this error the images should be registered on a linear combination (a subspace) of the the eigenvectors. An iterative algorithm that alternates between eigenvector computation and image registration is proposed Its output is the registered images and their eigenfeatures

The results are related to (and improve upon) previous work in two areas. The first is content based image indexing in terms of eigenfeatures (e.g., [TP91, Sch98b, Sch98a, SW96, PPS94]) Our contribution is the computation of improved eigenfeatures. The second area is image registration [Bro92]. Our contribution is an algorithm for multiple image registration. In particular, unlike previous work where images were registered on individual frames we propose registering multiple images on a subspace.

The proposed algorithm is also related to work on describing image sequences in terms of dominant motions, or motion layer (e.g., [SA96]) Our algorithm segments the scene into several components (the eigenfeatures), but it associates only one transformation (motion) with each frame

1.2 Paper organization

The paper is organized as follows. An error expression that measures the quality of image registration and eigenfeature computation is derived in Section 2. Algorithms for minimizing this error are discussed m Section 3 Experimental results are described in Section 4.

2 The error expression

A transformation (registration) of an image x that produces the registered image y is written as $y = \phi(x)$ We propose to measure the quality of multiple image registration by how well the registered images can be approximated in terms of a few eigenfeatures Specifically, for each image x_i from the given set of m images $x_1, \ldots, x_m$ we search for a transformation ϕ_i so that the set of m registered images $y_i = \phi_i(x_i)$, $i = 1, \ldots, m$, can be accurately approximated in terms of k (a small number of) eigenfeatures. Using the approximation in (1) and treating images as vectors we get the following expression for the global error.

$$E = \sum_{i=1}^{m} |\phi_i(x_i) - \sum_{j=1}^{k} f_{ij} v_j|^2 \qquad (3)$$

Thus, when given the images $x_1, \ldots, x_m$ as input our goal is to compute the transformations ϕ_i, the eigenfeatures v_j, and the scalar constants f_{ij} that produce small error in (3).

Observe that if all the transformations ϕ_i are chosen as the identity (unregistered images), the choice of vectors v_j and scalars f_{ij} that minimize the error (3) is reduced to optimizing the approximation in (1) Therefore, the optimal solution is for $v_1, \ldots, v_k$ to be the dominant eigenvectors of the images matrix of second moments. Another special case is with $k = 1$ Here the error in (3) is a measure to the quality of registering all the images on a single vector (template image) v_1 This shows that the minimization of the error (3) requires a generalization of standard image registration and eigenvector techniques.

With no restrictions on the allowable set of registration transformations there are trivial meaningless solutions. (For example, $\phi_i(x)$ = constant image independent of x.) To eiiminate these cases we will assume that there is a predefined fixed family of registration transformations. Examples are translations or planar affine transformations.

3 Error minimization

In this section we describe computational methods to minimize the error (3). To simplify notation we write $V = (v_1, \ldots, v_k)$ as an $n \times k$ matrix, and $f_i = (f_{i1}, \ldots, f_{ik})'$ as a k vector (We use " ' " for matrix transpose.) With this notation the global error to be minimized can be written as:

$$E = \sum_i |\phi_i(x_i) - V f_i|^2$$

This global error can be written as a sum of "local" errors. Let $y_i = \phi_i(x_i)$ be the ith registered image, and set $e_i = y_i - V f_i$ This gives: $E = \sum_i |e_i|^2$ We propose to minimize the global error E by iterating the following two steps:

Step A: Keep V fixed. For each i determine ϕ_i that minimizes the local error $|e_i|^2 = |\phi_i(x_i) - V f_i|^2$

Step B: Keep the ϕ_i fixed Determine V that minimizes the global error E

3.1 Algorithms for Step A

This step involves the registration of an image x on the subspace V The solution is the registration transformation ϕ_i that takes x as near as possible to the subspace spanned by the columns of V In our case V is orthonormal (since its columns are eigenvectors), and the local error can be expressed as:

$$\epsilon_i = |e_i|^2 = |\phi_i(x_i)|^2 + |f_i|^2 - 2\phi_i'(x_i) V f_i \qquad (4)$$

We describe two algorithms for computing image registration on a subspace. The first uses a "black-box" standard image registration routine, and the second uses a "black-box" nonlinear minimization routine.

3.1.1 First algorithm

This algorithm uses a "black-box" image registration routine to compute subspace registration. The idea is to iterate the following two steps:

(a) Keep ϕ_i fixed and determine f_i that minimizes ϵ_i in (4).

(b) Keep f_i fixed and determine ϕ_i that minimizes ϵ_i in (4)

Since ϵ_i is quadratic in f_i, the solution to Step (a) can be written in a closed form:

$$y_i = \phi_i(x_i), \quad f_i = V' y_i$$

Step (b) can be computed by:

$$u = V f_i, \quad \text{compute } \phi_i \text{ that minimizes } |\phi_i(x_i) - u|^2$$

Since u is a vector (a single image), the last part can be implemented with any registration routine that would register x_i on u. To summarize, a single local iteration cycle gets as input an estimate to the registered image $y_i = \phi(x_i)$, and proceeds to improve this estimate as follows:

1. $f_i = V' y_i$ (equivalently, $f_{ij} = v_j' y_i$ for $j = 1, \ldots, k$).
2. $u = V f_i$ (equivalently, $u = \sum_{j=1}^{k} f_{ij} v_j$).
3. A new registered image y_i is computed by registering x_i on u.

3.1.2 Second algorithm

This algorithm uses a "black-box" nonlinear minimization routine to compute subspace registration As mentioned above, the value $f_i = V' y_i$ minimizes (4). Substituting and simplifying using the orthonormality of V gives:

$$\epsilon_i = |\phi_i(x_i)|^2 - |V' \phi_i(x_i)|^2$$

Therefore, ϵ_i is a (nonlinear) function of the transformation parameters. The number of parameters is 2 for translations, and 6 for the general affine coordinate transformation. The minimizing parameters can be computed by applying standard minimization routines to ϵ_i

3.1.3 Comparing the two algorithms for Step A

In our experiments the two algorithms produced identical registration The first algorithm takes 4-5 iterations to converge. In our particular implementation the registration code uses function minimization, as described in [Sch95]. The same code was used for the minimization in the second algorithm that does not require iterations, so that the second algorithm was significantly faster The first algorithm should be the preferred choice when an efficient registration routine is available.

3.2 An algorithm for Step B

Here the transformations ϕ_i are known, so that the registered images $y_i = \phi_i(x_i)$ can be computed. Our goal is to compute V that minimizes the global error.

$$E = \sum_i |y_i - V f_i|^2$$

This is the error of approximating the images y_i in terms of k eigenvectors. Therefore, the optimal solution $V = (v_1, \ldots, v_k)$ is obtained by applying standard eigenvector decomposition to the matrix of second moments of the (registered) images y_i Specifically,

1. Compute $y_i = \phi_i(x_i)$ for $i = 1, \ldots, m$.
2. Compute $Y = \sum_{i=1}^{m} y_i y_i'$
3. The vectors $v_1, \ldots, v_k$ are the k eigenvectors of Y associated with its k largest eigenvalues.

Since Y is $n \times n$, where n is the number of pixels in each image, a direct computation of Y is impractical. Efficient algorithms that perform the eigenvector decomposition without explicitly computing Y are known See, for example, [ML95, Schss, Sch98a].

4 Experimental results

This section describes the results of experiments with the proposed algorithm In all cases, the registration transformations ϕ_i were computed as affine coordinate transformations. (A single affine transformation, specified in terms of 6 parameters, was computed for each image.) The minimization code was the implementation described in [Sch95]. The eigenvector routine was the one described in [Schss, Sch98a].

The algorithm of Section 3 that alternates between Step A and Step B converged in 4 6 iterations. The experiments show significant reduction in the error, as defined by Equation 3, for small values of k, typically in the range of 1-4 For larger values of k there is very little change in the eigenvectors. The reason appears to be that for large k values the images are so close to the eigenvector subspace that registration can yield very little improvement.

For $k = 2$ we observed results that appear as image registration on two distinct templates. This is an experimental observation, since "in theory" one may expect registration on a continuous range of linear combinations of the two eigenvectors. We describe in details the results of two experiments.

4.1 First experiment

The algorithm was applied to the sequence of 72 views of the first object from the Coil-20 image set [NNM96]. These are consecutive views of a wooden duck with varying pose of 5 degrees. The first row of Figure 1 shows 9 images from the sequence (with varying pose of 40 degrees), and the two principal eigenfeatures (computed from the entire sequence) The registration and the two principal eigenfeatures computed by applying the algorithm with $k = 1$ (registration on one eigenvector) are shown in the second row The algorithm appears to register the frames on the duck's head and body, and they are clearly visible in the first eigenvector The registration and the two principal eigenfeatures computed by applying the algorithm with k = 2 (registration on a subspace of two eigenvectors) are shown in the third row Here the results show registration either on a "left view", or on a "right view" of the duck All images of frontal, or back view, are tilted and shifted towards one of these views.

The error per pixel in the approximation (1) is summarized in the following table:

	before registration	after registration
k = 1	52.3	**27.8**
k = 2	38.9	24.5

4.2 Second experiment

The algorithm was applied to the first 29 frames of the *flower garden* sequence. (It is a standard sequence in video compression experiments.) The scene change involves background motion due to parallax of houses and garden, and a much larger motion (due again to parallax) of a tree in the foreground. Figure 2 shows several frames from the sequence, and the two principal eigenfeatures (computed from all 29 frames) Observe that the first eigenvector is a blurry image of the background. The registration and the two principal eigenfeatures computed by applying the algorithm with $k = 1$ (registration on one eigenvector) are shown in the second row The algorithm registers the frames on the background, roughly at the same position as the background is in the 15th frame. The second eigenvector contains mostly information about the tree. The registration and the two

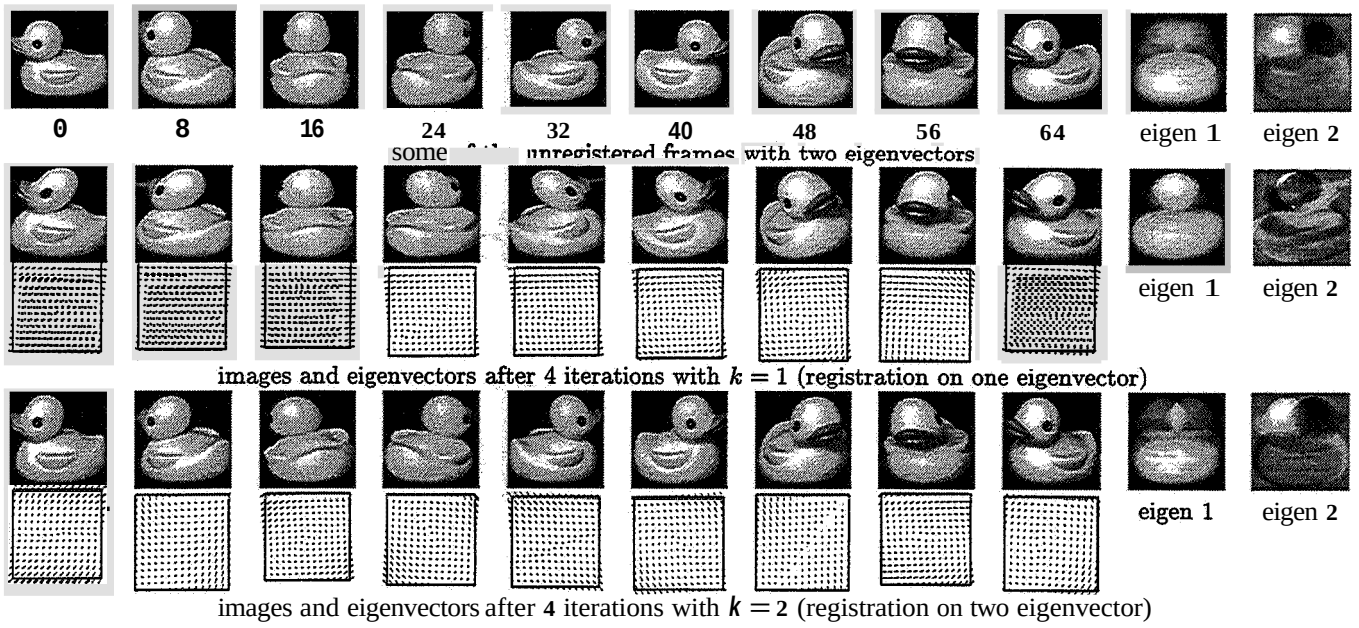


Figure 1: results from the first experiment.

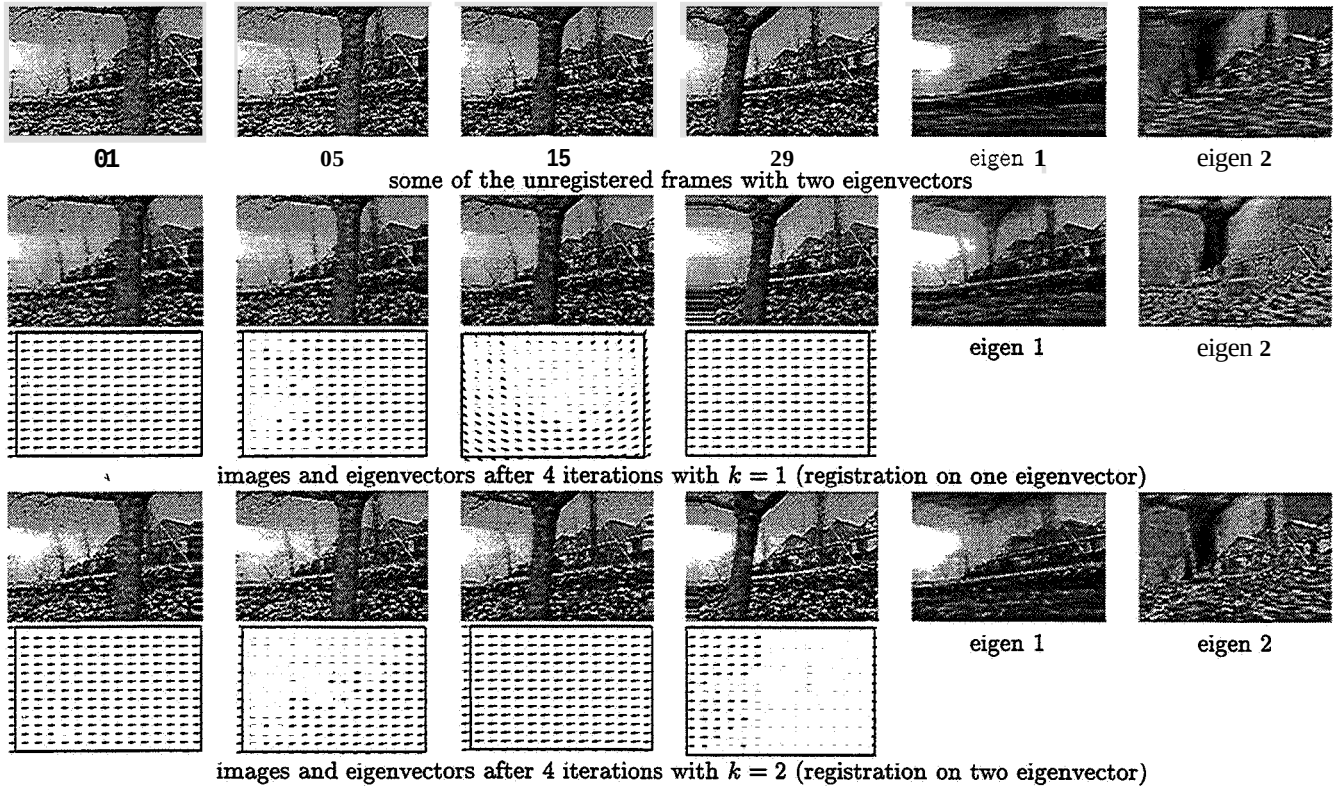


Figure 2: results from the second experiment.

principal eigenfeatures computed by applying the algorithm with k = 2 (registration on a subspace of two eigenvectors) are shown in the third row Here the results show registration on two background positions. The first eigenvector shows a "double image", with both of these locations.

The error per pixel in the approximation (1) is summarized in the following table:

	before registration	after registration
k = 1	77.4	59 7
k = 2	68.7	49.0

5 Concluding remarks

This paper describes an algorithm that registers multiple images and computes their eigenfeatures. The algorithm was tested on real images and appears to perform well, but it is clear that the iterative procedure may get "stuck" in a local minimum of the error expression (3) Better results may be obtained with improved minimization.

In our implementation the computation of image registration. was the most time consuming. The eigenvector decomposition was implemented with the algorithm described in [Sch98a, Schss] That particular algorithm was designed for computing eigenfeatures in a distributed environment such as the Internet We observe that the expensive registration step (Step A in Section 3) is "local", and can be computed independently at each image location. When combined with the distributed algorithm of [Sch98a, Schss] the entire computation can be implemented very efficiently on a network of loosely connected computers, or over the Internet.

The idea of registering images on a subspace appears to be novel, and may be useful even if one is not interested in computing eigenfeatures. Specifically, when given m images $x_1, \ldots, x_m$ and k template images $v_1, \ldots, v_k$, the following procedure registers the images on the template subspace (a linear combmation of the templates)

1. Orthonormalize the matrix $V = (v_1, \ldots, v_k)$. This can be done, for example, by using the modified Gram-Schmidt [GVL96].

2. Apply one of the algorithms of Section 3 1

Acknowledgment

This work was supported in part by the National Science Foundation under Grant IRI-9309135.

References

[BP96] D Beymer and T Poggio. Image representations for visual learning *Science*, 272:1905–1909, June 1996

[Bro92] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, December 1992

[CTCG95] T F Cootes, C. J Taylor, D H. Cooper, and J Graham. Active shape models — their training and application. *Computer Viszon and Image Understanding: CVIU*, 61(1):38–59, January 1995.

[Fuk90] K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, New York, second edition, 1990.

[GVL96] G H Golub and C. F Van-Loan. *Matrix computatzons.* The Johns Hopkins University Press, third edition, 1996.

[ML95] H Murase and M. Lindenbaum. Partial eigenvalue decomposition of large images using spatial-temporal adaptive method. *IEEE Transactions on Image Processing*, 4(5):620–629, May 1995.

[MN94] H. Murase and S. Nayar Illumination planning for object recognition using parametric eigenspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(12):1219 - 1227, December 1994.

[MN95] H Murase and S Nayar Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision*, 14:5 - 24, 1995

[NNM96] S. Nene, S Nayar, and H. Murase. Columbia object image library COIL-20. Technical Report CUCS-006-96, Columbia University, 1996.

[OVVSss] A O'Toole, T Vetter, H. Volz, and E. Salter Three-dimensional caricatures of human heads: Distinctiveness and the perception of facial age. *Perception*, in press

[PPS94] A Pentland, R. W Picard, and S. Sclaroff Photobook Tools for content-base manipulation of image databases In *Proc. SPIE Conf on Storage and Retrieval of Image and Video Databases II*, pages 34–47, San Jose, CA, February 1994.

[RB95] R Rao and D Ballard. Natural basis functions and topographic memory for face recognition. In *Proceedings of the 14th International Joznt Conference on Artificaal Intelligence(IJCAI'95)*, pages 10–17, Montreal, August 1995 Morgan Kaufman

[SA96] H Sawhney and S. Ayer Compact representation of videos through dominant and multiple motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):814–829, August 1996

[Sch95] H. (Shvaytser) Schweitzer *Occam* algorithms for computing visual motion. *IEEE Transactions on Pattern Analysis and Machine Intellzgence*, 17(11):1033–1042, November 1995.

[Sch98a] H. Schweitzer Computing Ritz approximations to primary images. In *Proceedings of the Sixth International Conference on Computer Vision (ICCV'98)*, page (in press), January 1998.

[Sch98b] H Schweitzer Indexing images by trees of visual content. In *Proceedings of the Sixth Internatzonal Conference on Computer Vision (ICCV'98)*, page (in press), January 1998.

[Schss] H Schweitzer **A** distributed algorithm for content based indexing of images by projections on Ritz primary images. *Data Mining and Knowledge Discovery*, in press.

[SW96] D. L Swets and J Weng. Using discriminant eigenfeatures for image retrieval *IEEE Bansactions on Pattern Analysas and Machine Intellzgence*, 18(8):831–836, August 1996.

[TP91] M **Turk** and **A.** Pentland. Eigenfaces for recognition *Journal of Cognitive Neurosczence*, 3(1):71–86, 1991

[VP96] T Vetter and T Poggio. Image synthesis from a single example image. In B Buxton and R. Cippola, editors, *Computer Vision ECCV'96*, number 1064 in Lecture Notes in Computer Science, page 652. Springer-Verlag, 1996.

Automatic Registration of Satellite Imagery

Leila M. G. Fonseca
National Institute for Space Research - INPE
São José dos Campos, SP, Brazil

Max H. M. Costa
State University of Campinas - UNICAMP
Campinas, SP, Brazil

B.S Manjunath and C. Kenney
Department of Electrical and Computer Engineering
University of California, Santa Barbara, **CA,** USA

Abstract

Image registration is one of the basic image processing operations in remote sensing. With the increase in the number of images collected every day from different sensors, automated registration of multi-sensor/multi-spectral images has become an important issue. A wide range of registration techniques has been developed for many different types of applications and data. The objective of this paper is to present an automatic registration algorithm which uses a multiresolution analysis procedure based upon the wavelet transform The procedure is completely automatic and relies on the grey level information content of the images and their local wavelet transform modulus maxima. The registration algorithm is very simple and easy to apply because it needs basically one parameter We have obtained very encouraging results on test data sets from the TM and SPOT sensor images of forest, urban and agricultural areas.

Keywords: wavelet transform, image registration, matching, satellite image.

1 Introduction

Image registration is the process of matching two images so that corresponding coordinate points in the two images correspond to the same physical region of the scene being imaged It is a classical problem in several image processing applications where it is necessary to match two or more images of the same scene [9] The registration process is usually carried out in three steps. The first step consists of selection of features in the images. Next each feature in one image is compared with potential corresponding features in the other image. A pair of features with similar attributes are accepted as matches and are called control points (CPs) Finally, using the control points, parameters of the best transformation which models the deformation between the images are estimated

A bottleneck in image registration has traditionally been the aquisition of control points. In remote sensing applications, users generally use manual reglstration which is not feasible in cases where there is a large amount of data. Thus, there is a need for automated techniques that require little or no operator supervision.

An important project at INPE is the study of the use dynamics and the land occupation through multitemporal data for the monitoring of the Amazon area [1]. The crucial phase in this process is image registration which is currently performed mannually, at a significant cost. This paper presents an automatic image registration algorithm which has obtained satisfactory results for registering multitemporal images of Amazon area. Some prelimary results has been recentely presented in [8]. An image pre-processing step was added whch helps in identifying a good set of control points for registration

The proposed algorithm is based on a multiresolution wavelet transform analysis. The registration procedure is realized basically in five steps: 1) processing of the images using a peer group filter which smooths the image while preserving the edges, 2) wavelet decomposition of the reference and sensed images to be registered, 3) extraction of feature points using the low-high and high-low components of the wavelet decomposition, 4)initial matching using the feature points and low-low component in the lower resolution, and 5) refinement matching in higher resolutions.

The matching process uses the correlation coefficient as a similarity measure and only the best pairwise fitting among all pairs of feature points are taken as control points. The matches pass by some strict tests in order to be considered correct. This conservative matching process reduces the number of mismatched pairs in the matching process and allows the use of smaller correlation window sizes. Nevertheless, some false matches will inevitably occur Therefore, a consistency-checking procedure is performed in order to eliminate incorrect matches and improve registration precision. After this we have a reliable initial guess for the registration transformation which is a crucial phase in the process. If the initial registration parameters are invalid the search for a registration transformation goes in a wrong direction, and the correct trend may not be recovered in later steps. A 2-D affine transformation is used to model the

deformation between the images and their parameters are estimated in a coarse-to-fine manner

The registration algorithm is very simple and easy to apply because it needs basically one parameter Because the matching is carried out only on the selected feature points and in a coarse-to-fine manner, a significant amount of computation is saved in comparison to traditional pixel-by-pixel searching methods.

Due to the fact that the registration procedure uses the grey level information content of the images in the matching process it is more adequate to register images of the same sensor or with similar spectral bands. In spite of this, it has demonstrated techmcal feasibility for many images of forest, urban and agricultural areas from Thematic Mapper (TM) and SPOT sensors taken in different times,

The organization of the paper is as follows, Section 2 gives an introduction to wavelet transform issues; Section 3 discusses the basic steps used in the registration algorithm; Section 4 shows preliminary experimental results of registering images taken from Landsat and SPOT satellites and Section 5 summarizes the work.

2 Wavelet Transforms

2.1 Definition

Wavelets are functions that satisfy certain mathematical requirements and are used in representing data or other functions.

Let $\psi(x) \in L^2(\mathbb{R})$ be a complex valued function. The basic wavelet $\psi(x)$ satisfies

$$\int_{-\infty}^{\infty} \psi(x)dx = 0. \tag{1}$$

The family of functions

$$\psi_{a,b}(x) = |a|^{-1}\psi(\frac{x-b}{a}) \; with \; a, b \in \mathbb{R}, a \neq 0 \tag{2}$$

generated from the basic wavelet under the operations of dilations (or scaling) by a and translations in time by b, are called wavelets.

The continuous wavelet Transform of a function $f(x) \in L^2(\mathbb{R})$ is given by the convolution

$$\begin{aligned} W_a[f(x)] &= |a|^{-1} \int_{-\infty}^{\infty} f(u)\psi(\frac{x-u}{a})du \\ &= f * \psi_a(x). \end{aligned} \tag{3}$$

where $\psi_a(x) = (1/a)\psi(x/a)$

The wavelet can be interpreted as the impulse response of a band-pass filter and the wavelet transform of a function as a convolution of this function with the dilated filter Thus the processing can be done at different scales or resolutions. If we look at the data with a large "window" (large a) we notice gross features. Similarly, if we look at the data with a small "window" (small a), we notice finer features [10].

In practice the scale a has to be discretized For a particular class of wavelets, the scale a can be sampled along a dyadic sequence $a = 2^j$ with $j \in \mathbb{Z}$, without modifying the overall properties of the transform [17]. The transforms corresponding to dyadic values of a are called discrete wavelet transform (DWT),

$$W_{2^j}[f(x)] = f * \psi_{2^j}(x) \tag{4}$$

The wavelet transform of 1-D can be easily extended to 2-D case. Interested readers can find more information on wavelets and the wavelet transform in [4], [7], [10] and [14].

2.2 Wavelet Transform Modulus Maxima

Let us call a smoothing function $\phi(x, y)$ the impulse response of a 2-D low-pass filter The first order derivative of $\phi(x, y)$ decomposed in two components along the x and y directions, respectively, are

$$\begin{aligned} \psi^1(x,y) &= \frac{\partial \phi(x,y)}{\partial x} \\ \psi^2(x,y) &= \frac{\partial \phi(x,y)}{\partial y}, \end{aligned} \tag{5}$$

and these functions can be used as wavelets.

For any function f, the wavelet transform at scale $a = 2^j$ defined with respect to these two wavelets has two components [16]:

$$\begin{aligned} W^1_{2^j}[f(x,y)] &= f * \psi^1_{2^j}(x,y) \\ &= f * (2^j \frac{\partial}{\partial x}\phi_{2^j}(x,y)) \\ &= 2^j \frac{\partial}{\partial x}(f * \phi_{2^j})(x,y) \\ W^2_{2^j}[f(x,y)] &= f * \psi^2_{2^j}(x,y) \\ &= f * (2^j \frac{\partial}{\partial y}\phi_{2^j}(x,y)) \\ &= 2^j \frac{\partial}{\partial y}(f * \phi_{2^j})(x,y) \end{aligned} \tag{6}$$

Therefore, these two components of the wavelet transform are proportional to the coordinates of the gradient vector of $f(x, y)$ smoothed by $\phi_{2^j}(x, y)$. They characterize the singularities along x and y directions, respectively [16]

We define the function

$$\begin{aligned} M[f(2^j, x, y)] = \Big(&\left|W^1_{2^j}[f(x,y)]\right|^2 \\ &+ \left|W^2_{2^j}[f(x,y)]\right|^2 \Big)^{1/2}, \end{aligned} \tag{7}$$

which is the modulus of the wavelet transform at the scale 2^j One can prove that for wavelets defined by Eq.(6), $M[f(2^j, x, y)]$ is proportional to the magnitude of the gradient vector

Mallat and Hwang [16] use the same approach as Canny [2] to detect the edge points of $f(x, y)$ at scale a They detect the points where the magnitude of the gradient is locally maximum in the direction where the gradient vector points to. At each scale 2^j, the modulus maxima of the wavelet transform are defined as points (x, y) where the modulus image $M[f(2^j, x, y)]$ is locally maximum, along the gradient direction

3 Registration Algorithm

Let us call the image to be warped the *sensed* image and the image to which the other image will be reduced the *reference* image. As in [6] we consider two cases for the image registration algorithm: 1)the images have the same ground resolution (pixel size); 2) the images are taken from different sensors and have different ground resolutions. In the second case the image with the highest resolution is reduced to the lower resolution This procedure is very simple since one knows a priori the spatial resolution of satellite images and the process of reducing the resolution can be realized automatically.

Next we describe each step involved in the registration process.

3.1 Image Preprocessing

Prior to registration, grey level preprocessing may be applied to the images in order to enhance their features. In this work we have used the Peer Group Averaging (PGA) filter which smooths the image while preserving the edges [11].

Peer group image processing identifies a 'peer group' for each pixel from a sliding window and then uses the peer group to make processing decisions at that pixel. The peer group for a pixel is selected from the window centered at the pixel and consists of the n pixels whose intensity values are closest to the center value. The center pixel is substituted by the average over the peer group. For our application the images were processed with PGA filter with a peer group of size $n = 4$ within a window 3x3.

3.2 Feature Point Detection

After reducing the images to the same spatial resolution we compute the discrete multiresolution wavelet transform (L levels) of the two images. For this computation we use the algorithm proposed in [15], whose implementation is computationally efficient

The wavelet decomposition of an image is similar to a quadrature mirror filter decomposition with the low-pass filter L and its mirror high-pass filter H [7] We call LL, LH, HL and HH the four images created at each level of decomposition, as in [12], where this decomposition is also used for the purpose of image registration.

The next phase aims to identify features that are present in both images in each level of the decomposition Here we use the modulus maxima of the wavelet transform to detect sharp variation points which correspond to edge points in the images The LH and HL subbands at each level of the wavelet transform are used to estimate the image gradient. For the wavelet decomposition we use the filters given in [3].

A thresholding procedure is applied on the wavelet transform modulus image in order to eliminate non significant feature points. Thus, a point (x, y) is recorded only if

$$M[f(2^j, x, y)] > \tau_{2j} , \tag{8}$$

where $\tau_{2j} = \alpha.\sigma_{2j} + \mu_{2j}$, α is a constant whose initial value is defined by the user and σ_{2j} and μ_{2j} are the standard deviation and mean of the wavelet transform modulus image at level 23, respectively The parameter α controls the number of feature points selected for the matching. Since the number of feature points increases in the finer resolutions the parameter α is also increased in the higher levels in order to select the most significant feature points in the images.

3.3 Initial Point Matching

The actual feature point matching is achieved by maximizing the correlation coefficient over small windows surrounding the points within the LL subbands of the wavelet transform Let the LL subbands of the sensed and reference images be f_s and f_r, respectively The correlation coefficient is given by

$$C_{f_s f_r}(x, y, X, Y) = \frac{1}{\sigma_s \sigma_r w_c^2} \sum_{i=-w_c/2}^{we12} \sum_{j=-w_c/2}^{wc/2} \{f_s(x+i, y+i) - \mu_s\}\{f_r(X+i, Y+i) - \mu_r\}, \tag{9}$$

where μ_s, σ_s, μ_r and σ_r are the local means (average intensity values) and standard deviation of the sensed and reference images, respectively, and w_c^2 is the area of matching window The correlation coefficient is one of the most traditional similarity measure and some results have indicated that its performance is one of the best [18]. This justifies its use in the matching process in this work.

The initial matching is performed on the lowest resolution images and is determined by the best pairwise fitting between the feature points in the two images. Let $P_s = \{f_s(x_i, y_j), i = 1, \ , N_s\}$ and $P_r = \{f_r(X_j, Y_j), j = 1, \ , N_r\}$ be the set of feature points detected in the sensed and reference images, respectively. Let T_c denote the threshold value above which two feature points are considered similar The point $f_r(X_k, Y_k)$ is the most similar feature point to $f_s(x_l, y_l)$ if

$$C_{f_s f_r}(x_l, y_l, X_k, Y_k) = \max_{1 \le j \le N_r} C_{f_s f_r}(x_l, y_l, X_j, Y_j). \tag{10}$$

Therefore, the matching process is achieved in the following way For each point $f_s(x_l, y_l) \in P_s$ all points $f_r(X_j, Y_j) \in P_r$ are examined and its most similar point $f_r(X_k, Y_k)$ is chosen. Next we test whether the achieved correlation is reasonably high. If $C_{f_s f_r}(x_l, y_l, X_k, Y_k) > T_c$, then $f_r(X_k, Y_k)$ is called "the best match" of $f_s(x_l, y_l)$ To verify that the match is consistent in the reverse direction, we test whether the best match of $f_r(X_k, Y_k)$ exists and is $f_s(x_l, y_l)$ If that is the case, both points are matched.

This two-way verification reduces the number of mismatched pairs in the matching process and allows the use of smaller window sizes. Nevertheless, some false matches will inevitably occur Therefore, a consistency-checking procedure similar to the one used in [13] is performed in order to eliminate incorrect matches and improve registration precision. The procedure is performed recursively in such a way that the most likely incorrect match is deleted first, followed by the next most likely incorrect match, and so on. This first part of the matching process is the crucial phase of the registration process. If the initial registration parameters are invalid the search for a registration transformation goes in a wrong direction, and the correct trend may not be recovered in later steps.

3.4 Image Warping

The above procedure provides a set of reliable matches which are used to determine a warping function that gives the best registration of the LL subbands to the precision available in level L of the wavelet transform

To model the deformation between the images a 2 - D affine transform with parameters (s, θ, Ax, Ay) is used [19]:

$$\begin{aligned} X = T_1(x, y) &= s[x\cos(\theta) + y\sin(\theta)] + Ax \\ Y = T_2(x, y) &= s[-x\sin(\theta) + y\cos(\theta)] + Ay, \end{aligned} \tag{11}$$

where (x,y) and (X,Y) are corresponding points in the sensed and reference image, respectively. This model is commonly used in remote sensing applications and is a good approximation for images taken under similar imaging directions [5], and which have been geometrically corrected (e.g., for Earth curvature and rotation) In most of remote sensing applications the images have a certain level of geometrical correction which enables the use of this kind of transformation

3.5 Refinement

The point matching and image warping steps can be performed at progressively higher resolutions in a similar fash-

ion to that described above. At each level $l < L$, the image f_s is transformed using the parameters estimated from lower resolution level $(l+1)$ In other words, the LL, LH, HL, and HH subbands at level $l+1$ are used to reconstruct the LL subband at level l, which is then warped by the transformation specified at the previous point matching operation.

Let f_s^t denote the warped (sensed) image. The refinement matching is achieved using the warped image and the set of feature points detected in the reference image. Each feature point $f_r(X_k, Y_k)$ detected in the image f_r at level l is matched to $f_s^t(x_l, y_l)$ if

$$C_{f_s^t f_r}(x_l, y_l, X_k, Y_k) = \max_{-w_r/2 \le m,n \le w_r/2} C_{f_s^t f_r}(x_l + m, y_l + n, X_k, Y_k), \quad (12)$$

where w_r is the size of the refinement matching window

The traditional measure of registration accuracy is the root mean square error (RMSE) between the matched points after the transformation, defined as:

$$RMSE = \left(\sum_{i=1}^{N} \left[\left(T_1(x_i, y_i) - X_i\right)^2 + \left(T_2(x_i, y_i) - Y_i\right)^2 \right] / N \right)^{1/2}, \quad (13)$$

where N is the number of matched points. This measure is used as a criterion to eliminate earlier matches which are considered imprecise. Poor matches are sequentially eliminated in a iterative fashion similar to that in [13] until the RMSE value is lower than 0.5 pixel

At each level the warping parameters are updated considering the refined list of matched points. After processing all levels the final parameters are determined and used to warp the original sensed image.

4 Experimental Results

In order to demonstrate the feasibility of the proposed algorithm for registering images of Amazon area some preliminary results are presented in this section In addition some results for images from urban and agricultural areas are also presented

We have used images from SPOT and Landsat-TM satellites in the experiments. All images tested are 512x512 pixels and were extracted from larger images to reduce processing time and disk usage.

The main parameters to be specified by the users are the number of levels of wavelet decomposition (L) and the value of a in Eq.(8). The wavelet decomposition is carried out up to the third level For all test images the process was realized with the parameters $a = 3$, $w_c = 11$, $w_r = 3$, and $T_c = 0.7$ The sensed images are warped using bilinear interpolation.

Experimental results are depicted in Figures 1 through 5. Some images shown in the figures are enhanced for purpose of display. Figure 1 shows the registration of two images taken from TM sensor, band 5 on different dates, 07/18/94 (TM945A) and 09/09/90 (TM905A) They correspond to an agricultural region near Itapeva, São Paulo. The image TM945A was taken as the reference one (Figure 1(a)) and TM905A as the sensed one (Figure 1(b)). The registration result is shown in Figure 1(e) Figures 1(b) and (c) show the control points obtained in the initial point matching phase which were used to estimate the initial parameters. These initial control points (denoted with "+" marks) are superimposed on the lowest LL subbands of the reference and sensed images. The images with the initial control points are zoomed up in order to vlsualize better the control points.

The images in Figure 2 are as in Figure 1 but the reference image (TM945RA) in Figure 2(a) is the result of registering manually TM945A with a map. Figures 2(c),(d) show the initial control points and Figure 2(e) shows TM945RA after transformation

Figure 3 shows the registration of two images from the urban area of São Paulo. A SPOT image , band 3, dated of 08/08/95 (SP953U), was reduced to a 30 m pixel size and taken as the sensed image (Figure 3(b)). A Landsat-TM image, band 4, dated of 06/07/94 (TM944U), was taken as the reference image (Figure 3(a)) The initial control points and the registration result are shown in Figures 3(c),(d) and Figures 3(e), respectively

Figure 4 shows the registration of Amazon region images taken from TM sensor, band 5, in different dates, 06/07/92 (TM925F) and 07/15/94 (TM945F). The image TM945F was taken as the reference one (Figure 4(a)) and TM925F as the sensed one (Figure 4(b)). Figures 4(c),(d) and (e) show the initial control points and the image TM925F after transformation, respectively.

Finally, Figure 5 shows the registration of Amazon region images taken from TM sensor, band 5, in different dates 07/07/97 (TM975F) and 08/03/95 (TM955F) The image TM955F was taken as the reference one (Figure 5(a)) and TM975F as the sensed one (Figure 5(b)). The initial control points and the image TM97F after transformation are depicted in Figures 5(c), (d) and Figure 5(e), respectively.

One can observe that although the images are of the same spectral band there are many differences between them whch make the registration problem difficult. For all test images the proposed registration algorithm obtained satis factory results. The use of the preprocessing filter helped in identifying a good set of control points, particularly for the Amazon images which show extensive texture data in the forest regions.

The parameters of the transformation and the number of control points (initial and final matching) for the test images shown in Figures 1-5 are listed in Table I, where the unit for Δx and Δy is in pixels and for θ is in degrees. The registration error (RMSE) is less than one pixel for all test images Our algorithm is reasonably efficient in terms of computational complexity The processing time depends on the image. Broadly speaking the process takes less than 1 minute on a SUN SPARC 20 workstation for a 512 x 512 pixel image.

5 Conclusion

We have described an automatic algorithm for the registration of satellite images. If the images have different spatial resolution the highest resolution image is reduced to the lower resolution and the processing is accomplished as for images of equal resolution The method is simple and requires few control parameters. The algorithm is performed at progressively higher resolution, which allows for faster implementation and higher regstration precision. It is more appropriate to register images taken from the same sensor or with similar spectral bands. Nevertheless, it worked well for images taken at different times and from urban, forest and agricultural areas, typical of remote sensing applications.

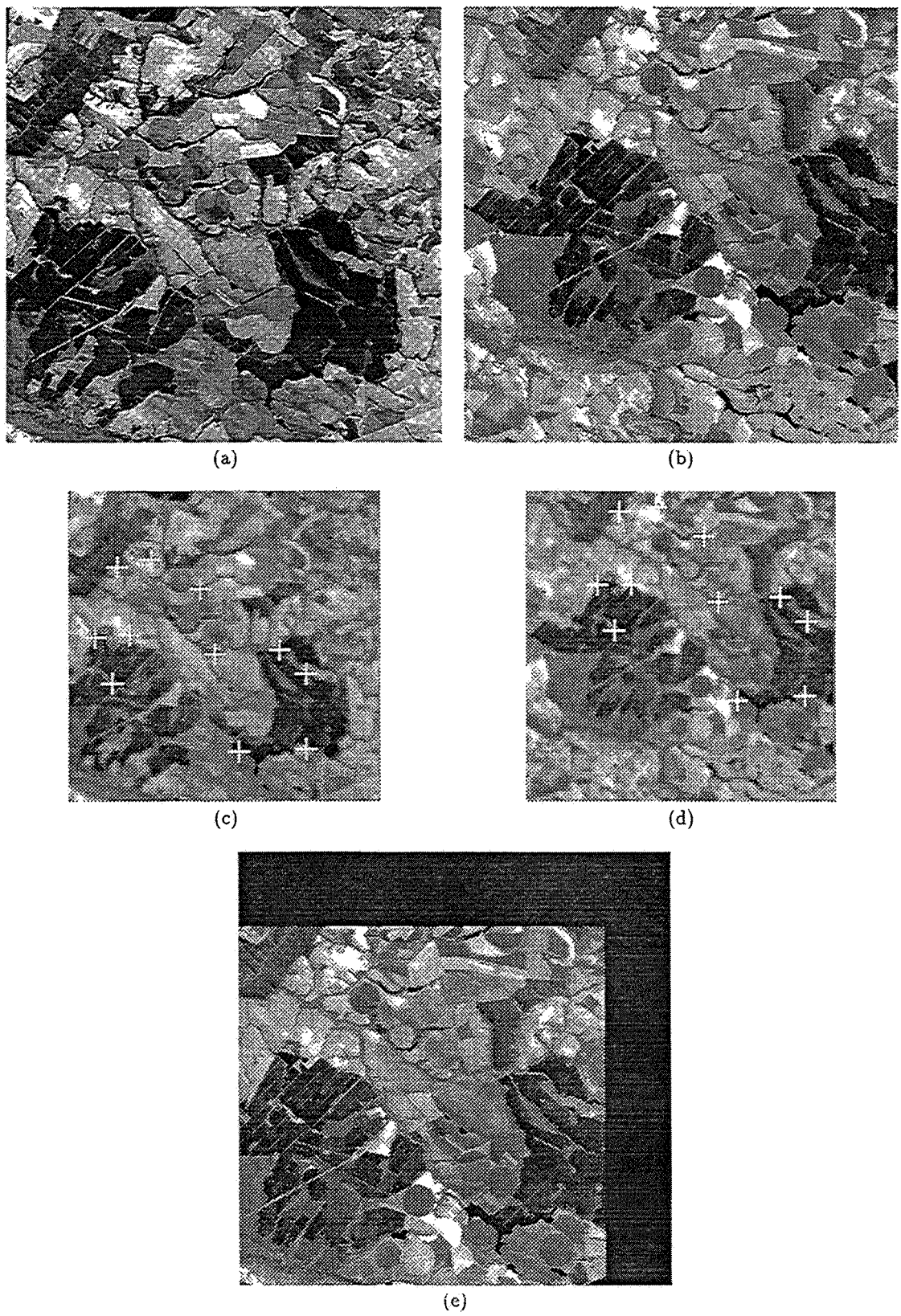


Figure 1 **Images from an agricultural area: (a) reference image** TM945A; (b) **sensed image** TM905A; **(c) and (d) show the initial CPs superimposed on LL subbands of** TM945A **and** TM905A, **respectively: (e) shows image** TM905A **after** final **transformation.**

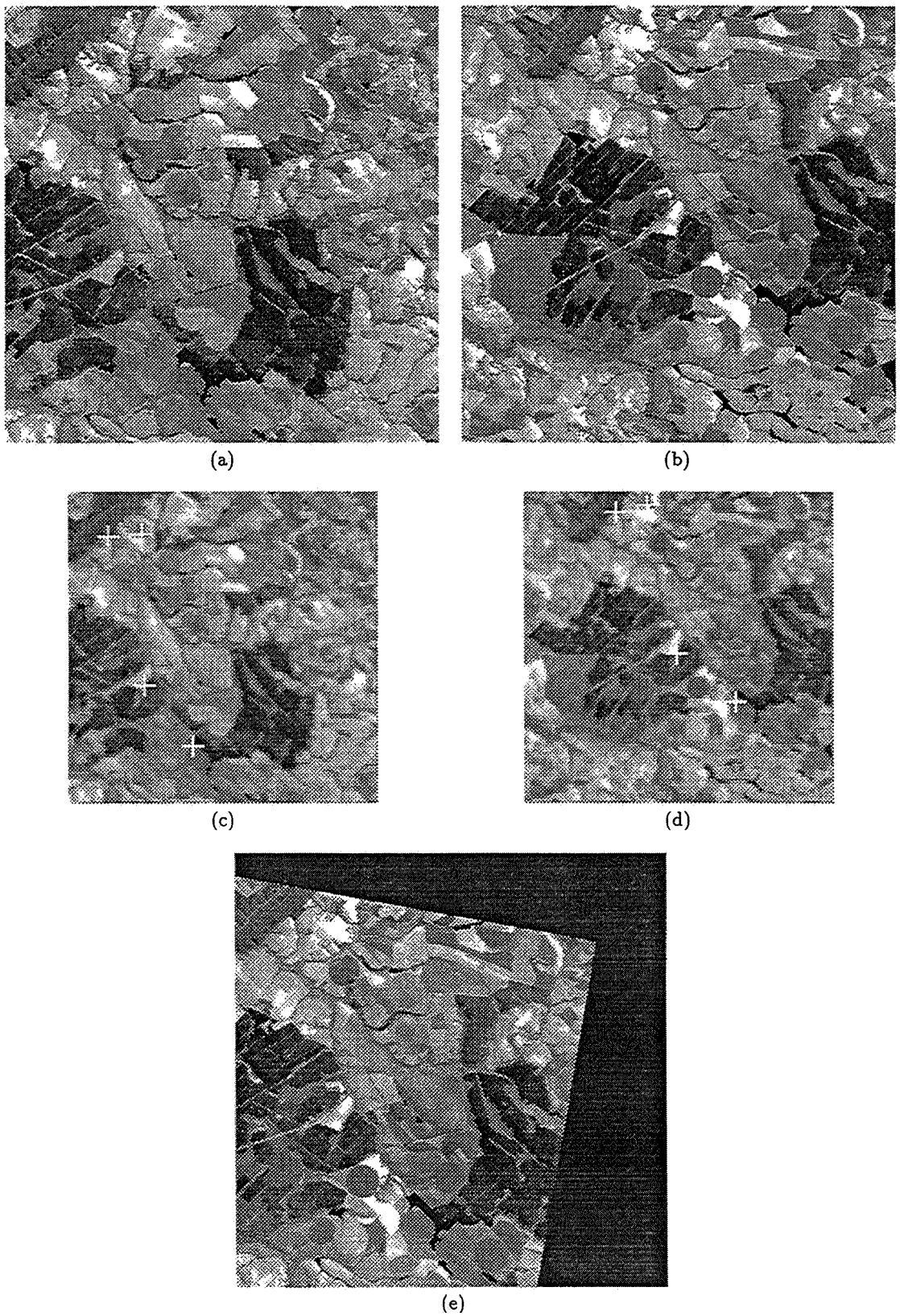


Figure 2: Images from an agricultural area: (a) reference image TM945RA: (b) sensed image TM905A; (c) and (d) show the initial CPs superimposed on LL subbands of TM945RA and TM905A, respectively; (e) shows image TM905A after final transformation.

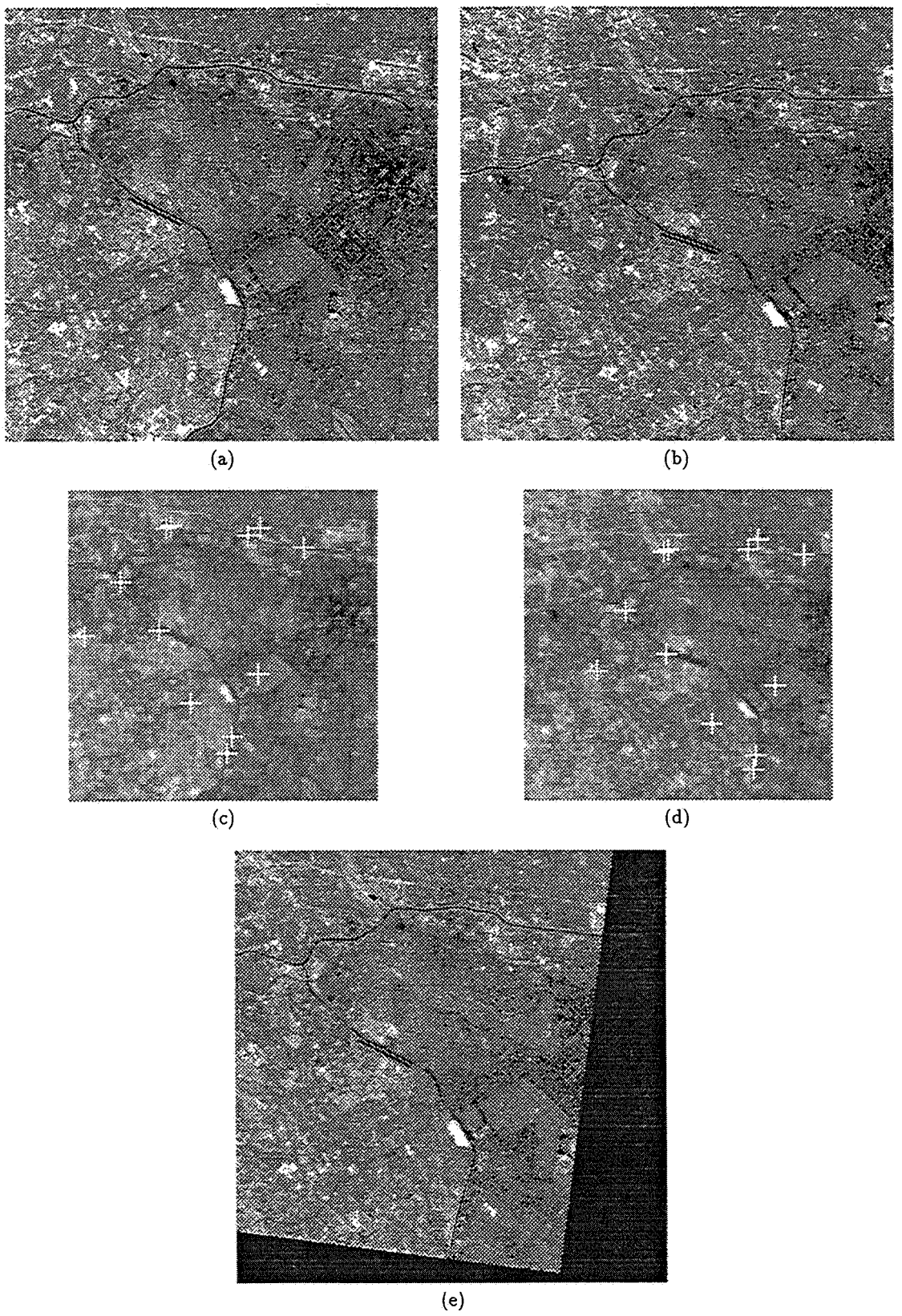


Figure 3: **Images** from **an urban area:(a) original TM image taken as reference one** (TM944U);(b) **resampled SPOT image taken as sensed one** (SP953U); **(c) and (d) show the initial CPs superimposed on LL subbands of** TM944U **and** SP953U, **respectively; (e) shows image** SP953U **after final transformation.**

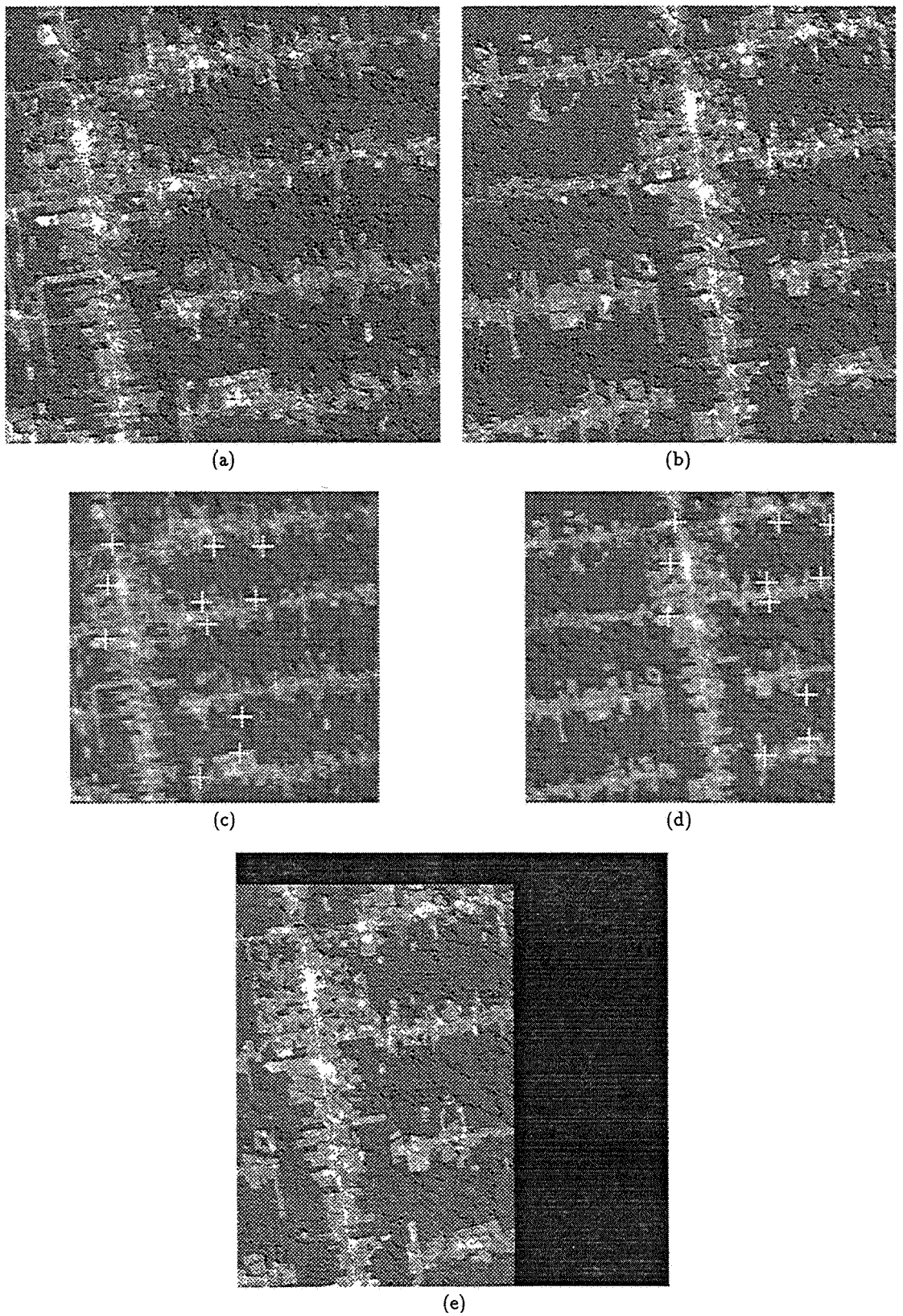


Figure 4. **Amazon region images: (a) reference image** TM945F; **(b) sensed image** TM925F: **(c) and (d) show the initial CPs superimposed on LL subbands of** TM945F **and** TM925F, **respectively: (e) shows image** TM925F **after** final **transformation.**

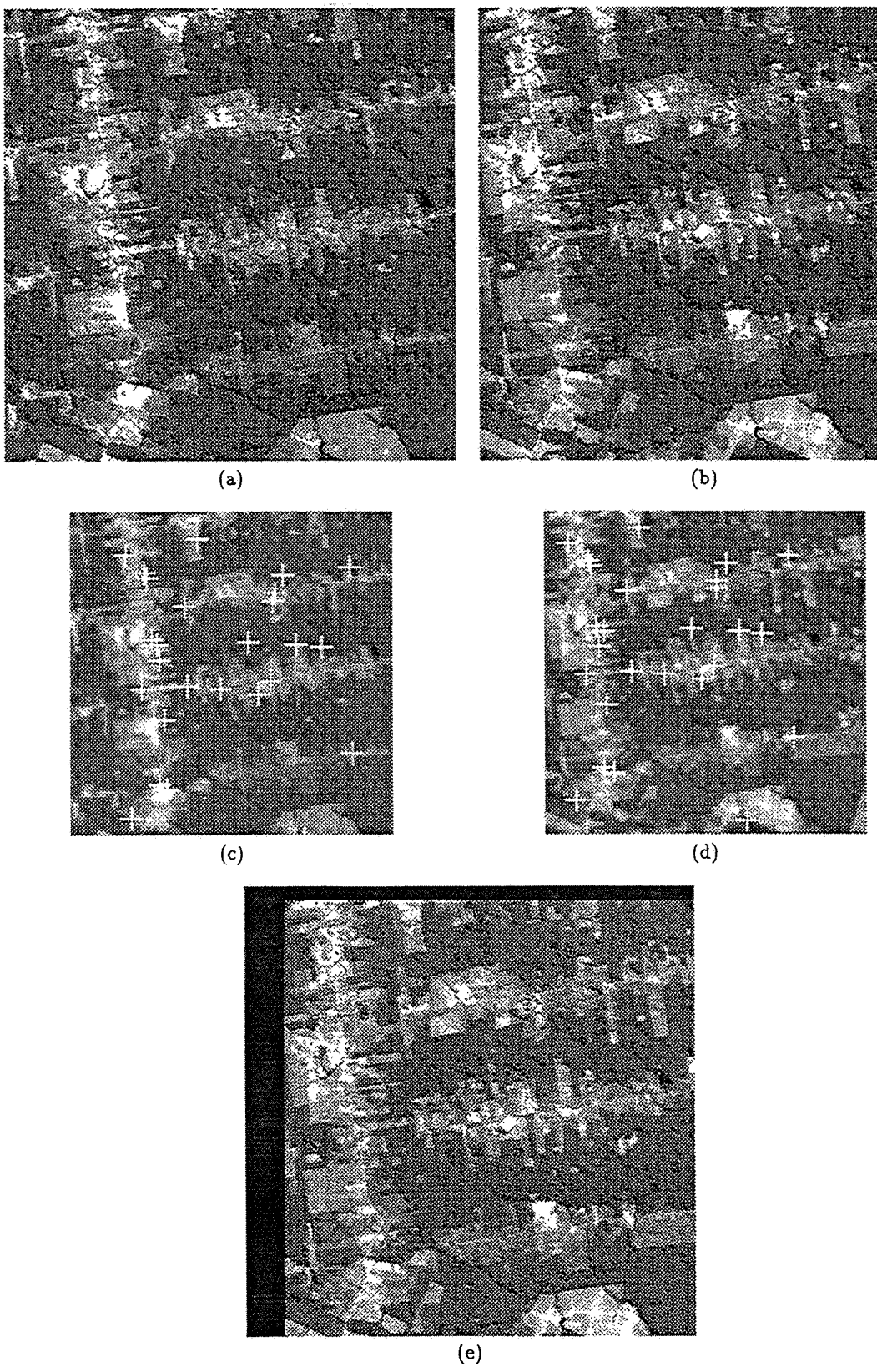


Figure 5: **Amazon region images: (a) reference image TM955F: (b) sensed image TM975F; (c) and (d) show the initial CPs superimposed on LL subbands of TM955F and TM975F respectively: (e) shows image TM975F after final transformation.**

Table I
Registration parameters for the examples shown in Figures 1 - 5

Images	θ	s	Δx	Δy	CPs	
TM905A TM945A	-0.02	1.00	87.8	-78.0	11	80
TM905A TM945RA	10.17	1.01	12.05	-82.3	4	42
SP953U TM944U	7.46	0.99	-70.9	-56.3	12	103
TM925F TM945F	-0.01	0.99	37.3	-182.1	11	39
TM975F TM955F	0.18	1.00	15.12	46.17	26	186

Acknowledgements

The authors would like to thank Dr Diogenes S Alves for providing the Amazon region images, Sergio D. Faria for providing the agricultural images and also Dr Gerard J F Banon for helpful discussions. This work was partially supported by CNPq (contract numbers 680061/94-0 and 301235/94-5), and in part by a grant from NASA (grant number NAGW-3951).

References

[1] D S. Alves. The amazon information system, In *ISPRS Congress,* pages 259–266, Washington, D C., 1992

[2] J F Canny A computational approach to edge detection *IEEE Trans on Pattern Anal. and Machzne Intell.,* 8(6):679–698, Nov 1986

[3] C Cheong, K Aizawa, T Saito, and M Hatori Subband image coding with biorthonormal wavelets. *EICE Trans. Fundamentals of Eletronecs, Communicatzons and Computer Sciences,* E75-A:871–881. July 1992

[4] C Chui. *An Introduction to wavelets.* C.Chui (editor), Academic Press, New York, 1992

[5] K Dana and P Anandan. Registering of visible and infrared images. In *Proceedings of the SPIE Conference on Architecture, Hardware and Forward-Looking Infrared Issues on Automatzc Target Recognitzon,* volume 1957, Orlando, FL, USA, 12-13, April 1993.

[6] J P Djamdji, A. Bijaoui, and R. Maniere. Geometrical registration of images: the multiresolution approach. *Photogramrnetrzc Engineering and Remote Senszng,* 59(5):645–653, May 1993

[7] N J Fliege. *Multirate Dzgital Szgnal Processing: multirate systems, filter banks, wavelets.* John Wiley & Sons, 1994.

[8] L. M. G. Fonseca and M Costa. Automatic registration of satellite images. In *Proceedings of the Tenth Brazilian Symposeum on Graphic Computatzon and Image Processing,* Campos de Jordão, SP, Brazil, To appear, 1997

[9] L. M G. Fonseca and B. S Manjunath Registration techniques for multisensor remotely sensed imagery. *Photogrammetric Engineerzng and Remote Senszng,* 62(9) 1049–1056, Sept 1996.

[10] A Graps An introduction to wavelets. *IEEE Computational Sczence and Engzneermg,* 2(2):50–61, Summer 1995.

[11] G. Hewer, C Kenney, B. S Manjunath, and R Schyoen. Peer group image processing. *ECE Technzcal Report TR # 97-02,* University of California, Santa Barbara, January 1997

[12] J Le Moigne. Parallel registration of multi-sensor remotely sensed imagery using wavelet coefficients. In *Proceedings of the SPIE - The International Society* for *Optical Engineering,* volume 2242. pages 432–43, Orlando, FL, USA, 5-8, April 1994

[13] H. Li, B S. Manjunath, and S. K Mitra, A contour-based approach to multisensor image registration *IEEE Transactzons on Image Processeng,* 4(3):320–334, March 1995.

[14] S G Mallat Multifrequency channel decompositions of images and wavelet models. *IEEE Transactions on Acoustics, Speech, and Szgnal processzng,* 37(12):2091–2110, Dec. 1989

[15] S. G Mallat. A theory for multi-resolution signal decomposition: the wavelet representation *IEEE Trans. on Pattern Anal and Machzne Intell.,* 11(7):674–693, July 1989

[16] S. G. Mallat and W H Hwang. Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory,* 38(2):617–643, March 1992.

[17] S. G Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Transactzons on Pattern Anal. Mach. Intell.,* 14(7):710–732, July 1992.

[18] R. Venkateswarlu Analysis of image registration algorithms for infrared images In *Proceedings of the SPIE - The International Society for Optzcal Engzneerzng,* volume 1699, pages 442–451, 1992.

[19] Q. Zheng and R A Chellappa. computational vision approach to image registration *IEEE Transactions on Image Processing,* 2(3):311–326, July 1993

Session II

General Methods

&

Resampling

Scope and applications of translation invariant wavelets to image registration

Samir Chettri [1], *Jacqueline LeMoigne* [2], *William Campbell* [3]

[1] GST and Code 935, [2] USRA/CESDIS, [3] Branch Chief, Code 935, NASA/GSFC.
Greenbelt, MD 20771.

ABSTRACT

The fist part of this article introduces the notion of translation invariance in wavelets and discusses several wavelets that have this property The second part discusses the possible applications of such wavelets to image registration. In the case of registration of affinely transformed images, we would conclude that the notion of translation invariance is not really necessary. What is needed is affine invariance and one way to do this is via the method of moment invariants. Wavelets or, in general, pyramid processing can then be combined with the method of moment invariants to reduce the computational load

Keywords: Image registration, translation invariant wavelets, Hu moment invariants, remote sensing.

1. INTRODUCTION

The fact that wavelets are not translation invariant was noted at least as early as [11]. Yet there are wavelets and wavelet algorithms that are provably translation invariant that have been applied in signal de–noising [3]. While the existence of translation invariant wavelets (TIW) is an important issue in its own right, the more important matter facing us is whether translation invariant wavelets (TIW) have any applications in image registration, in particular, those images that are obtained from earth orbiting satellites. Since what we really need is affine invariant image registration, we conclude that TIW have limited applications in image registration. Given this fact we suggest that an effective way to do affine invariant wavelet processing is by combining pyramid processing (including wavelet processing) and moment invariant based methods.

2. WAVELET PROCESSING

The concept of wavelets simultaneously arose in several fields including computer vision (image pyramids), signal processing (filter banks) and mathematics (atomic decompositions) but has reached its maturity with the work of Mallat, Meyer and Daubechies [4]. In wavelet analysis we consider the space of square integrable (finite energy) signals $V = \mathcal{L}^2(R)$. A multi–resolution is the decomposition of the space V into a sequence of subspaces

$$V_{-1} \subset V_0 \subset V_1 \subset V_2 \qquad (1)$$

such that each V_n consists of signal approximations at resolution 2^n We also require that the union of all the subspaces V_n, $n = -\infty, \ldots \infty$ is the space V and that the zero element is the only element common to all the subspaces. Now, if there exist functions ϕ that satisfy a translation–invariance ($\phi(x) \in V_m \Leftrightarrow \phi(x - 2^{-m}r) \in V_m$) and a scale–invariance ($\phi(x) \in V_m \Leftrightarrow \phi(2x) \in V_{m+1}$) constraint, and furthermore, if ϕ is an orthonormal basis for V_m then any finite energy signal can be written as a series expansion of $\phi_{n,m} = 2^{m/2}\phi(2^m x - n)$. The ϕ functions are called the "father wavelets" or, more commonly, the scaling functions, and correspond to a low–pass or averaging filter

Given the above assumptions one can show the existence of a basic ("mother") wavelet $\psi(x)$ which forms an orthonormal basis for the space V In keeping with the dyadic decomposition of the spaces the multi–resolution analysis splits the signal into band pass (the "wavelet" part) and low pass (the "scaling function" part) signals.

A useful way of representing the wavelet transform is through the use of the filter bank formalism In fact a detailed study of filter bank theory is essential for fast implementations of the wavelet transform in hardware or software. The filter bank representation is best seen as an unbalanced tree where one branch is recursively expanded as shown in Figure 1 In this figure h_0 is a low–pass filter and h_1 is a band–pass or high–pass filter Let us consider the upper branch of the tree. The input signal is $x(n)$. It is low–pass filtered and then decimated by 2, i.e., every other sample is discarded, as represented by the symbol $\downarrow 2$. The output after lowpass filtering is a_1 This is treated as the input signal to another set of low and band–pass filters to be followed by decimation. To inverse the transform we reverse our path along the different branches of the tree, placing an adder wherever two branches meet. The wavelet transform is known as a perfect reconstruction transform because the inverse transform gives us back the original signal exactly.

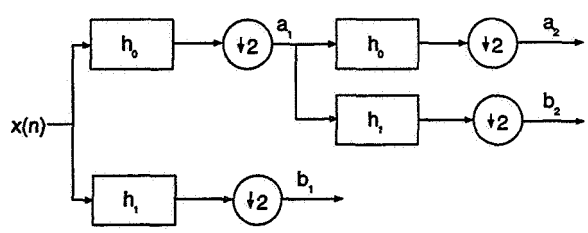


Figure 1. Filter bank representation of wavelet decomposition

To do the two dimensional wavelet transform, we treat rows and columns independently, i.e., we use tensor products of the one dimensional wavelets. This process is shown in Figure 2. Here all the rows are wavelet transformed into low and high

frequencies respectively. Next, the columns are transformed similarly. We can continue this process by applying the same decomposition to only the low frequencies which form a sub-image whose sides are half the length of the original image.

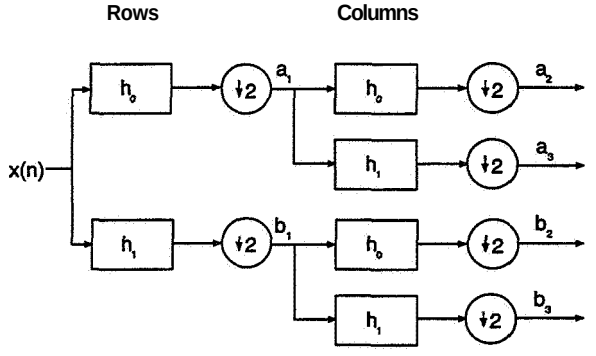


Figure 2: Two dimensional wavelet transform using filter bank representation

The resulting two dimensional wavelet decomposition is shown schematically in Figure 3 for a single level of the decomposition. In Figure 3, an N x N image is reduced to four $N/2$ x $N/2$ images after a wavelet transform. Here *LL* represents the averaged image, *LH* and *HL* represent horizontal and vertical edges respectively and *HH* represents all the high frequency components in the image and in addition may also represent some textural information. The wavelet analysis may then be performed recursively on the *LL* image, leading to a pyramid decompostion from which the onginal image can be completely recovered.

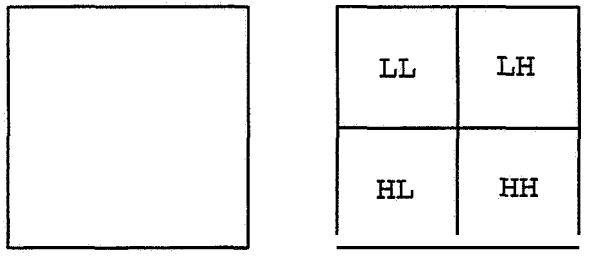


Figure 3: Image and its multi–resolution image decomposition

3. TRANSLATION INVARIANCE

The notions of translation invariance and variance in wavelets are best studied by an example. Figure 4 shows a signal and its unit–translated version in (a). Similarly (c) shows the same signal and a version of the signal that is translated by two samples to the left. Panel (b) shows the wavelet transform of the two signals in (a) Note that there is no correlation between the two wavelet transforms. F i i y panel (d) shows the wavelet transforms of the signals in (c). The wavelet transform of the original signal and its two–shift version are related by a shift. The wavelet we have used in this processing is the DB4 wavelet, i.e., the four coefficient Daubechies wavelet as described in [4]. This example illustrates the definition of translation invariance in the wavelet literature: a shift of the **original** signal, leads to the output being shifted by the same amount. By this definition, the wavelet transform using Daubechies' wavelets is not shift invariant though every signal that is shifted by integer multiples of 2 leads to a wavelet representation that is shifted by the corresponding integer multiple as illustrated in Figure 4.

In the following sub–sections, we study some approaches to translation invariance in the literature. In each case we present the basic theory (for more details, please read the corresponding references) as well as some of the computational issues.

3.1. Beylkin–Coifman–Donoho (BCD)

In this method the translation invariance is reduced to a problem of finding the wavelet transforms of all possible circular shifts of the input signal and then using an information theoretic approach to finding the best possible shift The speed of the algorithm comes from Beylkin's observation that all possible circular shifts of a signal can be wavelet transformed in $\mathcal{O}(N \log N)$ time as opposed to $\mathcal{O}(N^2)$ time. The speed of the method is due to the observation (as shown in Figure 4) that all even shifts (i.e., shifts by $2k$, $k \in \mathbb{Z}$) of a signal lead to the same wavelet transform output shifted by k, with a similar observation for all odd shifts. The more difficult job is in the choosing of the single shift that is the "most translation-invariant," An approach to translation invariant denoising of signals mentioned in [10] (page 53) and attributed to Coifman is as follows: Take several positive and negative shifts of the signal and form the wavelet transforms of each of them. Denoise each of them and invert the transform and then take the average of the reconstructed signals with the translations undone.

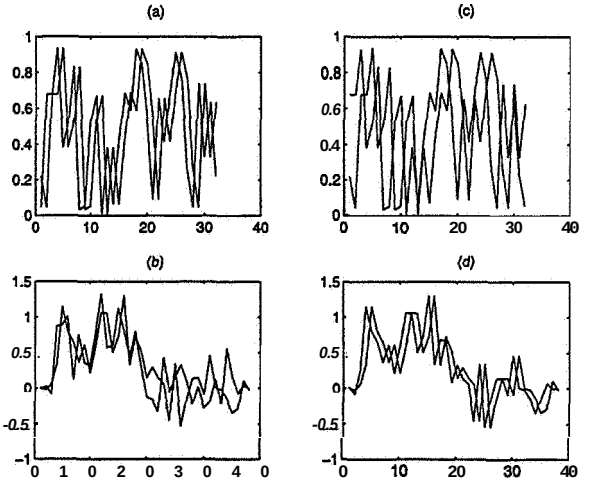


Figure 4: Translation invariance in wavelets. (a) A random signal and its unit shifted version. (b) The wavelet transforms of each signal in (a) plotted on the same graph Note the lack of correlation between the wavelet transforms, (c) The same signal and a version of it that is translated two samples to the right. (d) Wavelet transforms of each signal in (c) Note that one wavelet transform is a translated version of the other

We can use the Beylkin algorithm for fast computation of all

shifts of a signal to create a translation invariant registration algorithm. Given two signals that we wish to compare (i.e., register) we would take the first signal and compute the wavelet transform for several shifts and average them Similarly for the second signal. We would next compare the DWT coefficients for each signal if they are similar, then we can conclude that one signal is a translated version of the other

3.2. Walter

Walter [12] approaches the problem by introducing the notion of weak translation invariance in multi–resolution space. Here, each subspace V_m in not translation invariant, rather translations in V_m are only invariant in the next space up the multi–resolution ladder V_{m+1} Walter proves that the Meyer wavelets are weakly translation invariant and Madych [7] has shown that the Sinc wavelets (or Shannon wavelets) are translation and dilation invariant (i.e., not weakly invariant, but completely invariant) Below we present a basic summary of the theory as found in [12].

Definition. Let $\{V_m\}$ be a multiresolution analysis of $\mathcal{L}^2(R)$. Then $\{V_m\}$ is said to be weakly translation invariant if the translation operator T_α, $a \in \mathcal{R}$ maps $\{V_m\}$ into $\{V_{m+1}\}$ Note: $(T_\alpha f)(x) = f(x - \alpha)$.

Theorem[12] Let the function $\phi(t)$ be given by

$$\hat{\phi}(\omega) = \left\{ \int_{\omega-\pi}^{\omega+\pi} h \right\}^{1/2} \tag{2}$$

Where, h is a positive distribution (i.e., a measure) with support in $[-\epsilon, \epsilon]$, $0 \leq \epsilon \leq \pi/3$, such that $\hat{h}(0) = 1$ Then $\phi(t)$ is a scaling function whose Fourier transform has support on an interval and its associated MRA $\{V_m\}$ is weakly translation invariant.

As an example, if we choose $h(\omega) = \delta(\omega)$ we get the Sinc function $\phi(t) = \sin(\pi t)/\pi t = \text{sinc } t$ whose Fourier transform is the ideal lowpass filter We can show that the Sinc function or Shannon wavelet is completely translation invariant, i.e., the translation operator T_α maps $\{V_m\}$ into $\{V_{m+1}\}$, $\{V_{m+2}\}$ V_{m+3} Generalizations of such wavelets have been studied in [8] and [2] for time–frequency analysis of musical signals. Similarly, one can show that a slightly different choice of $h(w)$ leads to the Meyer wavelets and that these are only weakly translation invariant.

3.3. Simoncelli

Both the BCD and the Walter approaches still maintain the orthogonality of the basis functions. On the other hand Simoncelli [11] and others [6] give up this condition and create oversampled representations. The advantage of this approach is that one gains both translation and rotation invariance. The disadvantage is that exact reconstruction is no longer possible, but because they satisfy the frame condition [4] near-perfect reconstruction is possible. We would point out that in the case of image registration, exact reconstruction is not really necessary. Simoncelli's approach is based on the concept of steerable filters. A steerable filter is defined as a set of filters at arbitrary orientations. A filter at any orientation is constructed as a linear combination of a set of basis filters. The steering constraint is written in equation (3) as follows:

$$f^\theta(x, y) = \sum_{j=1}^{m} k_j(\theta) f^{\theta_j}(x, y) \tag{3}$$

Where, $f^\theta(x, y)$, the steerable filter at any orientation θ is written as a linear combination of some "basis" filters $f^{\theta_j}(x, y)$. As an example, the familiar derivative of gaussian filters is steerable,

3.4. Problems

The problem with any translation or even approximately translation invariant wavelet transform is that it might not be dilation invariant or rotation invariant etc. Generally speaking, most of the above transforms are not rotation, dilation or gray scale invariant. And while the steerable pyramid is both rotation and translation invariant, it is not dilation invariant. However the disadvantage of these steerable pyramids is that they require $4k/3$ elements of storage. Where, k is the number of orientations at which we filter the original image. Quite clearly one can find specific problems or problem areas where one signal/image is only a translated version of the other Under those circumstances we could use any of the above algorithms described above. However, in the general case none of the above methods will succeed Based on the above discussion we now discuss a combmation of wavelets and algebraic invariants for the registration problem

4. ALGEBRAIC INVARIANTS, WAVELETS AND REGISTRATION

As discussed earlier, in general, the orthogonal wavelets will not be translation, dilation, rotation and gray–scale invariant. While one can create wavelets with some of these properties, creating wavelets with all of these properties seems to be a difficult task. Due to this it is proposed that we use the orthogonal wavelets to perform a pyramid decomposition of the image. We note that since the LL band of the decomposed image is generally a smoothed version of the original image, large features in the original image are also seen in the decomposed image and they generally have the same shape. Of course we can quantify the statement "generally have the same shape," by using invariant moments designed by Hu in 1962 and revised in [9].

The Hu/Reiss moment invariants are obtained from the central moments as defined below:

$$\mu_{pq} = \int_{\mathcal{O}} (x - x_t)^p (y - y_t)^q dx dy. \tag{4}$$

In equation (4), x_t and y_t are the x and y coordinates of the centroid of the binary object $(\mathcal{O})$ under consideration. We also note that μ_{00} represents the area of $\mathcal{O}$

The μ_{pq}'s are not invariant to affine geometric distortions. However, different algebraic combinations of them are. These are listed below and may also be found in [5] and [1]

$$I_1 = \mu_{00}^{-4}(\mu_{20}\mu_{02} - \mu_{11}\mu_{11}) \quad (5)$$

$$I_2 = \mu_{00}^{-10}(\mu_{30}^2\mu_{03}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} + 4\mu_{30}\mu_{12}^3 - 3\mu_{21}^2\mu_{12}^2) \quad (6)$$

$$I_3 = \mu_{00}^{-7}(\mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) + \mu_{02}(\mu_{30}\mu_{12} - \mu_{21}^2)) \quad (7)$$

Thus, by combining algebraic invariants and wavelets we can do image registration. Given two images we wish to register, the method to do this is as follows:

1. Take the wavelet transforms of the two images and use the *LL* images of each.
2. In each *LL* image, perform a region segmentation, or a cluster analysis to identify closed objects in the scene.
3. Compute the moment invariants of these objects in *each LL* scene.
4. Match the objects with closest moment invariants in *each LL* scene.
5. Obtain the centroids of these objects and perform a coarse registration, i.e., transform one image to coarsely register with the other
6. Use the inverse wavelet transform to get the original images back (or if one has stored the original images, ignore this step)
7. The coarse objects in the *LL* scenes are now available at higher resolution. Use our prior knowledge of the objects in the *LL* scene to obtain the higher resolution objects. (It may be necessary to do the moment invariant calculations and comparisions of the objects again at this stage). Object centroids are determined again and these are used to get the final registration.

The advantage of such a method is basically the advantage of the pyramid methods in general, i.e., we work in a physically smaller space, thus reducing the total amount of computation. Also, the propagation of the results from a low resolution space to a higher resolution space iteratively increases accuracy. We note that the method listed above is a multiresolution version of the algorithm discussed in [5].

In order to test the method described above, we have worked with a TM image of the Pacific North West (Figure 5 (a)) and its rotated and translated version (Figure 6 (a)). We start with the *LL* versions of the original images (Figure 5 (c) and Figure 6 (c)). Since the scale is small, we also display zoomed versions of Figure 5 (c) and Figure 6 (c) in Figure 5 (d) and Figure 6 (d) respectively and similarly for the untransformed version in Figure 5 (b) Figure 6 (b). In particular we choose the "key shaped object" located at about [280 450] in Figure 5 (b) and correspondingly in the other images for moment invariant analysis.

The moment invariants I_1, I_2 and I_3 are computed for the key shaped figures in the all the four zoomed images and are tabulated below Observation of Table 1 shows that the first two invariant moments are not identical but very close

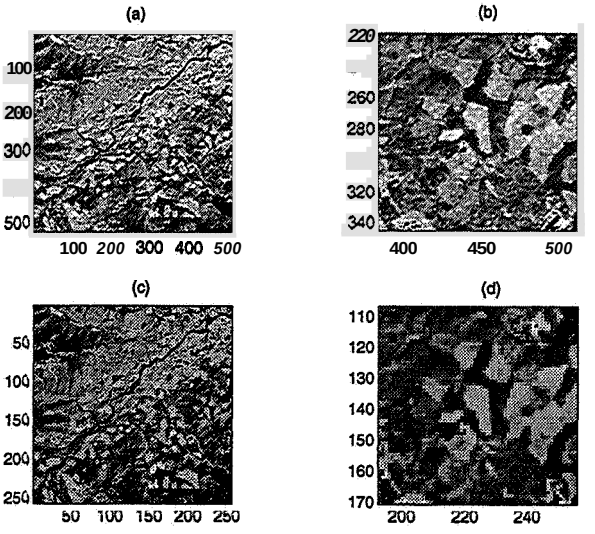


Figure 5: Wavelet transform preserves shape. (a) A TM image of the Pacific North West. (b) A zoomed version of the right middle part of (a). (c) A wavelet transformed version of (a) showing only the *LL* portion. (d) A zoomed version of the right middle part of (c). Note that even though (d) is a half resolution version of (b), object shapes are preserved. We chose the key shaped object at [280, 450] in panel (b) for analysis.

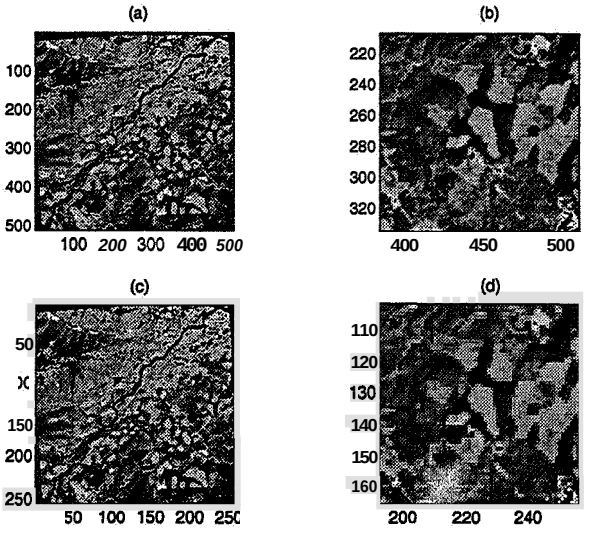


Figure 6: More on how wavelet transform preserves shape. (a) A TM image of the Pacific North West as in Figure 5 but rotated and translated. (b) A zoomed version of the right middle part of (a). (c) A wavelet transformed version of (a). showing only the *LL* portion. (d) A zoomed version of the right middle part of (c). Note that even though (d) is a half resolution version of (b), object shapes are preserved. *Also,* compare similar resolutions in this figure and Figure 5. We chose the key shaped object at [270, 445] in panel (b) for analysis.

Figure	I_1	I_2	I_3
Figure 6(b)	2.664111e-01	3.875606e-02	4.764094e-03
Figure 6(d)	2.764252e-01	3.938831e-02	4.719842e-03

Table 1 Invariant moments for the key shaped figure in Figures 5(b), 5(d), 6(b) and 6(d)

to one another and may be used for discrimination. What is surprising is that there is very little difference between the moment invariants of the keys in the LL images when compared to the keys in the corresponding higher resolution images.

5. CONCLUSIONS AND FUTURE WORK

We started our discussion on the use of translation invariant wavelets. Since this class of wavelets only exhibits translation invariance (or weak translation invariance), their application to the general registration problem is restricted However, general invariance may be obtained through the use of the algebraic moment invariants. Wavelets enter the picture because of their well known ability to do non–redundant (i.e., non over-sampled) pyramid decompositions. We discussed an algorithm for algebraic invariant based wavelet registration and showed its application to a remotely sensed image. We are looking to automate the image segmentation procedures that we currently use.

6. ACKNOWLEDGEMENTS

Many thanks to all the members of the registration group at NASA/GSFC for their help at various stages along the way. Allen Lunsford and Charlie Hoisington spent much time explaining the registration problem as it pertains to Code 935 data. Their help is greatly appreciated.

References

1. Belkasim, S. O. et. al. "Pattern recognition with moment invariants. A comparative study and new results." *Pattern Recognitton,* Vol 24, pp. 1117-1138, 1991.
2. Chettri, S. R. et al. "Harmonic wavelets, constant *Q* transforms and the cone kernel TFD." In *Wavelet Applications III,* ed. Szu, H H., Vol. 2762, pp. 446–451, 1996.
3. Coifman, R. R. and Donoho, D. L. "Translation-Invariant De–Noising." In *Wavelets and Statistics,* eds. Antoniadis, A and Oppenheim, G., pp. 125-150, Springer–Verlag, 1995.
4. Daubechies, I. *Ten lectures on wavelets.* SIAM, 1992
5. Flusser, J and Suk, T "A moment–based approach to registration of images with affine geometric distortion." *IEEE Trans. on Geosci. and Rem, Sensing,* Vol. 32, pp. 382-387, 1994.
6. Freeman, W T and Adelson, E. H. "The Design and Use of Steerable Filters." *IEEE Trans. PAMI,* Vol. 13, No. 9, pp. 891-906, Sept 1991
7. Madych, W R. "Some elementary properties of multiresolution analysis in $\mathcal{L}^2(R)$." In *Wavelets - A Tutorial in Theory and Applicatrons,* ed. Chui, C.K., pp. 259-294, Academic Press, 1992.
8. Newland, D. E. "Harmonic and musical wavelets." *Proc. R. Soc. Lond. A,* Vol. 444, pp. 605–620, 1994.
9. *Reiss,* T H. "*The*revised fundamental theorem of moment invariants." *IEEE Trans. PAMI,* Vol. 13, pp. 830–834, 1991.
10. Saito, N. *Local Feature Extmctron and Its Applications Using a Library of Bases,* Ph.D. Dissertation, Yale University, 1994.
11. Simoncelli, E. P et. al. "Shiftable Multi–scale Transforms." *IEEE Trans. Inf. Theory,* Vol. 38, No. 2, pp. 587–607, 1992.
12. Walter, G. G. "Translation and Dilation Invariance in Orthogonal Wavelets." *Applied and Computational Harmonic Analysis,* Vol. 1, pp. 344-349, 1994.

55-61

A Scale Space Feature Based Registration Technique for Fusion of Satellite Imagery

547753

340114

P2

Srini Raghavan, Robert F Cromp, and William C. Campbell
Code 935, NASA Goddard Space Flight Center
Greenbelt, Maryland 20771

Feature based registration is one of the most reliable methods to register multi-sensor images (both active and passive imagery) since features are often more reliable than intensity or radiometric values. The only situation where a feature based approach will fail is when the scene is completely homogenous or densely textural in which case a combination of feature and intensity based methods may yield better results. In this paper, we present some prelimmary results of testing our scale space feature based registration technique, a modified version of feature based method developed earlier for classification of multi-sensor imagery (Raghaven et. al 1996). The proposed approach removes the sensitivity in parameter selection experienced in the earlier version as explained later

Our approach makes use of a coarse-to-fine matching scheme to match features such as edges and line segments extracted at multiple scales fiom the wavelet transform of the image. This matching scheme allows one to match the imagery at arbitrary resolutions with arbitrary translation and rotation. For the preliminary investigation of this technique, we assume that the alignment between the two images is a similarity transformation, though the extension to a more complex transformation is fairly straightforward. The input to our technique are line segments in both images extracted using the feature points from the HIGH-LOW and LOW-HIGH components at multiple scales of the wavelet transform. The line segments are formed fiom the feature points using a simple cooperative-competitive neural network scheme (Raghavan et. al. 1995).

The following are the major computational steps of the algorithm.

Step 1 Compute the wavelet transform for the reference and input image.

Step 2: Fill-in the feature points using the neural network based oriented smoothing technique. Extract line segments from the filled edges.

Repeat the following steps 3:9 until completion a the finest level of the scale space

Step 3: For the line segments at the next coarsest level, perform a Hough transform on the rotation of the line segments in one image to the line segments in the other. For each line segment in the reference image, a complete list of possible parings of line segments in the other image (possibly within a certain spatial window of search) is obtained. For each potentia1 pairing the orientation difference is computed and histogrammed. Each bin of the histogram corresponds to a small angle of rotation, usually 2-4 degrees.

Step 4: The next step is to locate the peak bin(s) in the histogram. We can treat the neighboring bins as potential candidates as well. It may be noted that the peak bin contains the potential corresponding edge pairings and hence to a certain extent the correspondence problem can be viewed as solved this way

Step 5 De-rotate the second image by the angle corresponding to the peak bin.

Step 6: The next step is to take any three non-parallel matching pairs of line segments fiom the two images found in the peak bin and create a virtual triangle for each image for a given three-tuple. The vertices, area and the centroid of the virtual triangle, are computed.

Step 7: Use the area of the virtual triangles to estimate the scale parameter The estimated scale values are also histogrammed to prevent the outliers (due to accidental alignment of line segments in the peak bin) from influencing the estimation of scale.

Step 8: Estimate the translation parameters fiom the centroids of virtual triangles that correspond to the best scale and rotation values.

Step 9: Redefine the bin resolutions for the Hough transform to be performed at the next finer level.

A major advantage of using the scale space coarse-to-fine filtering method is that it helps mitigate the false peaking problem commonly encountered in Hough transform. Thus, the bin widths can be broad at the coarsest scale and can be progressively made smaller and sensitive as we approach the finest level.

The full length final version of this paper will include the results of our testing with synthetic and real data. Accuracies of feature based techniques are, in general, limited to one or two pixels due to different sources of error from the feature extracting techniques. However, they are quite robust due to their relative stability in the presence of noise, intensity changes, and view point difference in the scenes. Further improvement can be achieved by using a correlation approach used in localized fashion.

An Optical Systems Analysis Approach to Image Resampling

Richard G. Lyon
CESDIS/UMBC
Code 930.5
Greenbelt, MD 20771
(301)-2864302
lyon@nibbles.gsfc.nasa.gov

Abstract

All types of image registration require some type of resampling, either dunng the registration or as a fmal step in the registration process. Thus the image(s) must be regridded into a spatially uniform, or angularly uniform, coordinate system with some pre-defined resolution. Frequently the ending resolution is not the resolution at which the data was observed with. The registration algorithm designer and end product user are presented with a multitude of possible resampling methods each of which modify the spatial frequency content of the data in some way The purpose of this paper is threefold: (i) to show how an imaging system modifies the scene from an end to end optical systems analysis approach, (ii) to develop a generalized resampling model, and (iii) empirically apply the model to simulated radiometric scene data and tabulate the results.

A Hanning windowed sinc interpolator method will be developed based upon the optical characterization of the system. It will be discussed in terms of the effects and limitations of sampling, aliasing, spectral leakage, and computational complexity

Simulatedradiometric scene data will be used to demonstrate each of the algorithms. A high resolution scene will be "grown" using a fractal growth algorithm based on mid-point recursion techniques. The result scene data will be convolved with a point spread function representing the optical response. The resultant scene will be convolved with the detection systems response and subsampled to the desired resolution. The resultant data product will be subsequently resampled to the correct grid using the Hanning windowed sinc interpolator and the results and errors tabulated and discussed.

(1.0) Generalized Imaging Model

The generalized imaging model for a conventional incoherent imaging system is given by'

$$I(x,y;\lambda,t) = PSF(x,y;\lambda,t) **O(x,y;\lambda,t) \quad (1)$$

Where $O(x,y;\lambda,t)$ is the object or scene under study as a function of spatial coordinates (x,y) of the wavelength (λ) and if the scene is tune dependent then it can also be a function of time. $PSF(x,y,\lambda,t)$ is the optical system Point Spread Function and represents the imaging response due to an unresolved point source i.e. a delta function. "**" denotes 2 dimensional spatial convolution. The PSF is in general a strong function of wavelength and if one is viewing through the atmosphere or other turbulent medium then it is also a stochastic function of time. In general

when viewing through a turbulent medium the convolutional relation no longer strictly holds since the PSF will be a function of the field angle. However for small regions the convolutional relationship holds; the region over which the convolutional relationship holds is commonly referred to as the isoplanatic region or patch. Note that $I(x,y;\lambda,t)$ would be the image differentially close to the detector array, since none of detector effects, such as spectral response, finite pixel area and finite integration time as well as noise effects have yet to be folded in.

The PSF for an incoherent imaging system is given by:

$$PSF(x,y,\lambda,t) = \frac{1}{\lambda^2 F^2}\left|\iint A(u,v)e^{i\frac{2\pi}{\lambda}W(u,v,t)}e^{-i\frac{2x}{\lambda F}(ux+vy)}dudv\right|^2 \quad (2)$$

Where λ is the wavelength, F is the system focal length. A(u,v) represents the system aperture function. A(u,v) is unity where light passes through the system and zero outside the aperture. $W(u,v,t)$ represents the wavefront leaving the system exit pupil. The integration is taken over the area of the aperture function. Note that the time dependence occurs only in the wavefront. Thus the wavefront is usually divided up into 2 terms a deterministic term and a stochastic components representing atmospheric fluctuations, line of sight jitter and scatter.

The deterministic term generally represents optical aberrations due to design and manufacture residuals. If one looks at the mathematical form of equation (2) it is seen that it is a 2D Fourier transform of the complex function thus making evaluation of equation (2) relatively fast. It can also be shown that modulus of the Fourier transform of equation (2), commonly referred to as the optical Modulation Transfer Function (MTF) is a monotonically decaying function with a cutoff frequency and thus is a low pass filter Thus any incoherent imaging system low pass filters the scene data. Also any value of wavefront other than zero tends to lower the system response [1]

If we make the assumption that the source changes much slower than a detector integration period then the objects time dependence can be ignored. If we also make the assumption that the sources spectral is much broader than the passband of the instrument then we can neglect the convolution at each wavelength separately. *Also* if we fold in the effects of the finite spectral response either due to the detector or spectral filters or both and also fold in the finite area of a pixel then we arrive at the spectrally weighted PSF which I'll denote by ***PS<***

$$PSF_s(x,y) = \int PSF(x,y;\lambda)S(\lambda)d\lambda \quad (3)$$

where $S(\lambda)$ represents the instrument passband. If we fold in the effects of the finite pixel area and the finite sampling then we arrive at the Point Response Function (PRF):

$$PRF(x,y) = \{PSF_s(x,y) ** Pix(x,y,dx,dy)\} \sum_{j,k} \delta(x - j\Delta x)\delta(y - k\Delta y) \quad (4)$$

Where $Pix(x,y,dx,dy)$ represents a pixel of dimension dx,dy at position x,y and the **sum** over delta functions represents the sampling function spaced on gridpoints separated by $\Delta x, \Delta y$ Note that **this** model also handles the cases of a sparse sampling with respect to the pixel size and also the case of Time Delay and Integrate (TDI) since in the general case we don't require that $dx \neq \Delta x$ and $dy \neq \Delta y$.

Equation **(4)** contains our generic model of the combined imaging, detection and sampling process and represents the response of the entire system to an input unresolved source.

Figure 1 shows a graphcal representation of a simulated wavefront and it's associated PSF. The leftmost figure in Figure 1 shows **an** optical wavefront leaving the system exit pupil and propagating towards the focal plane. The edge of circular function represents A(u,v) i.e. the aperture function, it is zero outside the circle and **unity** inside. The structure within the bounded circle represents the wavefront function W(u,v), it is dark where it is less than zero and bright where it **is** greater If we were looking through a turbulent medium **this** function would be changing stochastically with time. The middle figure of Figure 1 shows the evaluation of equation (2) and is the PSF in the focal plane, not the strong asymmetries due to the asymmetries in the wavefront. The rightmost figure of Figure 1 **is** the same PSF but with a logarithmic color stretch to show the finer details. A detector pixel size is shown in the upper right of Figure 1 Thus if we convolved **this** detector area with the PSF we would amve at the **PRF,** i.e. equation (4).

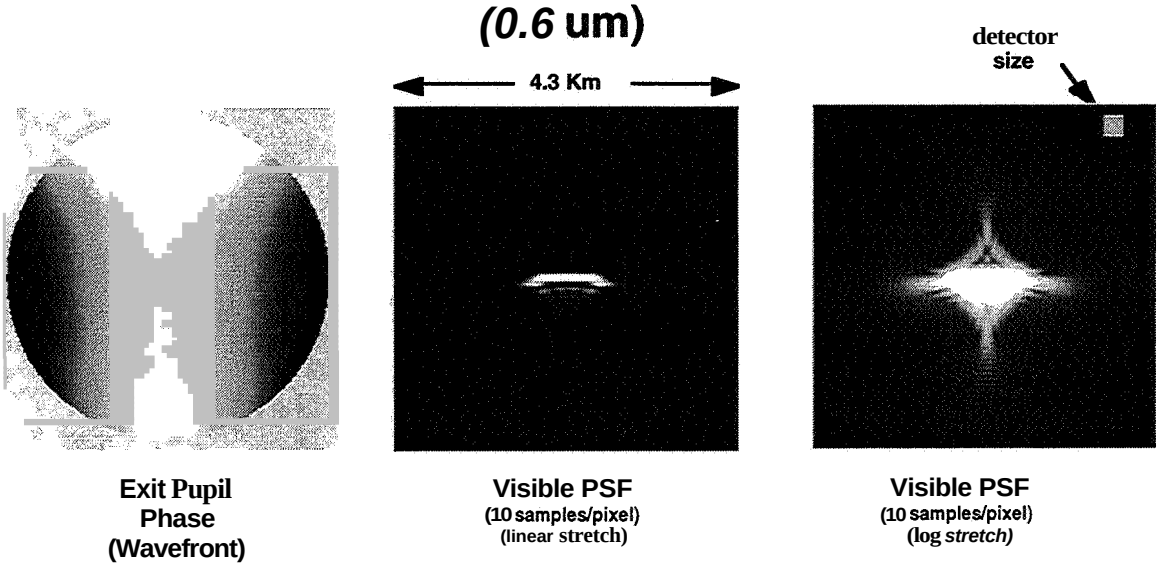


Figure 1
Wavefront and
Optical PSF

(2.0) Sampling and Aliasing

If we assume that each pixel element is square then $Pix(x,y,dx,dy)$ can be represented by a RECT function and we can adopt 1D notation without loss of generality The RECT function as given here is given by:

$$RECT(x/dx) = \begin{cases} 1.0 \text{ if } |x| \le dx/2 \\ 0.0 \quad \text{otherwise} \end{cases} \tag{5}$$

If we Fourier transform equation **(4)** into the frequency domain, to yield the system Modulation Transfer Function, some of the more standard effects become obvious:

$$MTF_s(f) = \{MTF_o(f) * SINC(f * dx)\} ** \sum_k \delta(f - k/\Delta x) \tag{6}$$

Where $MTF_o(f)$ is the optical MTF and $SINC(f * dx) \equiv \dfrac{\sin(\pi f dx)}{\pi f dx}$ Thus we get the optical MTF multiplied by a **SINC** function and replicated at equal intervals of $df = 1/\Delta x$ in frequency space. The optical MTF has a hard cutoff at DIU, and the first zero of the SINC function occurs at $f = 1/dx$. If for an example we assume an aperture size of D=30 cm,

a ground pixel size of p=30 meters square and wavelength of λ=4.0 microns and an orbital height of h=900 km then Figure 2 shows a plot of equation *(6)* for the this case with the pixel size equals the sample spacing, i.e. this is the most common case and represents a CCD array type detector with a fill factor of 100%. Figure *2* shows the optical MTF and the optical MTF multiplied by the pixel MTF. The units are cycles per pixel on the ground where a pixel is 30 meters. Note that from figure 2 that the first zero of **SINC** function in equation *(6)* occurs at 1 cycle/pixel but that the optical cutoff occurs at *2.5* cycles/pixel thus this system has the potential for aliasing if the scene data contains spatial frequencies at 0 S cycles/pixel or higher That is Nyquist's theorem requires that **we** sample at twice the highest frequency In general equation *(6)* should be starting point for any type of resampling analysis. In the next section we will discuss how to apply it and some of the effect due to resampling including how to minimize aliasing when resampling onto a finer grid.

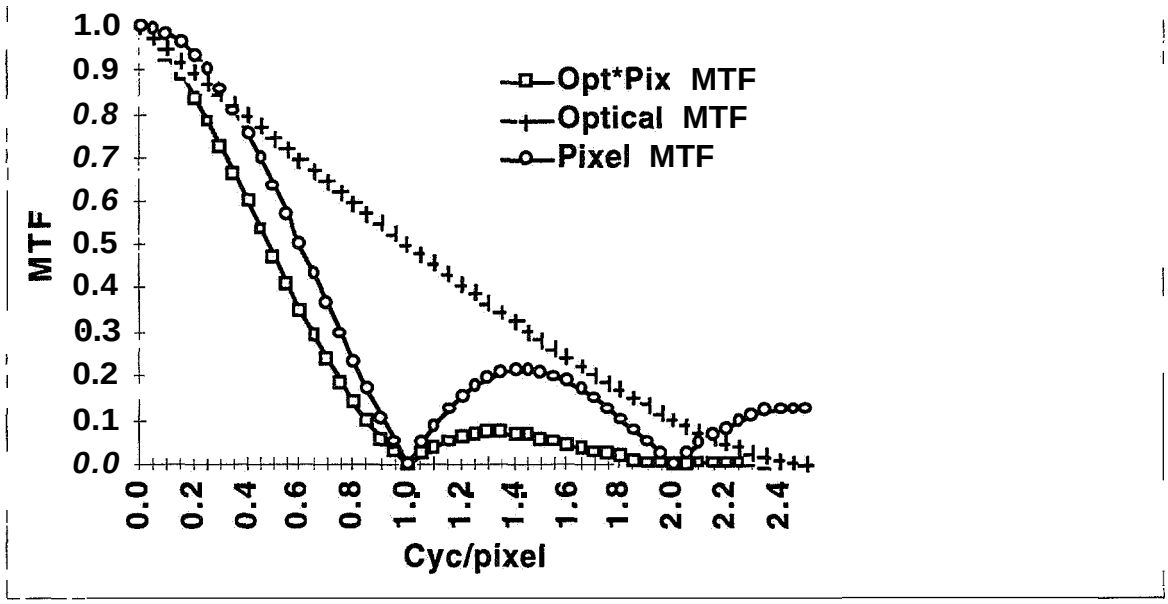


Figure 2
Optical and Pixel MTFs

(3.0) Bandlimited Interpolation

One of the main results to note from equation *(6)* is that all scenes that have been imaged by an imaging optical system are bandlimited. Thus a hard cutoff frequency exists and is known from the systems design. Resampling methods which are classified as interpolation can be recast as a discrete convolution of the data with an interpolation kernel. The ideal interpolation kernel for a band-limited function is the **SINC** function, see for example [2][3][4]. All other nearest neighbor and bi-linear resampling techniques are heuristic, cubic splines is a h t e d approximation of the theorectically optimal SINC resampling function. This is well known and is generally known as

the Whittaker-Shannon sampling theorem [1]. The **SINC** function can be used in practice since it requires an infinitely large pixel neighborhood. First discussed below is the general derivation of the Whittaker-Shannon sampling theorem. Then follows a derivation of a **SINC** like interpolation kernel with a correction term, i.e. a windowing function, to minimize a specific type of aliasing known as "spatial-spectral leakage" [4]. Other common forms of resampling will not be discussed here since there is already extensive literature available [2][3][4][5].

(3.1) Whittaker-Shannon Sampling and SINC Interpolation

Figure 3 shows a graphical representation of the idea behind **SINC** interpolation of bandlimited signals. With reference to Figure 3, given a bandlimited signal, S(x), then let it's Fourier transform be represented by the compactly supported function, $\tilde{S}(f)$ Let the discretely sampled version of S(x) be represented by:

$$S_j = S(j\Delta x)\sum \delta(x - j\Delta x) \qquad (7)$$

then it's Fourier transform would be given by:

$$\tilde{S}(f) ** \sum \delta(f - k/\Delta x) \qquad (8)$$

where "**" denotes convolution. Note that this replicates the spectrum at equal increments of $\Delta f = 1/\Delta x$ If the continuous signal spectrum is non-zero only over the interval from $\left[\frac{-1}{2\Delta x}, \frac{+1}{2\Delta x}\right]$ then, in principle, an ideal blocking filter of the form:

$$B(f) = RECT(f\Delta x) = \begin{cases} 1.0 & if\ |f| \le 1/2\Delta x \\ 0.0 & \text{otherwise} \end{cases} \qquad (9)$$

can be applied to "block" all the replicated spectrums leaving only the center spectrum. Thus transforming back into the spatial domain yields the following "idealized" method for interpolation:

$$S(x) = \sum_j S(j\Delta x) SINC\{(x - j\Delta x)/\Delta x\} \qquad (10)$$

Equation (10) is the mathematical recipe to exactly interpolate a bandlimited signal provided that we sample at twice the bandlimited cutoff.

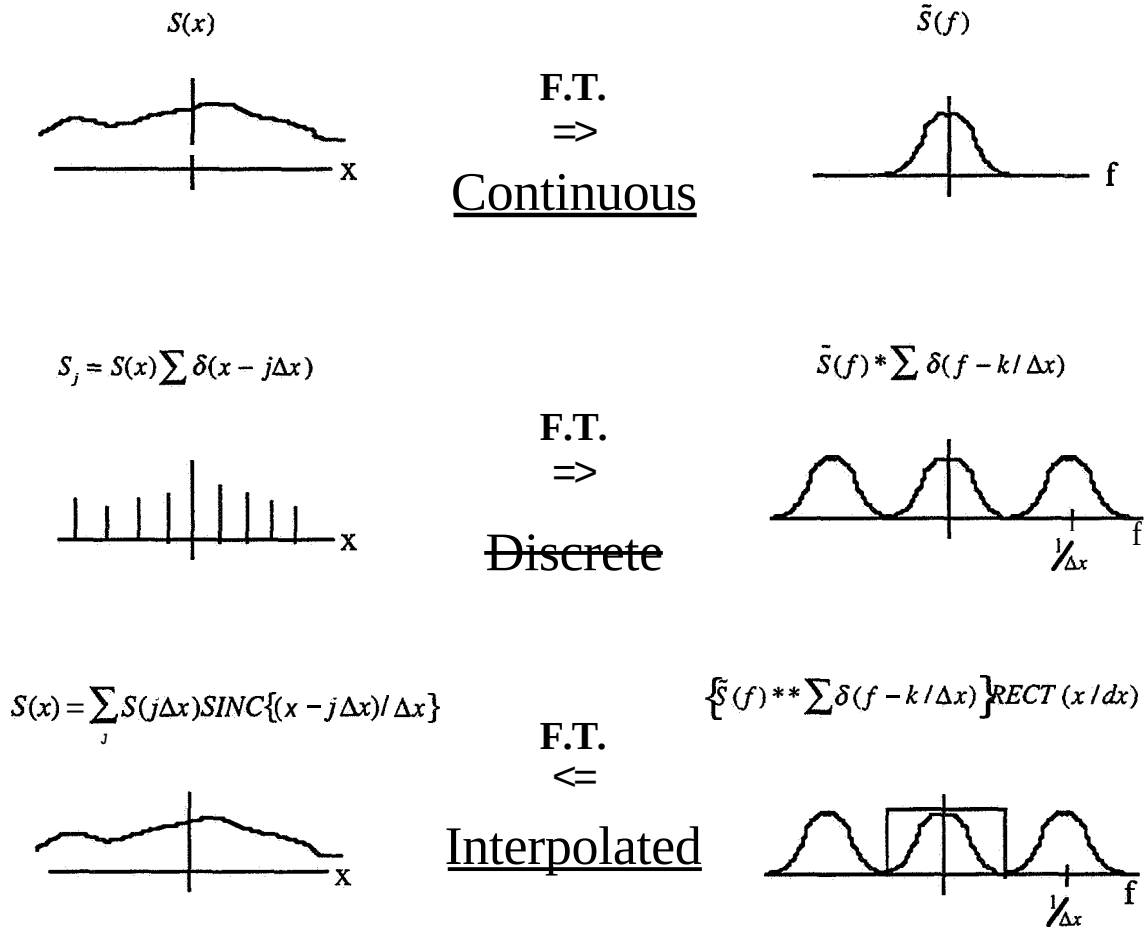


Figure 3
SINC Interpolation and Aliasing

(3.2) Hanning Windowed Sinc Interpolator

In practice equation (10) is generally not used since the index on the summation must extend from **[−∞,+∞].** Any truncation of the **SINC** interpolation kernel leads to spatial-spectral leakage. This problem can be **partially** overcome by the use of a windowing function discussed in this section. Figure **4** shows this effect in more graphic detail. If we truncate the **SINC** kernel onto a **3x3** grid and Fourier transform it back to the spectral domain then we get the-3 lobed function on the far left of Figure **4.** Note that there is also sideband leakage. This effect can be minimized via the used of a Hanning windowing function [5]. The Hanning window has fast sidelobe reduction in the spectral domain but not the fastest rise/fall time. This tends to soften the edges of the blocking filter The Hanning window modified construction kernel becomes:

$$K(x) = \frac{1}{2} SINC(x/\Delta x)\, RECT(x/D)\left\{1 - \cos\left(\frac{2\pi}{D}\left(x + \frac{D}{2}\right)\right)\right\} \qquad (11)$$

Where D is the size of convolution window in pixel units. Figure 4 is a plot of the truncated SINC and the Hanning windowed truncated SINC function in the spectral domain for D=3,5,7 and 10 pixel kernel sizes. In each case the filter is centered about f=0.0 cycles/pixel and extends to +/- 1.0 cycle/pixel. Overlayed on each curve are the non-windowed and windowed filters. The non-windowed filters contain ringing at both the top and in the wings. The Hanning windowed filters have much flatter tops and more rapid attenuation of the sidelobes, however, the rise/fall times are slower for the windowed filters. It was found that this approach to resampling works quite well in practice, however, it is slower than Bi-Linear interpolation and Bi-Cubic Splines, however, in general it gives better results since it more rigorously preserves spatial-spectral fidelity up to a cutoff frequency. Note that also a maximum likelihood weighting scheme can be employed in the presence of non-stationary noise. This work is still on-going.

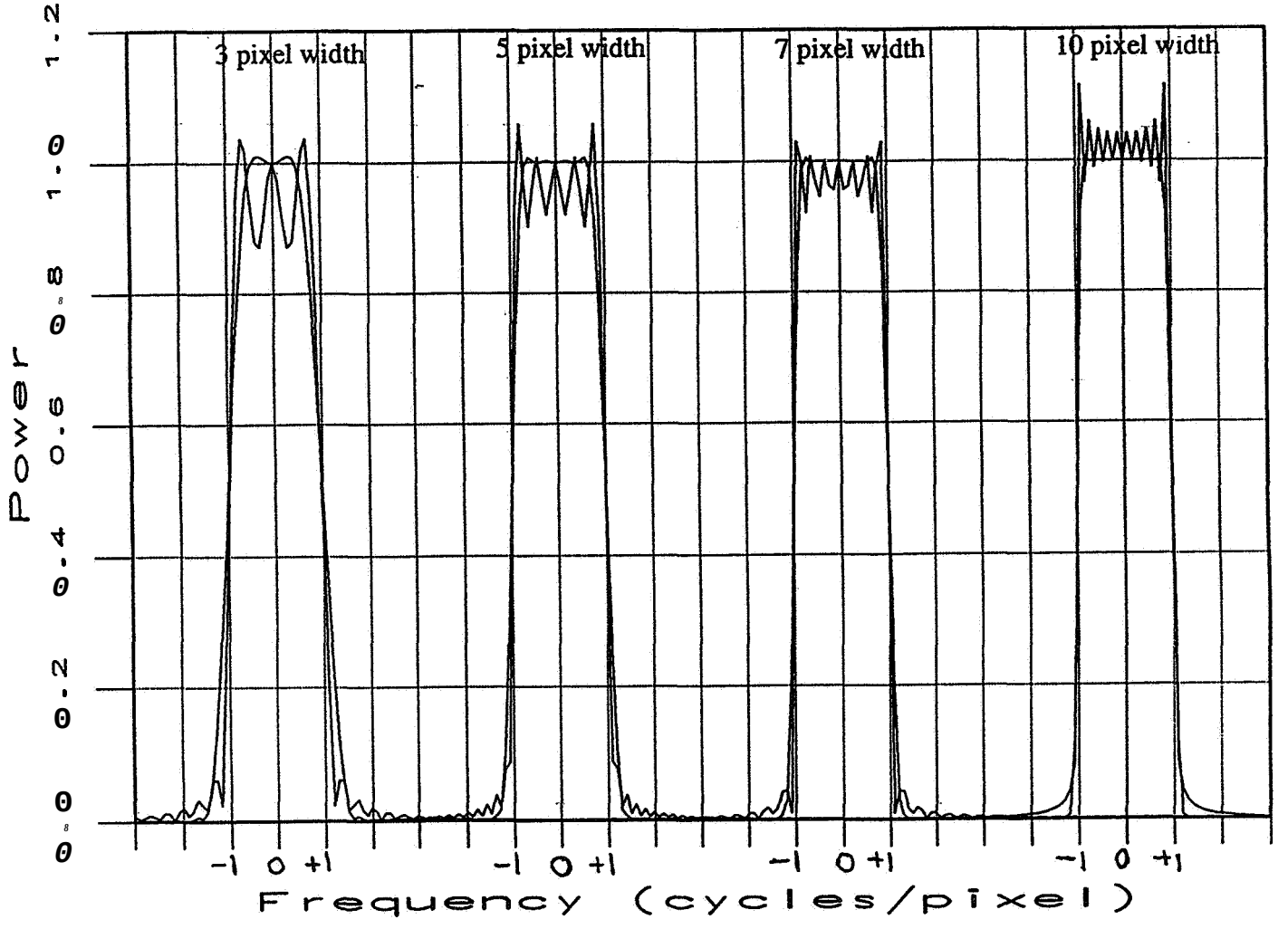


Figure 4
Hanning Windowed Blocking Filters

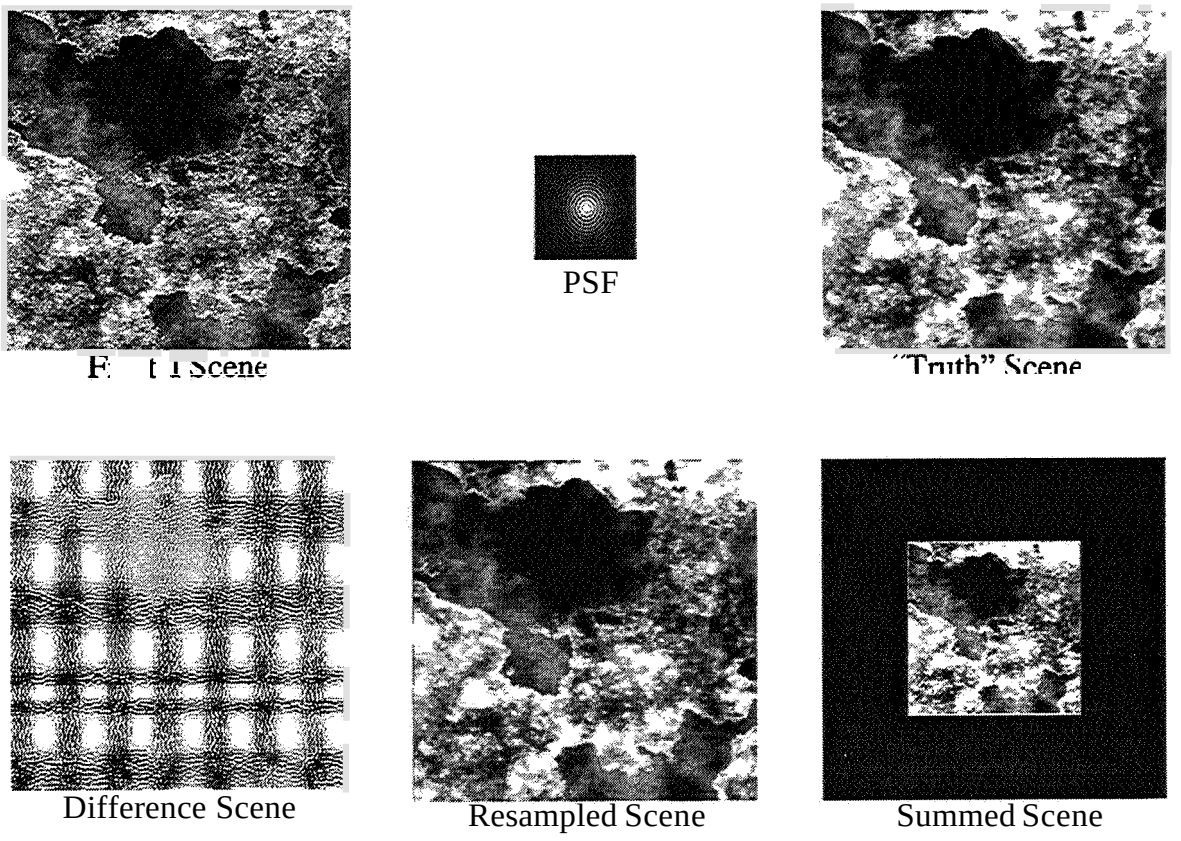


Figure 5
Scene Simulations

(4.0) Simulations

As a demonstration of the method a computer simulation was conducted. **A** fractal scene was "grown" simulating a downlooking earth sensor from low earth orbit. The scene contains land, coastlines and clouds. This scene is referred to as the "raw" scene. The raw scene is sampled on a 512 x 512 grid. The scene is convolved with an optical PSF such that the bandlimit is at 0.5 cycles/pixel and convolved with the detector response. This scene **is** referred to as the **"truth"** scene. The **"truth"** scene **is** subsequently subsampled to the final focal plane sampling and is referred to as the "image". The image represents the **data** as "seen" through the end to end sensor model, This image is then passed through the Hanning windowed sinc interpolator algorithm and resampled to twice the resolution **as** that of the focal plane scene and compared against the "truth" scene, hence the name truth scene. Note that the resampled data is not compared to the raw scene since this has yet to be low passed filtered by the sensor system. However methods are available whch can remove the sensor signature, see, for example, [6][7].

Figure 5 shows a mosaic of this simulation. The upper left is the raw fractal scene with clouds and coastlines and landmassess. The upper middle shows a diffraction limited PSF such that the diffraction limited spatial frequency cutoff **is** at 0.25 cycles/pixel. The upper right is the **PSF** convolved with the fractal scene and represents the truth scene, i.e. this is the best we could hope for without performing deconvolution of the data. The lower right is the truth scene summed down to the detector

resolution. The lower middle is resampled scene. Hence input to the algorithm is the scene on the lower right and the output is the lower mddle. The lower left scene is the difference between the Truth scene and the resampled scene. All scenes are on a linear greyscale except for the difference scene this has been histogram equalized to accentuate the low level structure.

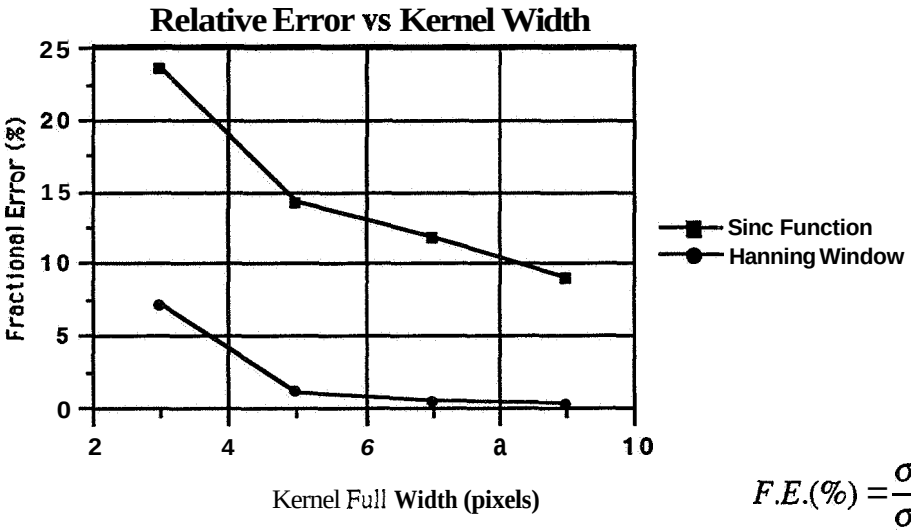
Figure 6 shows the fractional error as a function of a interpolation kernel width for both the Hanning windowed case and for the a truncated case. Fractional error is defined here by:

$$F.E.(\%) \equiv \frac{\sigma_d}{\sigma_T} * 100\% \qquad (12)$$

where σ_T is the standard deviation of the truth image, and σ_d is the standard deviation of the difference image. The kernel width is the full width of the interpolation kernel in units of the summed down image. Note that the error decays as the kernel width grows, however, for the Hanning windowed the error is always much less, indeed a 3x3 Hanning windowed interpolation kernel is still better than a 9x9 truncated sinc interpolator. Note that for completeness this method needs to be quantitatively compared to other standard resampling methods such as bi-linear and bi-cubic splines.

Figure 7 shows some preliminary computer timings versus interpolation kernel size for the fractal scene data. The runs were all performed on a Silicon Graphics 175Mhz Octane. The data was regridded from a 256x256 to a 512x512 grid and all operations were done in 4 byte floating point. The Hanning windowed sinc was slower over the entire range of kernel size as would be expected since more calculations are required. Both show an approximately quadratic dependence with kernel size.

Figure 8 shows some of the preliminary results for the power spectral densities **(PSD).** The results are plotted in log-log space. Actually what is being shown is a periodgram of the respective data sets since a **PSD** requires an ensemble average. Shown, with reference to the legend on Figure 8, is the raw scene **PSD,** the truth scene **PSD,** the resampled scene **PSD** and the modulation transfer function. The raw scene follows an approximate power law with respect to spatial frequency This is as would be expected for a fractal scene. The truth scene **PSD** is essentially the raw scene **PSD** multiplied by the MTF, and this attenuation can easily be scene on Figure 8.



$$F.E.(\%) = \frac{\sigma_d}{\sigma_T} * 100\%$$

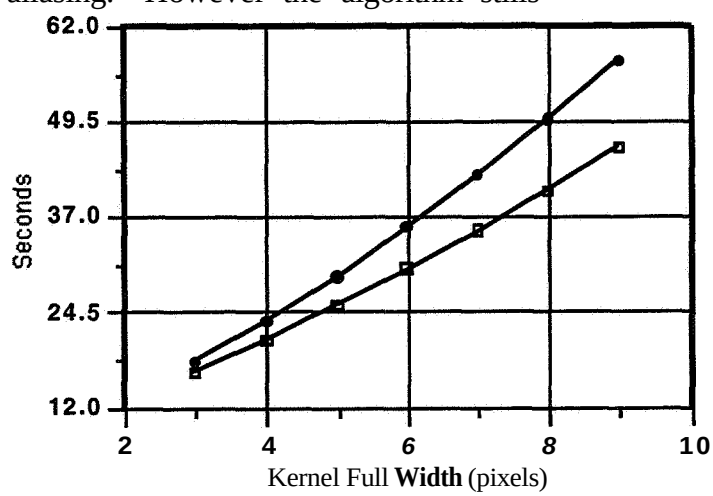
$\sigma_d \equiv S.Dev$ of Truth - Resampled

$\sigma_T \equiv$ S Dev of Truth Image

Figure 6
Fractional Error

(5.0) Conclusions

In conclusion **a** preliminary bandlimited resampling algorithm has been designed, developed and **is** currently **in** the testing and optimization stages. The algorithms **has** been tested to a limited extent on synthetic scene **data** for the effects of aliasing. However the algorithm stills needs to .be tested with respect to jitter, noise, registration effects and "rubber-sheeting" These tests are currently in the progress. Also the method needs to be quantitatively compared with other standard resampling methods such as bi-linear and bi-cubic splines to determine if there **is** still some deficiencies.



- **512 x 512 Grid**
- **175 Mhz SGI Octane**
- **4 Byte floating point**

Figure 7
Computational Timings

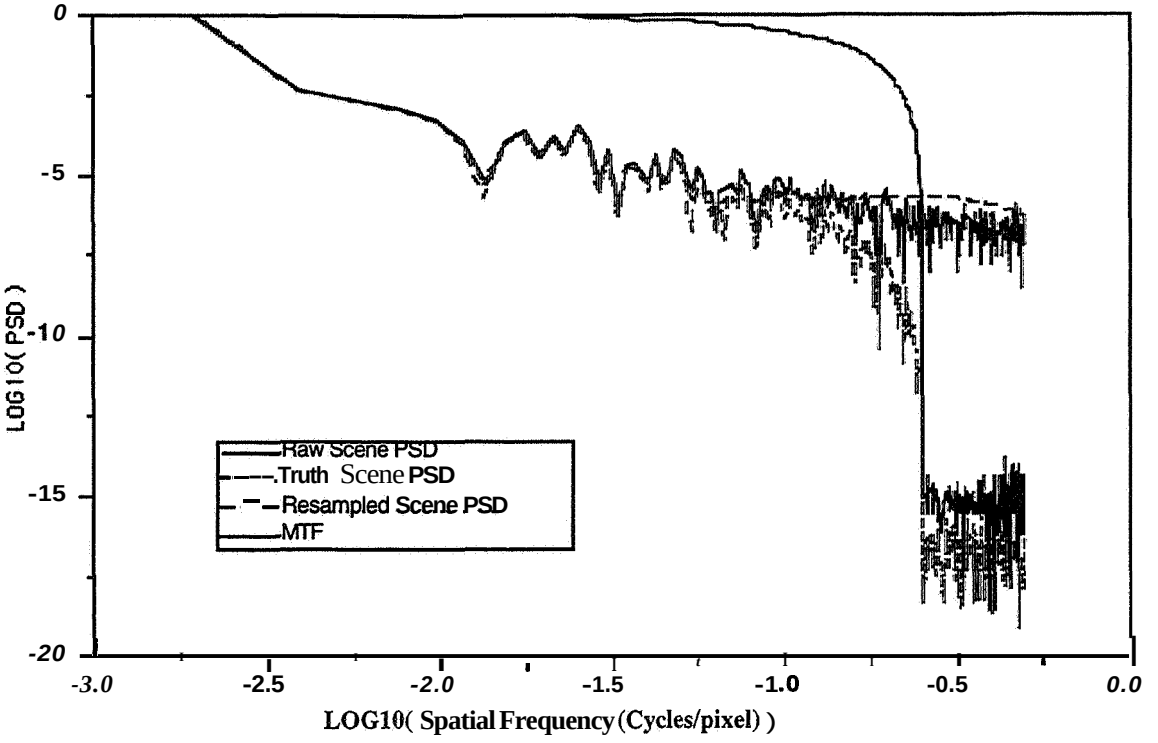


Figure 8
Power Spectral Densities

(6.0) References

[1] Goodman ,J W, " Introduction to Fourier Optics ", McGraw-Hill Inc, 1986

[2] Diksht, O and Roy,D.P *"An* Empirical Investigation of Image Resampling Effects Upon the Spectral and Textural Supervised Classification of a High Spatial Resolution Multispectral Image", Photogrammetric Engineering & Remote Sensing, Vol 62, No 9, September 1996, pp. 1085-1092

[3] Shlien, S, "Geometric Correction, Registration and Resampling of Landsat Imagery", Canadian Journal of Remote Sensing, Vol 5, No 1, pp. 74-89

[4] Lyon, R.G "Scanned Image Construction via Optimal Interpolation", Proceedings of **SPIE,** Vol. 2298, July 1994

[5] Press, W H., Flannery, B. P , Teukolsky, S. A., and Vetterling W T., "Numerical Recipes in C", Cambridge, 1988, pp. 443-444

[6] Lyon, R. G., Hollis, J. M., Dorband, J. E., "Image Restoration Using Maximum Entropy Method with Maximum Likelihood Constraints", Submitted to Applied Optics September 12, 1997

[7] Lyon, R. G., Hollis, J M, Dorband, J. E.,"A Maximum Entropy Method with A Priori Maximum Likelihood Constraints", ApJ, **478:** 658-662, April 1,1997.

Restoration and Reconstruction From Overlapping Images

Stephen E. Reichenbach*
Computer Science and Engineering Department
University of Nebraska–Lincoln

Daniel J. Kaiser
Information Science and Technology Department
Doane College

Andrew L. Hanson
Jing Li
Computer Science and Engineering Department
University of Nebraska–Lincoln

Abstract

This paper describes a technique for restoring and reconstructing a scene from overlapping images. In situations where there are multiple, overlapping images of the same scene, it may be desirable to create a single image that most closely approximates the scene, based on all of the data in the available images. For example, successive swaths acquired by NASA's planned Moderate Imaging Spectrometer (MODIS) will overlap, particularly at wide scan angles, creating a severe visual artifact in the output image. Resampling the overlapping swaths to produce a more accurate image on a uniform grid requires restoration and reconstruction. The one-pass restoration and reconstruction technique developed in this paper yields mean-square optimal resampling, based on a comprehensive end-to-end system model that accounts for image overlap, and subject to user-defined and data-availability constraints on the spatial support of the filter.

1 Introduction

Resampling is required in many imaging applications and is particularly important in remote sensing to correct for geometric distortion and to register, rescale, or otherwise remap. Resampling requires reconstruction — the process of determining values at arbitrary spatial locations. Determining the best value requires restoration — the process of correcting for degradations that are introduced during the imaging process in order to obtain more accurate estimates of the scene radiance field. With multiple images, a resampling technique should use all available data in restoring and reconstructing the scene.

This paper develops a technique for multi-image fusion that, in one-pass through overlapping input images, restores and reconstructs the scene radiance field. The technique is effective because it maximizes fidelity based on a comprehensive end-to-end system model that accounts for scene statistics, acquisition blurring, sampling, and noise. The technique is efficient because the filter is derived subject to user-defined and data-availability constraints on spatial support for spatial-domain processing.

Section 2 presents the end-to-end imaging system model that is the basis for the derivation. Section 3 extends the model to account for overlapping images. Section 4 describes the problem of overlapping swaths in MODIS images. Section 5 defines the performance criteria for the derivation. Section 6 presents the derivation of the optimal unconstrained filter for overlapping images. Section 7 imposes user-defined and spatial-availability constraints on the derivation. Section 8 describes performance evaluation of the optimal and constrained filters. Section 9 looks briefly at the more general problem of multi-iie fusion.

2 Digital Imaging System Model

Using the end-to-end, imaging system model illustrated in Figure 1, the digital image p is

$$p(x_n) = \int_{-\infty}^{\infty} h(x_n - x)\, s(x)\, dx + e(x_n) \qquad (1)$$

where the continuous scene radiance field s is convolved with the pre-sampling acquisition point spread function (PSF) h (e.g., characterizing blurring from the optics and the spatial integration of the photodetector), sampled, and degraded by noise e (e.g., shot noise, circuit noise, and quantization error). In reality, noise is caused by spatially continuous processes, but one can define a discrete noise process that has statistically identical effects on the digital image. This and subsequent equations are written in one dimension for notational convenience, but generalize directly to multiple dimensions. As a practical matter, only a finite portion of the image p is available. Spatial coordinates x are normalized to the field-of-view and pixels have integer indices n that determine spatial location x_n within the field-of-view.

The filter f restores and reconstructs a continuous image that then can be resampled. The resulting (continuous)

*This research was supported by NASA through the Landsat 7 Science Office at the Goddard Space Flight Center (Grant NAGS-3126) and the Human Resources and Education Office at NASA Headquarters (Grant NCC5-169). Special thanks to John Barker, Darrel Williams, and other members of the Landsat 7 science team. Professor Reichenbach can be contacted by phone at (402) 472-3200, by email at reich@unl.edu. and on the WWW at http://cse.unl.edu/~reich.

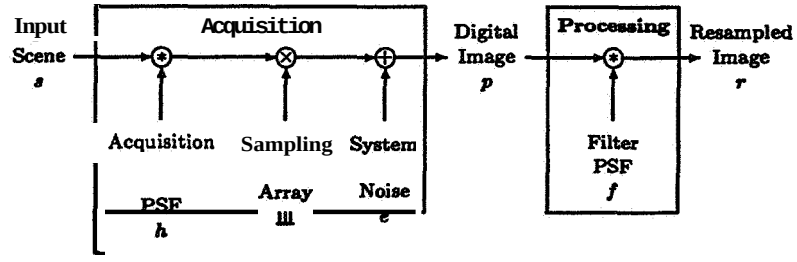


Figure 1: Digital Imaging System Model.

image r is

$$r(x) = \sum_{n=-\infty}^{\infty} f(x - x_n)\, p(x_n). \quad (2)$$

To address the limited field-of-view and to facilitate Fourier analysis, it is common to assume the scene and hence the image are periodic with period equal to the field-of-view and to constrain the filter support to the extent of the image (or d e r). The effect of this assumption in the filter derivation is negligible because the image field-of-view is large relative to the scene mean-spatial-detail and system PSF.

3 Overlapping Images

In overlapping images, the pixels may not be uniformly spaced. consider two overlapping sets of uniformly spaced pixels, such as would be produced by successive swaths of MODIS. Referencing the coordinate system to one of the sample sets, the pixel locations are

$$x_n = \begin{cases} \Delta_x \frac{n}{2} & \text{if } n \bmod 2 = 0 \\ \Delta_x \frac{n}{2} + \epsilon & \text{if } n \bmod 2 = 1 \end{cases} \quad (3)$$

where Δ_x is the inter-pixel spacing in both sample sets and the even indices reference locations in one sample set and the odd indices reference locations in the other sample set.

Aliasing introduced by sampling can be analyzed more directiy in the Fourier, spatial-frequency domain. The Fourier transform or spatial-frequency spectrum @ of a single (periodic) image with uniform sampling is

$$\hat{p}(v_n) = \sum_{k=-\infty}^{\infty} \hat{h}(v_n - k)\, \hat{s}(v_n - k) + \hat{e}(v_n) \quad (4)$$

where spatial frequencies v are in cycles per field-of-view, $\hat{h}$ is the acquisition transfer function, $\hat{s}$ is the spatial-frequency spectrum of the scene, and $\hat{e}$ is the spatial-frequency spectrum of the noise. For a periodic scene and image, the spectra coefficients v_n are discrete, uniformly spaced, and have integer indices a. Sampling causes the folding of the components of the spatial-frequency spectrum (hence, the summation with index k).

With two overlapping images, the image transform $\hat{p}$ is

$$\hat{p}(v_n) = \sum_{k=-\infty}^{\infty} \hat{h}(v_n - k)\hat{s}(v_n - \mathrm{k})\,(1 + \exp(-i2\pi\epsilon k)) + \hat{e}_0(v_n) + \hat{e}_\epsilon(v_n) \quad (5)$$

where E is the offset between the sampling sets. The offset (or shift) ϵ in the second sample set, does not change the stochastic nature of the noise process. Unlike the image spectrum in Equation 4, the image spectrum in Equation 5 is not necessarily periodic.

The frequency-domain equation for filtering (corresponding to Equation 2) is

$$\hat{r}(v_n) = \hat{f}(v_n)\hat{p}(v_n). \quad (6)$$

4 Overlapping Swaths in MODIS Images

The along-track dimension of the MODIS scan swath is 10km at nadir (0 degrees) but reaches 20km at the maximum scan angle (55 degrees) at the end of each swath. Because of its geometric shape, narrow at the center and broad at the left and right, this is referred to as the *bow-tie effect*[1]. The along-track ground speed is 10km per swath, so successive scans overlap. At the maximum scan angles, the overlap is as much as 50% with the swath above and 50% with the *swath* below.

To illustrate the bow-tie effect, we have constructed a simulation of the MODIS system. The simulation uses a Landsat TM image (Band 5) with approximately 30m resolution, a portion of which is pictured in F i e 2, as the scene. Following the imaging system model illustrated in Figure 1, the Landsat TM scene is blurred, sampled, and corrupted by noise. In this simulation, the ground sampling interval is 250m at nadir with 40 pixel swaths.

In order to facilitate comparison of the effects at different scan angles and because the ground projection of the MODIS scan is much larger than the Landsat TM scene, the Landsat TM image is tiled edge-to-edge to produce a scene as wide as the MODIS scan half-width (i.e., nadir to one end of the scan). The Landsat TM image is 4096 x 4096. The size of the MODIS half-width, 1165km, is approximately 10 tiles x 4096 pixels/tile x 28.5 meters/pixel.

The simulated MODIS image produced by this process is illustrated in Figure 3. The left edge of the image is at nadir and the right edge is at the end of the swath. The left-*most* tile, with its center at a scan-angle of 0.0826 radians, is enlarged in Figure 4. A tile from the middle of the half-scan, with its center at a scan-angle of 0.5170 radians, is enlarged in Figure 5. The right-most tile, with its center at a scan-angle of 0.9428 radians, is enlarged in Figure 6. At larger scan angles, the ground sampling distance is larger and so fewer MODIS pixels are sampled in the same ground area. The bow-tie effect is a dominant artifact in the portion of the swath pictured in F i e 6. In this part of the image, the delineation between swaths is clear and overlap makes it difficult to discern spatial structures in the scene.

Figure 2 Landsat TM Scene

Resampling the image on a uniform rectangular raster eliminates the troublesome bow-tie effect. One of the simplest algorithms for resampling is to take the value of the nearest pixel to the resampling point. This reconstruction method is called *nearest-neighbor interpolation.* Figure 7 illustrates the image of Figure 3 resampled using nearest-neighbor interpolation. Figures 8-10 are the resampled tiles with centers at scan-angles 0.0826, 0.5170, and 0.9428 respectively.

Nearest-neighbor interpolation is simple and easily computed, but produces images with visible blockiness and low fidelity. These artifacts of nearest-neighbor interpolation are clearest in Figure 10. Nearest-neighbor interpolation reconstructs but does not restore. More sophisticated reconstruction techniques, such as linear interpolation or cubic convolution, yield better results. Techniques that restore and reconstruct[2, 3] can yield much better fidelity.

5 Digital Imaging System Performance

System performance is measured by how closely the output image r matches the scene s. Linfoot[4] used the *expected* mean-square *error* of an *imaging* system (with stochastic scene and noise)

$$S^2 = E\left\{\int_{-\infty}^{\infty} |s(x) - r(x)|^2 \, dx\right\} = E\left\{\int_{-\infty}^{\infty} |\hat{s}(v) - \hat{r}(v)|^2 \, dv\right\} \quad (7)$$

to define *image* fidelity as

$$F = 1 - \frac{S^2}{\sigma_s^2} \quad (8)$$

where σ_s^2 is the expected (ensemble average) variance of the scene radiance field

$$\sigma_s^2 = E\left\{\int_{-\infty}^{\infty} |s(x)|^2 \, dx\right\} = E\left\{\int_{-\infty}^{\infty} |\hat{s}(v)|^2 \, dv\right\}. \quad (9)$$

For notational simplicity, equations in this paper assume the scene radiance field is a zero-mean process; in practice, the mean can be accounted for during filtering. Fidelity is bounded by 1, with $F = 1$ if and only if the output image is identical to the scene radiance field. Mean-square-error metrics such as fidelity facilitate mathematical analyses, but do not correspond directly to subjective visual quality. However, a more definitive objective measure of image quality has proven elusive.

6 Optimal Filter Derivation

The filter design criterion is to minimize mean-square error (or equivalently to maximize fidelity) of the system. The mean-square error after filter-kg (Equation 7) can be written as

$$S^2 = \sigma_s^2 - 2\sigma_{s,r} + \sigma_r^2. \quad (10)$$

where σ_r^2 is the expected variance of the output *image* and $\sigma_{s,r}$ is the expected covariance of the scene radiance field and image. These terms can be expressed as

$$\sigma_s^2 = \sum_{n=-\infty}^{m} \hat{\Phi}_s(v_n) \quad (11)$$

$$\sigma_{s,r} = \sum_{n=-\infty}^{\infty} \hat{\Phi}_{s,p}(v_n) \, \hat{f}^*(v_n) \quad (12)$$

$$\sigma_r^2 = \sum_{n=-\infty}^{m} \hat{\Phi}_p(v_n) \, |\hat{f}(v_n)|^2 \quad (13)$$

where $\hat{\Phi}_s$ is the power spectrum of the scene, $\hat{\Phi}_{s,p}$ is the cross-power spectrum of the scene and image, $\hat{\Phi}_p$ is the power spectrum of the image, and the '*' superscript denotes complex conjugation.

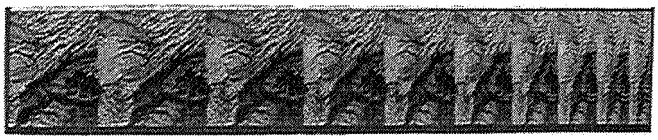
If the noise is random and signal-independent and the sidebands of the scene spectrum that alias to the same frequency are uncorrelated, then for the overlapping images of Equation 5

$$\hat{\Phi}_s(v_n) = E\{|\hat{s}(v_n)|^2\} \quad (14)$$

$$\hat{\Phi}_{s,p}(v_n) = E\{\hat{s}(v_n)\hat{p}^*(v_n)\} = 2\hat{\Phi}_s(v_n)\hat{h}^*(v_n) \quad (15)$$

$$\hat{\Phi}_p(v_n) = E\{|\hat{p}(v_n)|^2\} = 2\sum_{k=-\infty}^{\infty} \hat{\Phi}_s(v_n - k)\,|\hat{h}(v_n - k)|^2 \times (1 + \cos(2\pi\epsilon k)) + 2\hat{\Phi}_e(v_n). \quad (16)$$

Because all spatial functions are real, the frequency-domain transforms are hermitian and the power spectra $\hat{\Phi}_s$, $\hat{\Phi}_p$, and $\hat{\Phi}_e$ are real, non-negative, and even.



F i e 3: Simulated MODIS Image (Half-Swath — Nadir to Swath Edge)

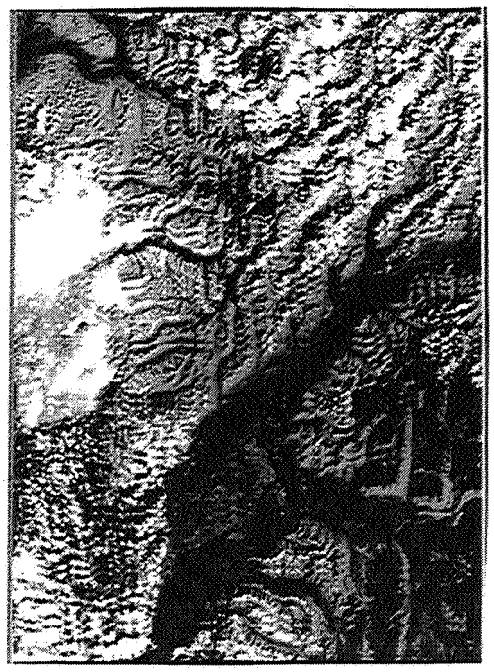


Figure 5: Simulated MODIS Image Tile **with** *Scan* **Angle 0.5170**

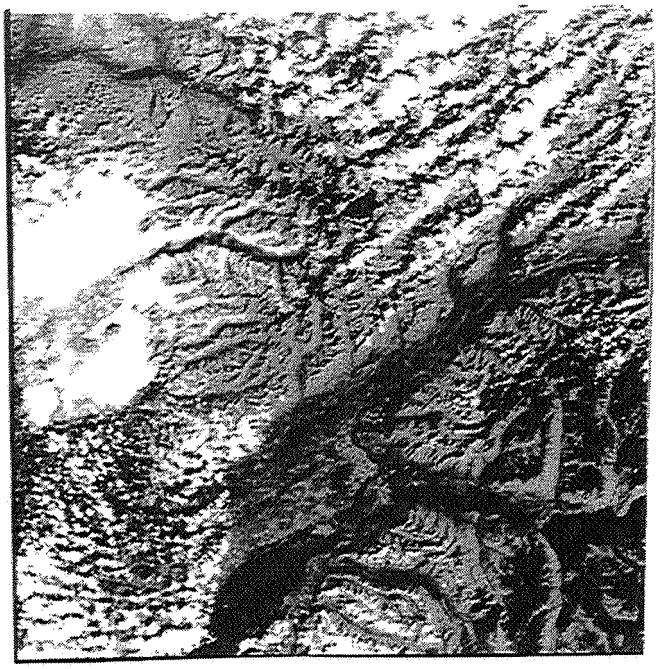


Figure 4 Simulated MODIS Image Tile with *Scan* **Angle 0.0826**

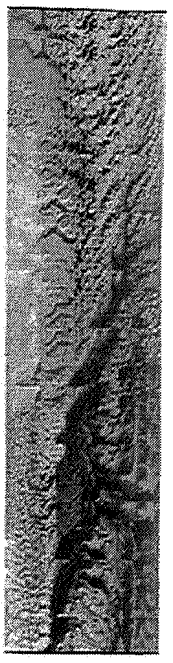


Figure 6 Simulated MODIS Image Tile **with** *Scan* **Angle 0.9428**

Figure 7: Resampled MODIS Image (Half-Swath — Nadir to Swath Edge)

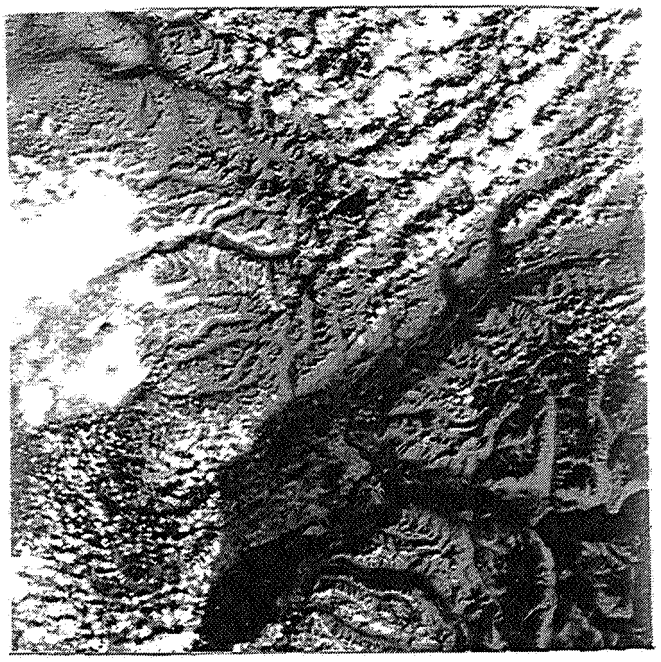


Figure 8: Resampled MODIS Image Tile with *Scan* Angle 0.0826

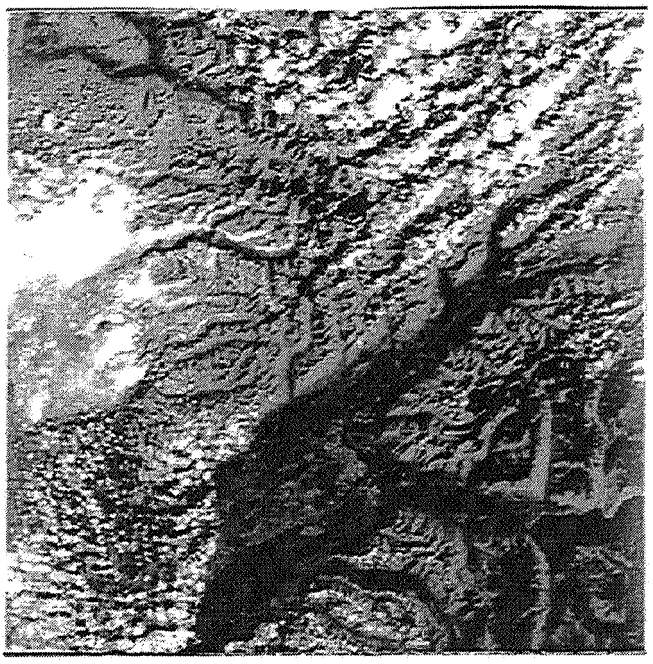


Figure 9: Resampled MODIS Image Tile with *Scan* Angle 0.5170

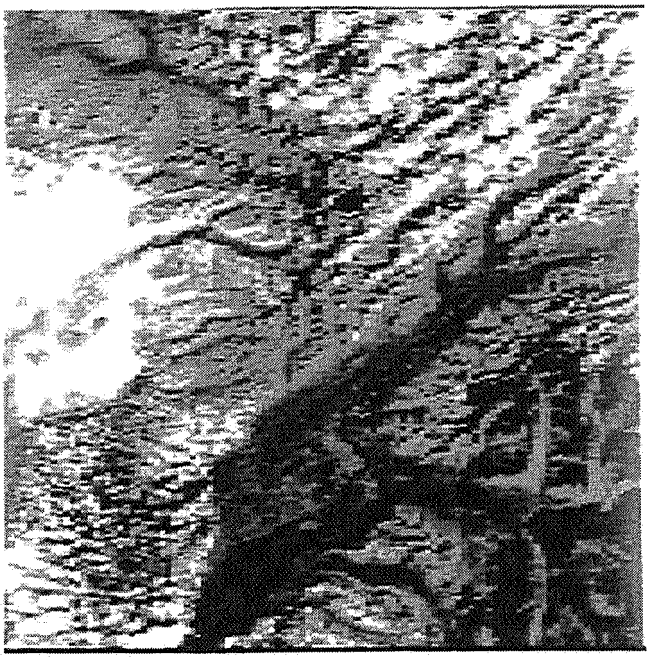


Figure 10 Resampled MODIS Image Tile with Scan Angle 0.9428

With no restrictions on the spatial support of the filter, the mean-square error is minimized when the filter transfer function is[5]

$$\hat{f}_w(\nu_n) = \frac{\hat{\Phi}_{s,p}(\nu_n)}{\hat{\Phi}_p(\nu_n)}. \quad (17)$$

We refer to this filter as the CDC Wiener filter because it is based on the continuous-discrete-continuous (CDC) model — acquisition converts from a continuous scene to a digital image and the filter converts from a digital image to a continuous result.

The CDC Wiener filter *cannot* be implemented practically via spatial convolution because it is continuous and its support is the *extent of* the field-of-view. Moreover, it has been assumed to this point that both sample sets are fully populated, but as with the overlapping swaths of the MODIS instrument, the overlap may occur in only part of the image. To produce a filter that is both more efficient and weights only available sample dues, we impose support constraints on the filter — limiting the filter to weight a subset of the image values during convolution — prior to the derivation.

7 Optimal Constrained Filter Derivation

The derivation of the optimal, constrained filter f_c is conditioned on constraints imposed on its spatial support. The support of the kernel is a xion-empty set of spatial locations, C, for which filter dues *can* be nonzero. Except for locations in the support set, the filter value is 0

$$f_c(x) = 0 \qquad \text{if } x \notin C. \quad (18)$$

The optimal, spatially-constrained kernel is derived by minimizing the mean-square *error* S^2 with respect to the elements in the support set. Minimization of S^2 with respect to the kernel elements requires that

$$\frac{\partial S^2}{\partial f_c(x)} = 0 \qquad x \in C. \quad (19)$$

The transfer function of the spatially-constrained kernel is

$$\hat{f}_c(\nu_n) = \sum_{x \in C} f_c(x) \exp(-i2\pi\nu_n x) \quad (20)$$

After substituting this expression for $\hat{f}$ into Equation 10 (with Equations 11-16), it *can* be shown[6] that minimization of S^2 with respect to the kernel elements requires

$$\sum_{x' \in C} \Phi_p(x - x') f_c(x') = \Phi_{s,p}(x) \qquad x \in C \quad (21)$$

where Φ_p is the auto-covariance of the image and $\Phi_{s,p}$ is the cross-covariance of the scene and image (again, presuming zero-mean processes). The number of unknowns in Equation 21 is equal to the number of elements in the support set of the kernel—there are $|C|$ equations in $|C|$ unknowns, solving for f_c yields the optimal constrained filter.

8 Filter Performance

Equation 10, with the expressions in Equations 11-16, *can* be used to compute the expected mean-square error from $\hat{\Phi}_s$, $\hat{\Phi}_e$, $\hat{h}$, and $\hat{f}$. For the CDC Wiener filter f_w in Equation 17,

$$S_w^2 = \sum_{n=-\infty}^{\infty} \hat{\Phi}_s[\nu_n] - \frac{|\hat{\Phi}_{s,p}[\nu_n]|^2}{\hat{\Phi}_p[\nu_n]}. \quad (22)$$

Therefore, the fidelity (Equation 8) for the CDC Wiener filter is

$$F_w = \frac{1}{\sigma_s^2} \sum_{n=-\infty}^{\infty} \frac{|\hat{\Phi}_{s,p}(\nu_n)|^2}{\hat{\Phi}_p(\nu_n)}. \quad (23)$$

No filter *can restore* with higher fidelity (smaller mean-square error) than F_w in Equation 23. However, for typical *imaging* systems, a small filter with a few centrally located elements *can* perform nearly as well [6]. For any filter f, including the constrained filter f_c, the expected mean square *error* is

$$S^2 = S_w^2 + \sum_{n=-\infty}^{\infty} \hat{\Phi}_p(\nu_n) |\hat{f}(\nu_n) - \hat{f}_w(\nu_n)|^2. \quad (24)$$

As *can* be seen in Equation 24, $S^2 \geq S_w^2$ and F_w is the upper bound *on* fidelity.

9 Conclusion

The restoration and reconstruction technique in this paper is described in terms of its applicabiity to overlapping swaths in MODIS images. The approach used in this technique has wider applicabity to multi-image fusion[7], the more general problem *of* combining images from multiple data sources to form a single image. Other multi-image fusion problems may be complicated by sources with different imaging geometries, spatial resolution. spectral response, time of acquisition, or other variables. With respect to these issues, the problem of resampling MODIS *images* on a uniform grid is a fairly straightforward problem, but provides a useful arena for developing, demonstrating, and evaluating various approaches.

References

[1] MODIS Level 1 Geolocation, Characterization and Calibration Algorithm Theoretical Basis Document. NASA TM-10594 MODIS Technical Report Series, Volume 2, Version 1, May 1994.

[2] Stephen E. Reichenbach, Daniel E. Koehler, and Dennis W. Strelow. Restoration and reconstruction *of* AVHRR *images. IEEE Transactions on Geoscience* and *Remote* Sensing, 33(4):997–1007, 1995.

[3] Stephen E. Reichenbach and Kevin Haake. Cubic convolution for one-pass restoration and resampling. In *International Symposium* on *Geoscience* and *Reomote Sensing*, volume III, pages 1597–1599. IEEE, 1996.

[4] E. H. Linfoot. Transmission factors and optical design. *Journal of* the *Optical Society of America*, 46(9):740–752, 1956.

[5] Friedrich *O.* Huck, Carl L. Fales, Nesim Halyo, Richard W. Samms, and *Kathryn* Stacy. Image gathering and processing: *Information* and fidelity. *Journal of the Optical* Society *of America A*, 2(10):1644–1666, 1985.

[6] Stephen E. Reichenbach and Stephen K. Park. Small convolution kernels *for* high-fidelity image restoration. *IEEE Transactions* on *Signal Processing*, 39(10):2263–2274, 1991.

[7] Robert A. Schowengerdt. *Remote Sensing: Models* and *Methods* for *Image* Processing. Academic Press, San Diego, CA, second edition, 1997.

Session III

Applications to Satellite Sensors

Automated Navigation Assessment for Earth Survey Sensors using Island Targets

FREDERICK S. PATT
ROBERT H. WOODWARD
General Sciences Corporation
Laurel, Maryland 20702

WATSON W GREGG
Global Change Data Center
NASA/Goddard Space Flight Center
Greenbelt, Maryland 20771

Abstract. An automated method has been developed for performing navigation assessment on satellite-based Earth sensor data The method utilizes islands as targets which can be readily located in the sensor data and identified with reference locations. The essential elements are an algonthm for classifymg the sensor data according to source, a reference catalog of island locations, and a robust pattern-matching algonthm for island identification. The algorithms were developed and tested for the Sea-vieunng Wide Field-of-view Sensor (SeaWiFS), an ocean color sensor This method will allow navigation error statistics to be automatically generated for large numbers of points, supporting analysis over large spatial and temporal ranges.

1. Introduction

Assessing navigation accuracy has been a challenge for nearly every Earth sensing satellite mission since the beginning of the space program. Extensive analysis efforts have been devoted to this problem, as evidenced by many publications in the open literature (e.g., Emery, Brown and Novak, 1989; Marsouin and Brunel, 1991). By and large, navigation assessment has relied on manual selection of ground control points based on overlaying coastlines in sensor image coordinates (Krasnopolsky and Breaker, 1994; Rosborough, Baldwin and Emery, 1994). However, the effort required to perform manual navigation quality control is considerable, and requires human resources which are better used for other purposes. Manual assessment is also extremely tedious and error-prone. For a global survey mission, these problems are exacerbated.

In developing a navigation assessment strategy for the Sea-viewing Wide Field-of-view Sensor (SeaWiFS), a global ocean color sensor launched in August 1997, we sought to develop a system to minimize the need for manual intervention. Automated methods based on autocorrelation techniques have been developed, but with limited success (O'Brien and Turner, 1992). One method based upon a catalog of features (islands, capes) over a local area has met with considerable success (Bordes, Brunel and Marsouin, 1992).

However, an automatic navigation assessment technique must meet rather stringent requirements Earth features to be used for computing nawgation errors must be distinguished in the data stream with a typical error of less than one pixel, and must be readily associated with a reference location whose coordinates are accurately known The processing must be efficient in its use of computer resources

We believe that islands represent a useful set of targets for this purpose. Islands can be readily located in sensor data as clusters of contiguous land pixels surrounded by water (assuming that land and water can be easily distinguished). The geometric centroid of an island defines a point which can be computed from sensor data and also from a map, simplifying both the identification process and the error computation. There are thousands of islands of a suitable size for SeaWiFS navigation (roughly 1 to 100 km in the largest dimension), well distnbuted in latitude and longitude.

To automate navigation assessment using island targets, we need to solve four distinct problems

1 Reliable classification of viewed locations in the sensor data as land, water, or clouds/ice; the first two to locate the islands, and the last to ensure the data are unobscured;
2 Location of contiguous land pixel clusters corresponding to islands,

3 Generation of a catalog of accurate reference island locations,
4 Robust matching of islands located in the sensor data with the reference locations, allowing for initial navigation errors which may be significant.

We have implemented this approach at NASA Goddard Space Flight Center (GSFC) in a navigation assessment system for the SeaWiFS mission. SeaWiFS will collect global ocean color data over a 5-year mission. The sensor measures radiances in eight visible and near-infrared bands; the instrument characteristics are summarized in Table 1 The sensor resolution is 1 12 km. at nadir, and the navigation accuracy requirement is 1 pixel (2 sigma). The primary data collected are Global Area Coverage (GAC), which will be subsampled every fourth line and pixel, with a pixel spacing of 4.5 km at nadir A limited amount of full-resolution Local Area Coverage (LAC) data are collected; in addition, full-resolution data are direct-broadcast to ground stations in High-Resolution Picture Transmission (HRPT) format.

The results of automated SeaWiFS navigation assessment will be used determine the accuracy of navigation on an ongoing basis and to develop improvements to the navigation segment of the scientific data processing. In the early mission phase, the results have been used to detect any static offsets and to characterize the spacecraft attitude control system. Over longer time scales, we plan to analyze periodic and secular errors and incorporate these results as refinements to the navigation system.

Our methods were onginally developed using simulated SeaWiFS data (Patt, Woodward and Gregg, 1997). Since the SeaWiFS launch we have refined and verified the data classification algonthm using flight data. The details are presented in Section 2, along with a summary of the methods described in the previous publication. In Section 3, we discuss the initial results of testing with SeaWiFS flight data.

Table 1
SeaWiFS instrument characteristics, including geometric characteristics and the band center wavelengths.

Instantaneous Field-of-View (IFOV)	1 5835 mrad 0.09"	
Scan half-width	58.3" 45"	(LAC/HRPT) (GAC)
Ground IFOV at nadir	1 12 km	
Pixels along scan	1285 248	(LAC/HRPT) (GAC)
Scan period	0.167 sec 0.667 sec	(LAC/HRPT) (GAC)
Tilt	± 20"	
Scan ground coverage	2802 km 1502 km	(LAC/HRPT) (GAC)
Band Center Wavelengths		
1	412 nm	
2	443 nm	
3	490 nm	
4	510 nm	
5	555 nm	
6	670 nm	
7	765 nm	
8	865 nm	

2. Methods

2.1 Sensor Data Classification

The goal of the classification algorithm is to locate islands that are not obscured from the sensor view by clouds Therefore we need a reliable technique for determining the source of the incoming radiance, by analysis of the data fiom one or more of the sensor bands. The assumption is implicit that each source has a unique spectral signature, that can be detected with the band complement of SeaWiFS, which would enable reliable classification.

Our method of data classificatron, previously determined using simulated SeaWiFS data, has now been refined using the flight data. We have followed the same overall approach as with the simulated data. We illustrate the revised method which uses data from Bands 2 and 8. Part of a SeaWiFS scene, which covers a region north and west of Australia, is shown for these bands (figures 1(a) and (b)). This image shows that for Band 2 the ocean and land are essentially indistinguishable. The clouds are visible as the bright regions in the image.

For Band 8 the relative bnghtness of the land and water are distinct, with land being generally bnghter However, some of the land radiances are darker than average, and some of the water radiances are brighter, resulting in overlap between the ranges Thus, this band alone does not provide useful land/water discrimination.

As before, we investigated correlations between the bands for the various data types. To begin this investigation we generated scatter plots of band vs. band to look for patterns that correspond to data types. A scatterplot of Band 8 vs. Band 2 radiances is shown for a subset of this region (figure 2(a)). The overall distribution appears as a long diagonal streak which is divided at the lower end By examination of the images we found that the divided region represents the land and water, with the points below the diagonal being the water The points to the nght of the divided region represent the clouds. These features show indications of a strong correlation between the two bands between these data types. We have examined numerous data scenes and found similar patterns.

Although the same basic pattern is present throughout a scene, the magnitudes of the radiances vary considerably due to changes in illumination and atmospheric path length. We found that the cloud radiances in Band 2 appeared to depend mainly on illumination, as represented by the solar zenith angle. Therefore, for cloud thresholding we normalized the Band 2 radiances to a constant solar zenith angle as follows

$$R_{2N} = R_2 / \cos \theta_0 \qquad (1)$$

where θ_0 is the solar zenith angle

The normalization for land/water thresholding used both the solar and sensor zenith angle to compensate for the combined effects of illumination and path length:

$$R_{NP} = R / (1 - \cos \theta_0 / \cos \theta) \qquad (2)$$

where θ is the sensor zenith angle.

The division of the land and water data along the diagonal in figure 2(a) suggests the final step in the analysis. if the average slope is removed fiom the distribution, then the classification can be performed using simple thresholds.

The slope of the distribution is estimated visually to be 0.7 We can remove this slope by computing the weighted difference of the normalized radiances,

$$R_D = R_{2NP} - 0.7\, R_{8NP} \qquad (3)$$

and plotting this difference vs R_{2N} (figure 2(b)). This plot shows the desired result, with the division between land and water radiances distributed roughly horizontally The data can now be classified using simple thresholds, as shown on the figure. As seen in the figure, the divisions are still not completely clean, and in fact the cloud threshold for Band 2 was arrived at after some experimentation. Currently we are using the following;. all data with R_{2N} greater than 18 are classified as clouds; then, all remaining data with R_D less than -3.2 are water, finally, all remaining data are classified as land. These values may be refined further after additional data are analyzed

This process can be applied rapidly to all pixels in a scene, with the data type stored as a flag for each pixel The results are shown for the same partial scene (figure 3), with white being clouds, grey being land and black being water Thls image shows that the land, water and clouds have been clearly identlfied, and several small islands are visible in the data, and thus the classificatron algorithm is performing as desired.

2.2 Island Location

Once the sensor data have been classified, there still remains the task of locating contiguous clusters of isolated land pixels in the sensor data which can be identlfied as islands. We also need to consider the requirement of ensuring that the island view is not contaminated by clouds We have chosen an algorithm which identifies candidate pixel segments in senes with the data classification, so that the total number of pixels which need to be considered as potential islands is a small percentage of the total.

Once an island has been located as a cluster of land pixels, longitudes and latitudes for all pixels in the cluster are computed using the navigation data for the scan lmes, and the centroid coordinates (mean values of the longitude and latitude) are computed for the pixel cluster, along with the size of the cluster in each dimension For a given cluster of N pixels, the calculations are performed as follows.

$$C_{lon} = \sum_{i=1}^{N} lon_i / N \quad (4)$$

$$C_{lat} = \sum_{i=1}^{N} lat_i / N \quad (5)$$

$$S_{lon} = \max(lon_i, i=1,N) - \min(lon_i, i=1,N) \quad (6)$$

$$S_{lat} = \max(lat_i, i=1,N) - \min(lat_i, i=1,N) \quad (7)$$

where C_{lon}, C_{lat}, S_{lon} and S_{lat} are the centroid coordinates and size in longitude and latitude, respectively.

2.3 Generatlon of the Reference Island Catalog

We have selected a number of criteria for the islands to be used as references for target identification.

1 Islands need to fall within a useful size range. For SeaWiFS the minimum island size is about 0.5 pixels in area, where a pixel is 1 12 km square at nadir We have selected approximately 50 km as the upper limit.
2 Islands need to be sufficiently separated from other islands and land masses to be readily distinguishable in the sensor data.
3 Islands need to be reasonably compact in shape.
4. Island locations must be well determined t
5 Islands must be reasonably distributed worldwide for global navigation assessment

We generated a catalog which would serve our purpose. We used the World Vector Shoreline (WVS) database distributed by the Natlonal Geophysical Data Center (NGDC). Altogether we cataloged nearly 7800 islands worldwide.

We checked the islands against the criteria (1) and (2) above. We generated separate catalogs for use with SeaWiFS GAC and LAC data by applying the appropriate minimum separation cntena for each. The resultant d o g s contain approximately 4300 islands for GAC and 6500 for LAC.

2.4 Island Matching

We sought a robust matching algorithm based on the expectation that the initial errors in the satellite navigation might be large compared to the SeaWiFS pixel resolution These navigation errors could result ftom timing, orbit position or instrument pointing errors, or some combination of these.

The simplest matching algonthm is direct, or "nearest neighbor" matching, in which each island located in the data is associated with the closest cataloged island, within a specified matching tolerance. This method is reliable if the expected navigation errors are much less than the typical separation of catalog islands, but degrades rapidly if the errors approach the minimum island separation specified for the catalog.

If the errors are large, the probability of correctly matching the islands can be improved substantially by use of geometric pattern matching. Our implementation of this approach is known as tnplet matching, in which separations for three islands located in the data are compared with the separations of cataloged islands. Triplet matching provides more robust matching of islands in the presence of large navigatlon errors, although the computational requirements are correspondingly greater

In addition to the locations or angular separations of islands, the sizes *can* also be used to aid the matching algonthms. This is mostly useful for the larger islands, since the vast majority of islands, in both the data and the catalog, are a few pixels or less in size.

This pattern-matching approach has already been applied in other situations. The GSFC Flight Dynamics Facility (FDF) has developed algorithms and software to match star tracker observations to a reference catalog. The FDF routines perform direct and triplet matchmg, and also use stellar magnitudes to improve the reliability of the matching.

We obtained FORTRAN source code for the matching algorithms ftom the FDF and were able to incorporate it into our navigation assessment software with minimal modifications. We substituted island sizes directly for stellar magnitudes, and computed the island position uncertainties based upon the position in the scan line and the number of pixels.

The details of the island matching are as follows. The matching algorithm (direct or tnplet) is selected via an input parameter. The first step for both algorithms is

the selection of candidate islands from the catalog. For each island located in the data, candidates are selected from the catalog based upon a specified search radius. The size of each candidate is also required to match that of the located island within a tolerance, to eliminate obvious mismatches.

At this point the algonthm branches according to the matching method selected For direct matching, each located island with candidates is matched with the nearest candidate from the catalog.

If tnplet matching is selected, the triplets are formed starting from the top of the sorted island list (i.e., the islands with the fewest candidates). Triplets are also required to pass a colinearity test, i.e., the smallest angle of the spherical mangle formed by the three island locations must be greater than a specified minimum value. This check ensures that the angular separations for the tnplet are geometrically independent measurements of the islands' relative positions. The angular separations for the three islands are then compared with the corresponding separations of all possible combinations of candidates. If the separations match within a specified tolerance the islands are considered to be matched. The process continues until either all of the islands are matched or all possible triplets are checked.

The computational burden of matchmg is reduced by sorting the islands according to the number of candidates prior to any attempted matching. By forming the initial triplets using islands with the fewest candidates (presumably one per island at the top of the list), the number of candidate combinations which needs to be checked is minimized. After a few of these islands are matched, they can be used to form triplets with other islands having more candidates, while still keeping the combinations of candidates to a minimum. In addifion, once an island is matched with a candidate, all of the remaining candidates are removed from the list. By using this logic nearly all of the tnplets can be formed with at least one of the islands having only one candidate.

As stated, the choice of matching algorithm depends upon the expected navigation errors. If the errors are larger than the pixel spacing, the tnplet algorithm provides robust matching, but is more computationally intensive and may result in mismatches and unmatched islands. For small navigation errors the direct matching algorithm is both fast and accurate. The reliability of these methods depends upon choosing values for the parameters (candidate search radius, matching tolerance and tnplet minimum colinearity angle).

3. Results and Discussion

We have applied the island location and matching methods to SeaWiFS GAC data and to HRPT scenes, following the start of routine SeaWiFS data collection on August 18,1997 The questions which we wished to answer were: 1) What is the number and distribution of islands located in SeaWiFS data7 2) How well can the results be used to indicate the typical navigation errors7 Our test results and discussion are presented in the following sections.

3 1 Island Location Results

The average number of islands located for one day of GAC data is 38.4 islands per orbit, with a range of 0 to 115 (table 2). These results show reasonable coverage throughout the day despite varying cloud cover The variation from orbit to orbit mainly reflects the vanations in the global distribution of islands, although the distribution of clouds for a given day also has a large effect Orbits which pass through the Southwest Pacific/Southeast Asian region, the East Mediterranean/Arabian Peninsula region and the Caribbean/Central America region tend to produce large numbers of islands.

The table also shows the results of island matching. For this test the direct matching method was used, with the tolerance set at 0.1 degree. The average number of islands matched is 17.6 per orbit, with a range of 0 to 70

Table 2.
Islands located in 1 day of GAC data

Orbit	Number of Islands Located	Number of Islands Matched
1	115	70
2	41	16
3	16	3
4	53	24
5	37	25
6	32	16
7	23	5
8	31	1
9	15	6
10	0	0
11	17	2
12	21	9
13	54	29
14	56	22
15	53	36

We also need to ensure that the located islands are well distributed geographically, to support analysis of navigation errors which may correlate with other variables (e.g., orbit position). We examined the distribution of located islands, to ensure that the navigation errors are being adequately sampled over the full range of latitudes and longitudes used for GAC data collection. The distribution of islands for the day (figure 4) shows that most of the islands are concentrated in the mid-latitudes, as expected, but a significant number are found near the northern extremes. Thus we conclude that the islands located in the GAC data are reasonably well distnbuted.

We also tested the approach using several HRPT passes collected a station located at NASNGSFC in Greenbelt, MD. Islands were located in all passes (table 3) and were matched in all but 1

The number of islands observed in HRPT data depends upon the location of the receiving station. The area viewed from the GSFC station mcludes much of the Caribbean Sea, which results in large numbers of islands being observed. Again, cloud cover has a significant effect on the variations in the results The average number of islands for these passes is 176, and the average number of matches is 42 These results represent a substantial number of targets for both GAC and HRPT data

Table 3.
Islands located in each of 7 passes of HRPT data.

Day / Pass	Number of Islands Located	Number of Islands Matched
271 / 1	209	81
272 / 2	143	19
273 / 1	68	27
273 / 2	489	117
274 / 1	34	0
274 / 2	137	44
274 / 3	155	7

The seemingly low rate of matching for HRPT data merits further explanation. We examined several sets of located islands with low matching rates. We concluded that the unmatched "islands" correspond to coastal regions with either numerous, closely spaced islands or long, slender barrier islands. In the former case, the islands have been rejected fiom the catalog as being too densely distributed for reliable matching. In the latter case, the extended barrier islands are viewed in the data as multiple small islands, which do not correspond to the cataloged islands. The higher resoluhon of the HRPT data results in many more such objects being located in coastal areas. Table 6 shows that, where islands were located in multiple daily passes, the matching rates are generally lower for the second and third passes of the day, which include more coastal regions around the Gulf of Mexico.

In all cases examined we found that the islands outside of the congested areas were located and matched with high reliability We believe that the algorithms are performing as designed, by first locating all potential targets and then matching those which can be reliably used for assessing navigation accuracy The number of unmatched targets in congested regions reflects the effectiveness of the data classification and island locahon algorithms rather than a failure of the matching algorithm. This also justified the size of the reference catalog; clearly a smaller catalog based on more restrictive criteria would result in fewer matches. Most importantly, the typical number of matched targets is substantially larger than would be obtained by manual selection of ground control points, and allows for detailed and consistent navigation assessment.

3.2 Navigation Assessment Results

The ultimate test of our methods is the success in determining and improving the navigation accuracy of SeaWiFS data. As of this writing the navigation assessment for SeaWiFS is still in the early stages Here we briefly illustrate the us of the island matching results for navigation assessment

The results of the island matches are a set of coordinate differences (latitude and longitude) for each island matched. A scatter plot of these differences for a sample orbit of GAC data (figure 4(a) provides an overall assessment of the navigation accuracy The figure shows a cluster of points centered at roughly (0.01, 0.03), indicating that this is the average navigation error for the scene. The size of the cluster is consistent with the GAC pixel spacing. The points outside of the cluster may represent either larger navigation errors or simply lower-quality matches. A more detailed analysis would show whether these errors are correlated with latitude, scan angle or other factors. Combining this information fiom multiple orbits and days would allow the results to be evaluated for consistency and analyzed for systematic errors.

A similar plot for an HRPT scene (figure 4(b)) shows a smaller cluster, with the center near that for the GAC data (0, 0.04). The size of the cluster in longitude is consistent with the LAC pixel size, while a larger spread is seen in latitude. This is likely due to variations in navigation over the scene. Again, a number of points are scattered around the cluster

Although preliminary, these results indicate the usefulness of the island matches for navigation assessment, both globally using the GAC data and to higher accuracy over local regions using HRPT data

4. Conclusion

We have developed a system for automated assessment of satellite-based Earth sensor navigaoon using island targets. The method involved filtering the sensor data to locate islands, and matching them with reference island locations in a catalog. We developed algorithms for classifymg sensor data and locating the islands using the results of classification; generated a global catalog of island locations and sizes based on a precise vector shoreline database; and adapted pattern-matching software to perform robust matching of located and reference islands.

We tested the methods using SeaWiFS sensor data representing both GAC and direct-broadcastHRPT data collection. The targets located in the GAC data are reasonably well distributed, with islands found in nearly every orbit and distributed over the full range of latitudes and longitudes. The tests with the HRPT data, most of which included parts of the Caribbean Sea, produced large numbers of islands. The island matching algorithm succeeded in matching almost half of the islands found in the GAC data and about one quarter of the those in the HRPT data, the unmatched islands correspond to congested coastal regions or bamer islands. The average number of matched islands in the initial tests is about 17 per GAC orbit or 42 per HRPT pass. This represents a substantial number of targets which can be generated automatically for navtgabon assessment during the SeaWiFS mission.

Acknowledgements

The work presented here was performed in support of the SeaWiFS Project at Goddard Space Flight Center in Greenbelt, Maryland. The authors wish to express their appreciation to NASA/GSFC, the Project office and the Global Change Data Center for the continuing support of this work We would also like to thank the GSFC Flight Dynamics Facility for providing the pattern-matching algorithm source code.

References

BORDES, P ,BRUNEL, P ,and MARSOUIN, A, 1992, Automatic adjustment of AVHRR navigation, *Journal of Atmospheric and Oceanic Technology,9*, 15-27

EMERY, W J., BROWN, J., and NOVAK, Z.P., 1989, AVHRR image navigation: *summary* and review, *Photogrammetric Engineering and Remote Sensing, 55,* 1175-1183

KRASNOPOLSKY, V M, and BREAKER, L.C., 1994, The problem of AVHRR image namgation revisited, *International Journal of Remote Sensing,* 15,979-1008

MARSOUIN, A., and BRUNEL, P ,1991, Navigation of AVHRR images using ARGOS or TBUS bulletins, *International Journal of Remote Sensing,* 12, 1575-1592.

O'BRIEN, D.M. and TURNER, P.J., 1992, Navtgation of coastal AVHRR images, *International Journal of Remote Sensing,* 13,509-514.

PATT, F ., WOODWARD, R and GREGG, W , 1997, An automated method for navigation assessment for Earth survey sensors using island targets, *International Journal of Remote Sensing,* in press.

ROSBOROUGH, BALDWIN and EMERY, 1994, Precise AVHRR image navigation, *IEEE Transactions on Geoscience and Remote Sensing,* 32,644-657

Figure Captions

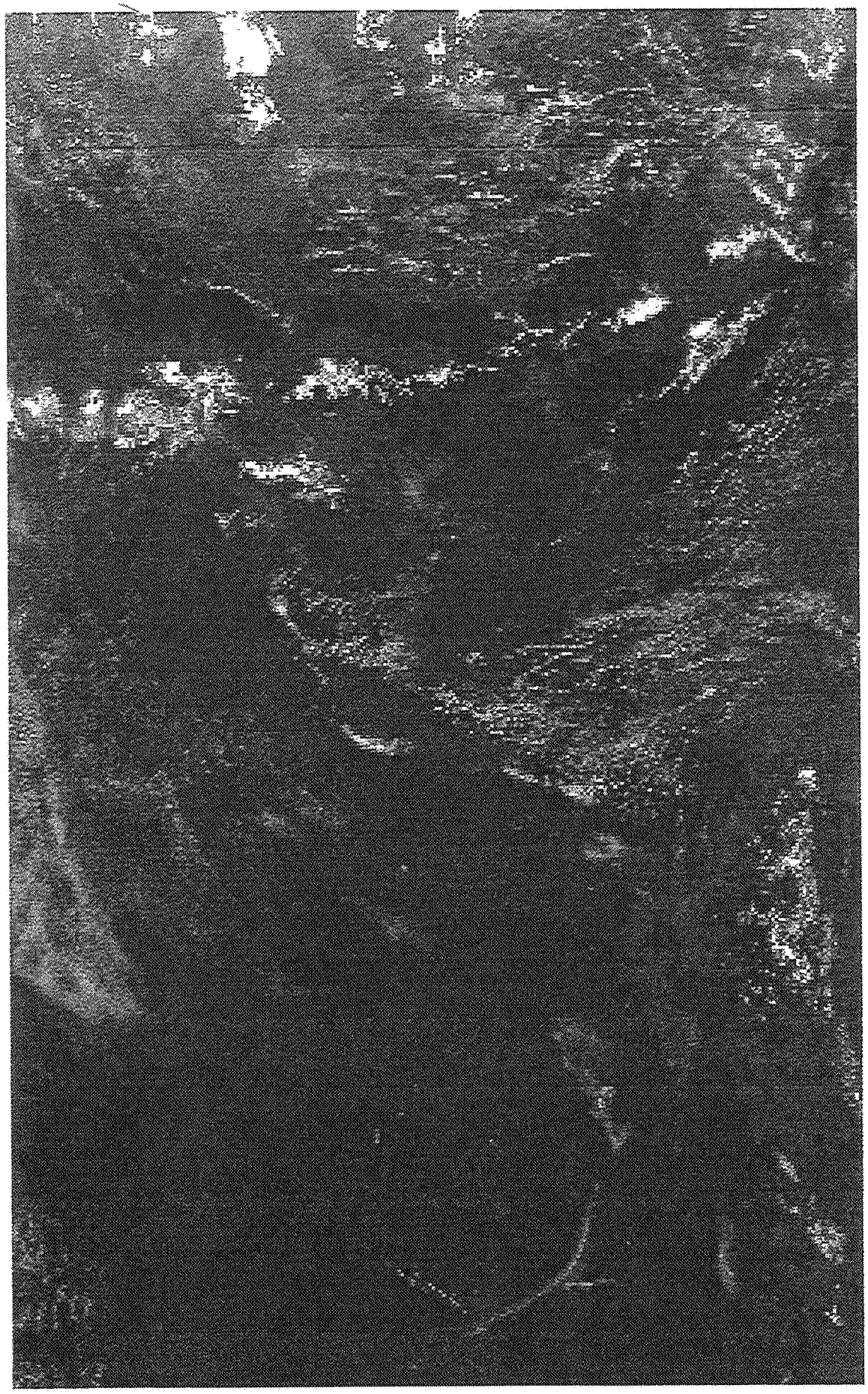
Figure 1 SeaWiFS GAC data for a scene including northwest Australia, (a) Band 2; (b) Band 8

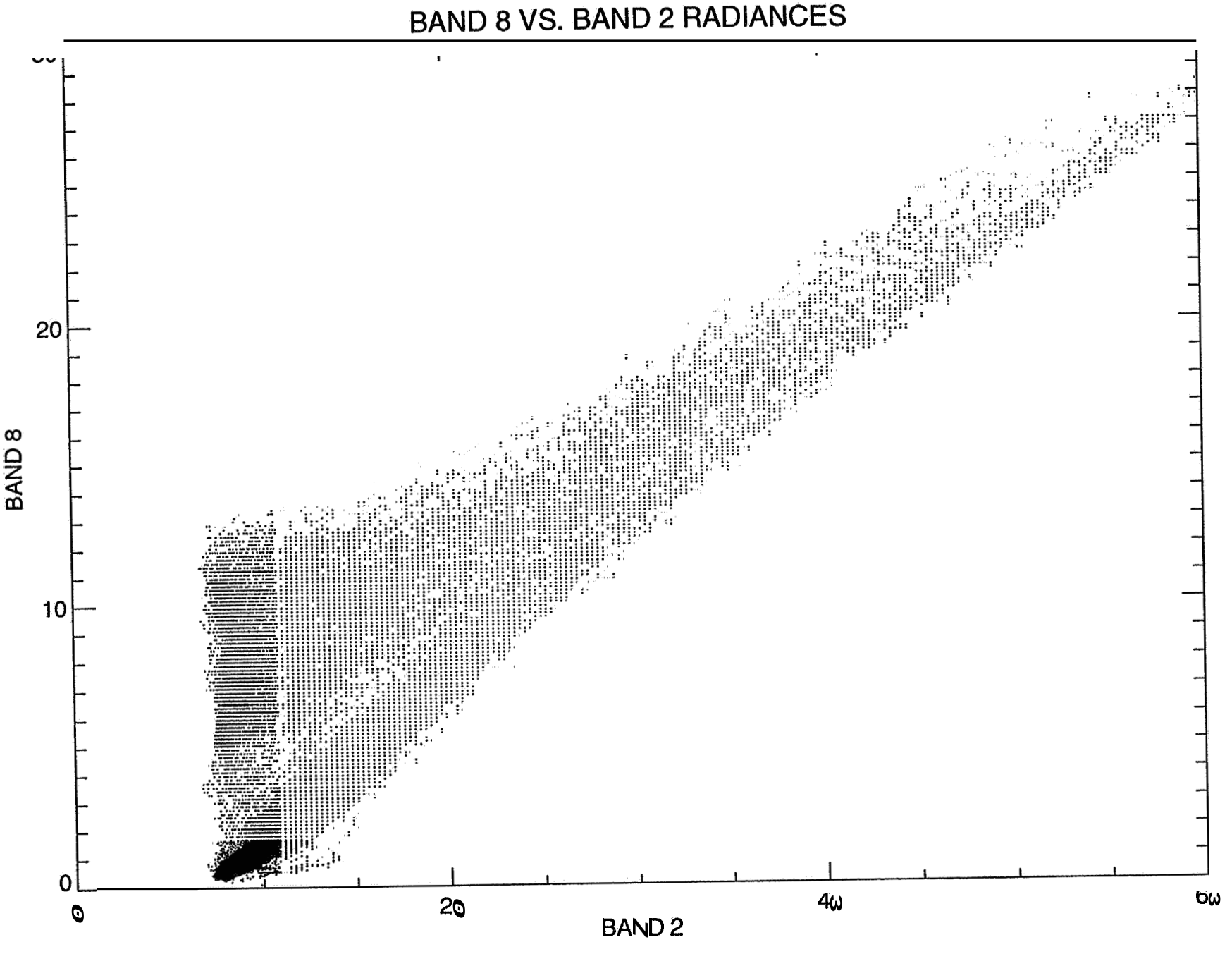
Figure 2 These plots illustrate the method for classifying SeaWiFS data using Bands 2 and 8, (a) Band 8 vs Band 2 radiances; (d) Weighted difference of normalized Bands 8 and 2 vs Band 2, with thresholds indicated on plot.

Figure 3 SeaWiFS scene showing the results of data classificaoon; data types computed using thresholds described in text, with white representing clouds, grey representing land and black representing water

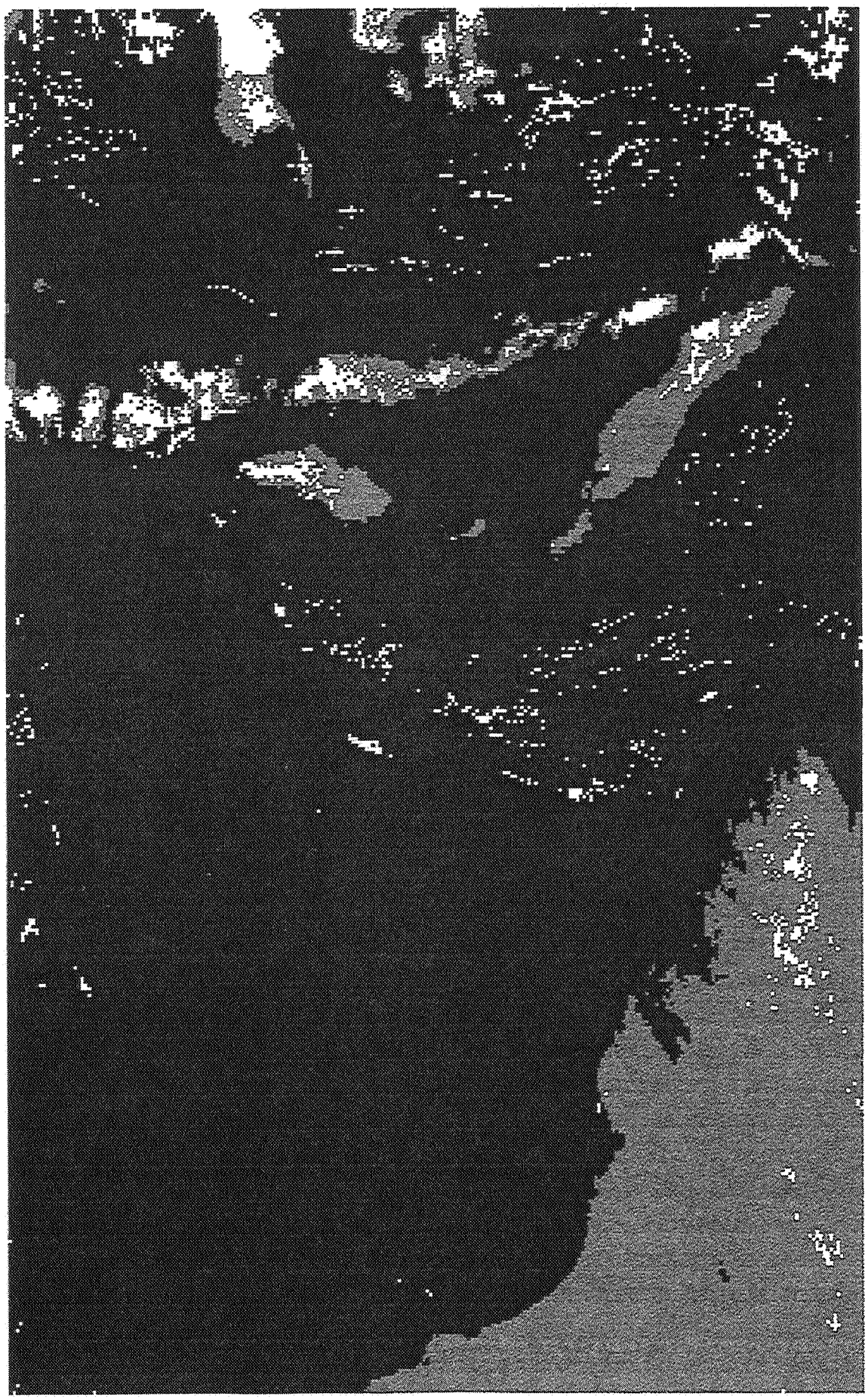
Figure 4. Geographic distribution of islands located in 1 day of GAC data.

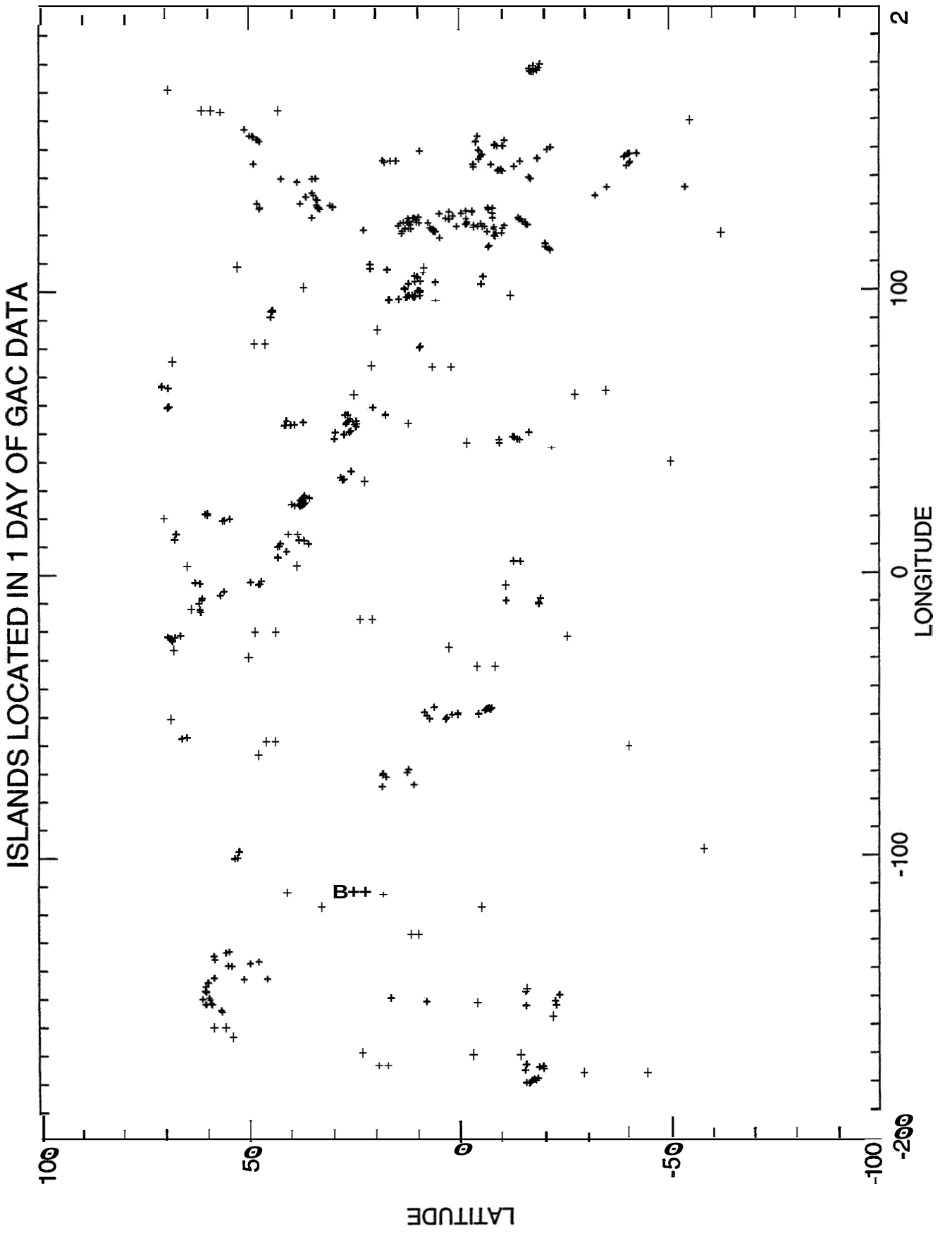
Figure 5 Latitude vs. longitude coordinate differences for matched islands, (a) in one GAC orbit; (b) in one HRPT scene.

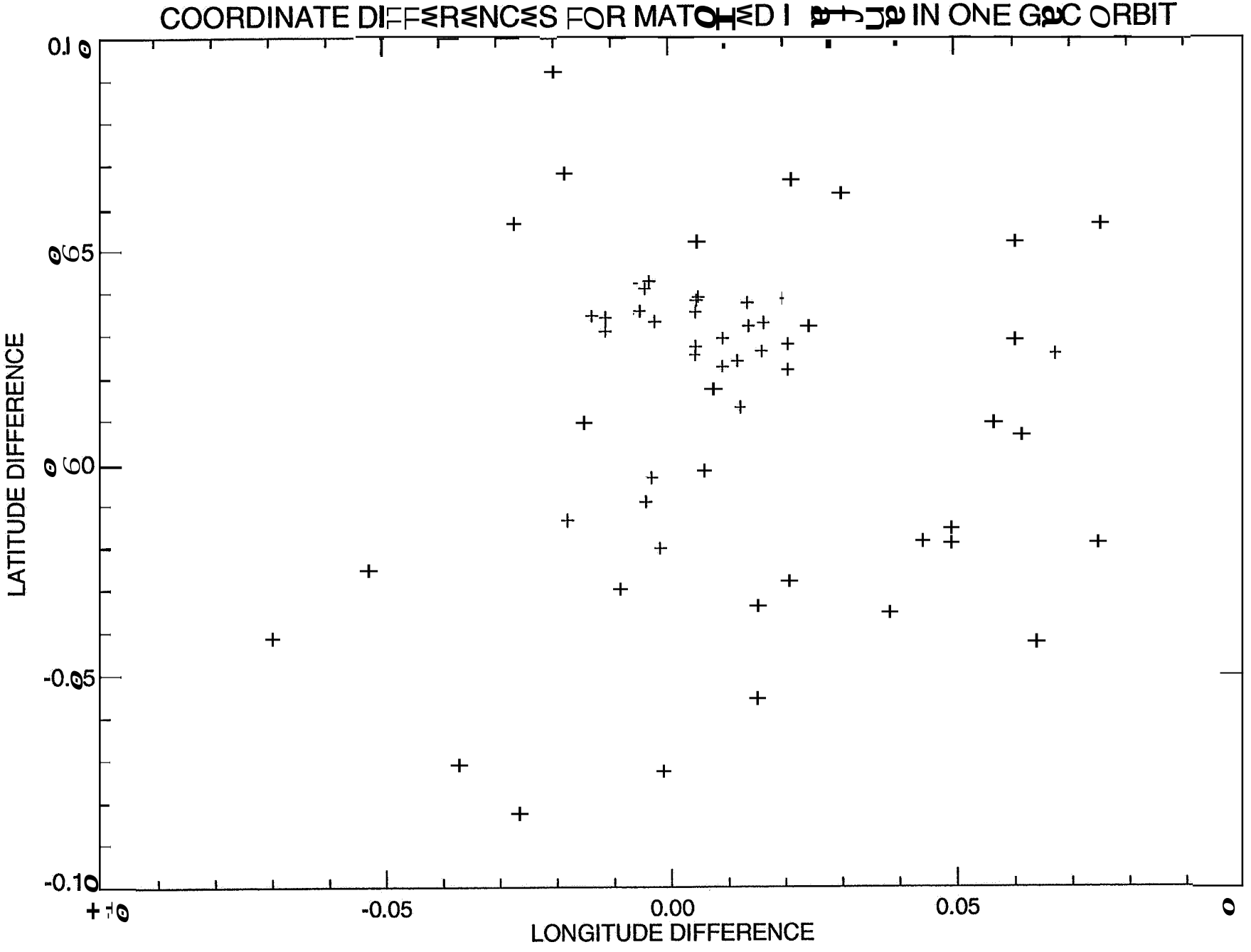


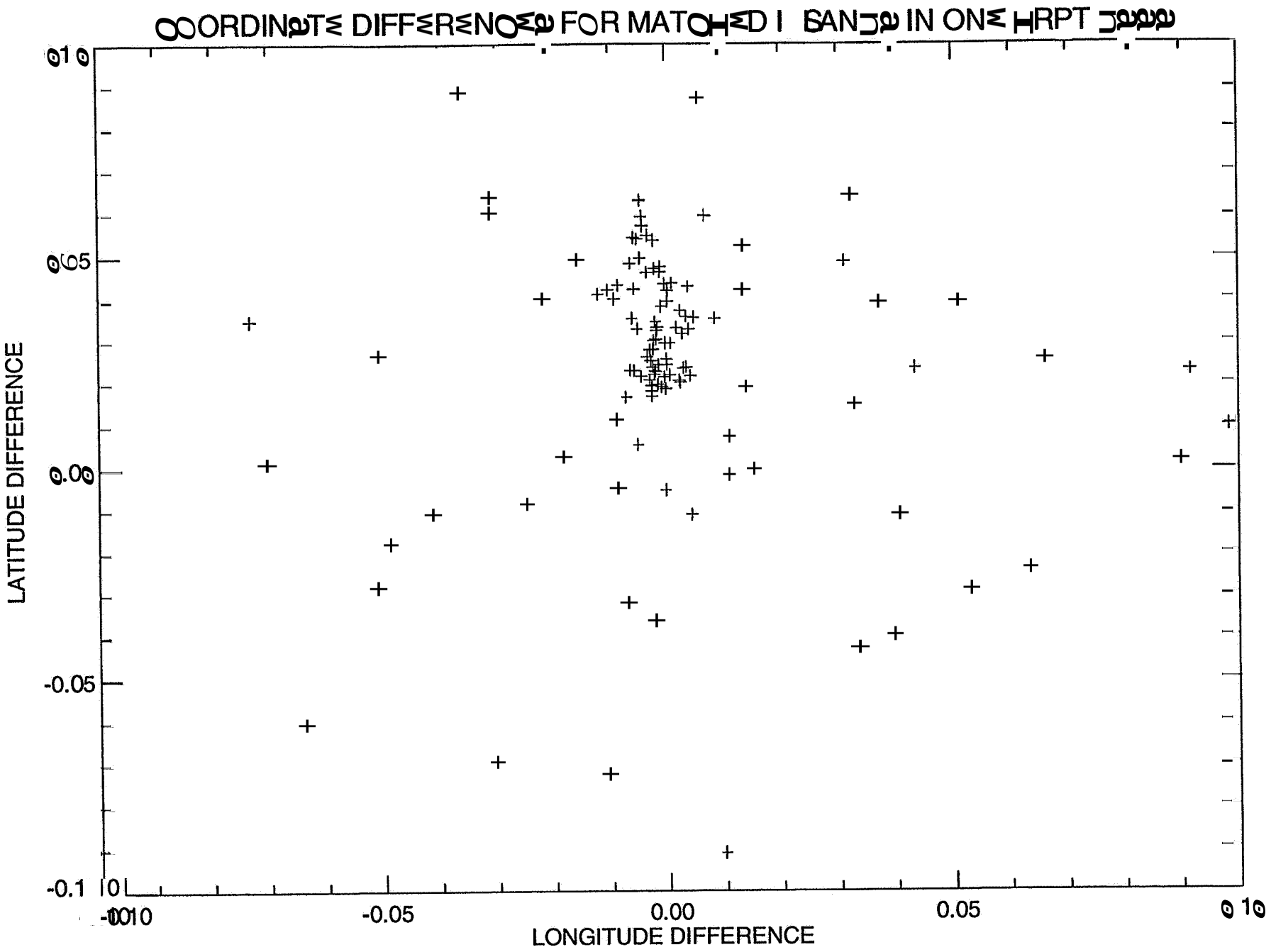

BAND 8 VS. BAND 2 RADIANCES
BAND 8
0
10
20
BAND 2
0
20
40
60

WEIGHTED DIFFERENCE OF BANDS 8 AND 2 VS. BAND 2
BAND 8 - 0.7*BAND 2
BAND 2
0
20
60




ISLANDS LOCATED IN 1 DAY OF GAC DATA
B++
LONGITUDE
LATITUDE
200
100
0
-100
100
50
0
-50
-100


COORDINATE DIFFERENCES FOR MATCHED
IN ONE GOC ORBIT
LATITUDE DIFFERENCE
LONGITUDE DIFFERENCE
-0.05
0.00
0.05


LATITUDE DIFFERENCE
LONGITUDE DIFFERENCE
-0.05
0.00
0.05

MODIS Land Ground Control Point Matching Algorithm

Robert E. Wolfe
robert.wolfe@gsfc.nasa.gov

Mash Nishihama
mash@ltpmail.gsfc.nasa.gov

and David Solomon
dsolomon@ratmos.gsfc.nasa.gov

Hughes STX
Greenbelt, Maryland 20770

Abstract:

The Moderate Resolution Imaging Spectroradiometer (MODIS) Land Ground Control Matching Algorithm performs image correlation to determine residual geo-location errors. MODIS's response and resolution **is** simulated by applying a model of MODIS's viewing geometry and transfer function to high resolution image data. The source for the ground control point reference image data is precision and terrain corrected Landsat TM data and a corresponding high resolution elevation model. First, the location of each pixel in the TM ground control point image **is** transformed into a along scan and across scan angles based on MODIS's location at a given time. This transformation includes the pixel's elevation. Then, the MODIS modulation transfer function (MTF) is applied in this angular space to generate the simulated pixels which are correlated with the actual MODIS pixels. The best match is determined by shifting small fractions of a MODIS pixel in the angular space, reapplying the MTF and performing the correlation. This method is expected to produce a measurement of the residual geo-location errors with an accuracy of a small fraction of a MODIS pixel.

Algorithm Development for the Automatic Registration of Satellite Images

Paul M. Dare, Renaud Ruskone, Ian 3 Dowman

Department of Geomatic Engineering
University College London, Gower Street, London WC1E 6BT
United Kingdom

ABSTRACT

The concept of a flexible system for fully automatic multisensor image registration is being developed in the framework of a European project (no. ENV4-CT96-0306) called ARCHANGEL (Automatic Registration and CHANGE Location). It is designed to accept many different types of data (SPOT, SAR, maps etc.) and to automatically find common features that can be identified in the different data sets. When these common features have been identified, they can be used to determine common points which are in turn used to register the images. Eventually, this system aims at accommodating registration between these different data sources, and detecting the changes. This paper describes the evolution of the ARCHANGEL project from a system that has already been developed for the automatic registration of pairs of SPOT images, known as PAIRS (Prototype Automatic Image Registration System). Some details of the ARCHANGEL project are described, including the concept of a "registrability" measurement for predicting the quality of features to be matched

I. INTRODUCTION

The concept of using multisensor image data for improved land cover monitoring/mapping has been around for many years now However, as the range of different sensor types increases and the resolution of the images gets better, the difficulties involved with multisensor image registration become more pronounced. Both accuracy of registration and speed of processing are becoming more and more important. Over the past few years this has led to the development of many novel image processing techniques in the field of image registration.

When two images from different sensors are registered, errors in the registration cause misalignments and hence make analysis of the merged data set more difficult and less accurate. The sources of these registration errors include:

- differing sensor geometry,
- platform orientation;
- atmospheric effects; and
- relief.

It is generally possible to accurately model the sensor geometry and, if the state vectors of the platform are accurately known, the platform orientation can be determined as well. However, it is much more difficult to correct for errors in the registration caused by atmospheric effects, and if no accurate maps and digital terrain models (DEMs) of the area being imaged are available, then correcting for effects caused by terrain is very difficult. Photogrammetric techniques originally developed for the rectification and analysis of aerial photographs can generally be extended to satellite data (Toutin, 1995) but high quality maps and DEMs are required. If these auxiliary data sets are not available, an alternative approach must be used to minimise the errors that occur when multisensor data is registered.

In the spatial domain there are two main methods for automatically registering images: feature based matching and area based matching (Fonseca and Munjunath, 1996).

For area based matching to produce accurate results the radiometric distribution of the two images must be very similar However, if these distributions are very different (e.g. optical and SAR images) then area based matchers will fail. In this case, feature based matching must be used.

This paper describes work under way at UCL and other European partners to develop a fully automatic end user system for integrating multisensor image data. This project, called ARCHANGEL, evolved from work done on the automatic registration of SPOT stereo images. Section II describes this project, called PAIRS, in more detail, whilst section III gives an overview of the ARCHANGEL project. Section III also describes in detail some aspects of the ARCHANGEL project, including segmentation schemes for automatic feature extraction, feature matching, and change location.

II. THE PAIRS PROJECT

The PAIRS project, funded by the Western European Union Satellite Centre and involving a number of European partners, aimed to develop a system to automatically register multisensor data sets. The two important aspects of the project were full automation and end-user applicability The PAIRS system is based on published algorithms but draws particularly on work done at the University of Stuttgart for the strategy for determining conjugate points (Haala et al, 1993) and on the use of robust estimation for removing erroneous points. The individual algorithms used in the system are used in many image understanding problems. The Förstner operator (Fiirsmer and Gulch, 1987) is well known for selecting interest points and this or similar algorithms have been used in a number of applications.

The first phase of the project concentrated on registering similar pairs of images (e.g. SPOT stereo pairs) using two dimensional plane transformations; i.e. with no correction for terrain. After the images have been loaded into the PAIRS system the region of interest and an initial approximate registration are defined. A large number of prominent well-defined points, known as interest points, are then selected using the Förstner interest operator (Fiirstner and Gülch, 1987). Then the corresponding points in each of the images have to be matched. It is possible to predict where an interest point from one image will appear in the other, but there may be some ambiguity regarding exactly which corresponding pixels should be

matched to each other. This is resolved by applying a cross correlation to the area around the pixels in question, the result of which is used to accept or reject a match. When all the conjugate pairs of pixels have been determined in this manner, they are used to determine the parameters of an affine transformation using a least squares method. Results showed that a root mean square error of about twenty metres was achievable when a pair of 1024x1024 scenes were used (Dowman et al, 1996a). The two main sources of these errors were the varying terrain in the image which was not corrected for, and some changes which took place on the ground between the acquisition of the first and second images.

The second phase of the PAIRS project was designed to accept more types of imagery and correct for terrain errors too. The overall layout of the PAIRS system is shown in figure 1

When images from different sensors are to be registered, interest points are not always the most appropriate matching entities because of the radiometric differences between the images. Instead features that can be extracted from both types of image are needed. Polygons were chosen as the matching entities since they should represent real world objects and therefore they should appear in both image types.

A number of methods have been used to extract polygons. Work has been done with aerial photographs and with KVR-1000 satellite data (Dowman et al, 1996b) as well as with ERS-1 SAR and SPOT panchromatic data (Dare and Dowman, 1997). The various methods include smoothing, segmentation, edge enhancement, edge thinning and removal of small polygons.

The basic techniques of matching polygons in the PAIRS system are adapted from a method where polygons are characterised by a number of parameters such as size, shape, area and location in the image (Abbasi-Dezfouli and Freeman, 1994).

The initial translation and orientation must first be fixed by defining a few polygons that have good matches based on a first pass through the images. An iterative approach then allows corresponding polygons to be identified. A large number of polygons are not necessary but it is important that the selected ones are evenly distributed across the pair of images.

Once established the corresponding polygons must then be exactly matched in order to extract conjugate points. A variation on the method of dynamic programming was used to do this (Newton et al, 1994). The first step of the algorithm stores the edges of the extracted patches as line segments. The algorithm then works through the line segments in a stepwise manner, projecting each line segment, one by one, from the secondary image to the reference image. A search area is set up around each pixel in the projected line segments, and a search is conducted for a corresponding matching pixel in the reference image. When a match is found (by minimising a cost function based on the edge strength and edge direction) the co-ordinates of the conjugate pair of matched pixels are used to improve the initial registration for projecting the line segments. This process is repeated until all the pixels in all the line segments have been considered. The output of the algorithm is a text file containing conjugate pairs of matched pixels, or tie points. This tie point data is used to generate the parameters of an affie transformation which in turn is used to register the secondary image to the reference image.

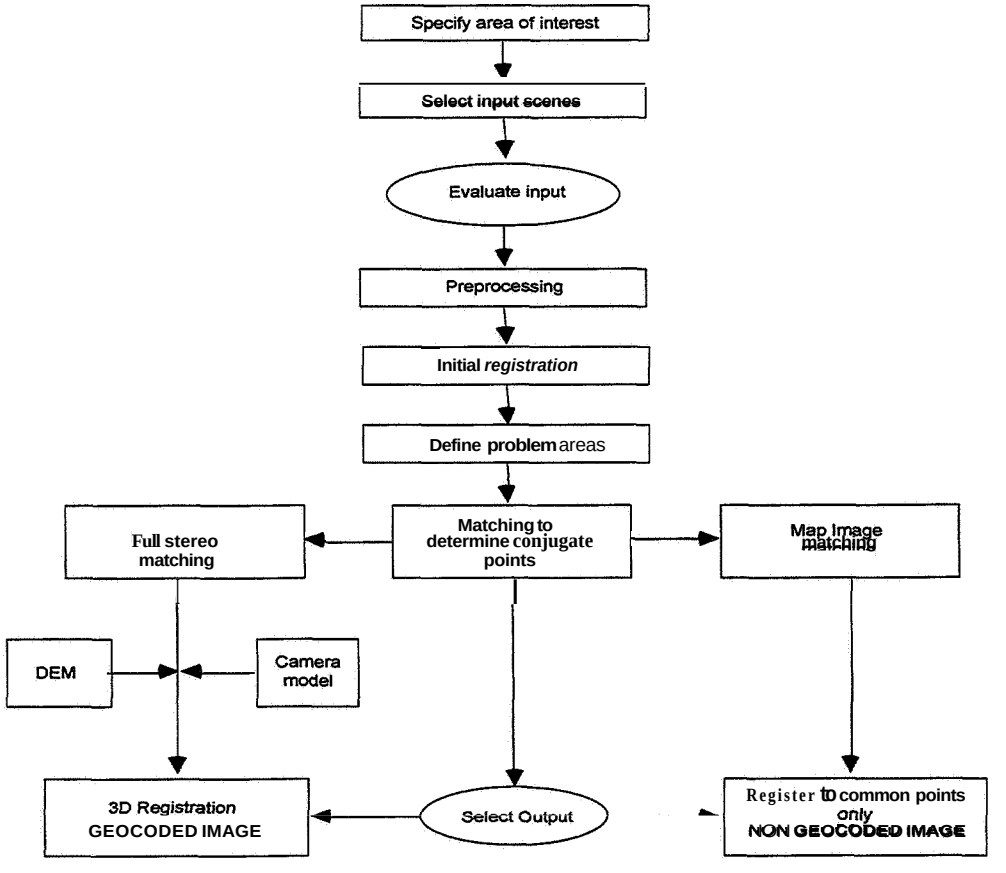


Figure 1 Flow diagram of PAIRS

PAIRS is built in an open architecture concept and is hosted on a SUN platform using the UNIX operating system. The system architecture is based on a software package called XPECT, a visual programming environment in which modules (processes) are represented iconographically and can be visually strung together. The modular design of the PAIRS within XPECT results in a high degree of flexibility All the procedures need not be run to produce an image registration. For example, the intensity based matching can be omitted if required. The system can be easily expanded or modified and XPECT is ideal for system development

III. THE ARCHANGEL PROJECT

The ARCHANGEL project aims to extend the work of the PAIRS project by incorporating algorithms that can deal with data coming from a topographic database or a scanned map. The result will be a robust system for registering a range of images to topographic data. However, in aiming to register a variety of image types to topographic data, a number of new problems not encountered in the PAIRS project come to light. Although some of the algorithms developed in PAIRS are used in ARCHANGEL, the main difference between the two systems is the extraction and matching of similar features from an image and the corresponding topographic dataset of the same area. This problem was not tackled in PAIRS, but it *forms* a cornerstone of ARCHANGEL.

ARCHANGEL is funded by the European Union 4th Framework Environment and Climate Programme. The project members include University College London, University of Stuttgart, University of Stockholm, University of Oporto, Earth Observation Sciences Ltd., UK, and the Centre for Environmental Satellite Data, Sweden. A full representation of ARCHANGEL is given in figure **2.**

The work under way at University College London centres on the extraction and matching of features from the various image types. University of Stuttgart is responsible for developing the tools for the processing of the topographic data for matching, as well as working on the matching of linear features. University of Stockholm is concentrating on researching edge detection techniques and studying change detection algorithms. Towards the end of the project, all the algorithms developed will be integrated into one architecture by Earth Observation Sciences Ltd, and validated by Centre for Environmental Satellite Data.

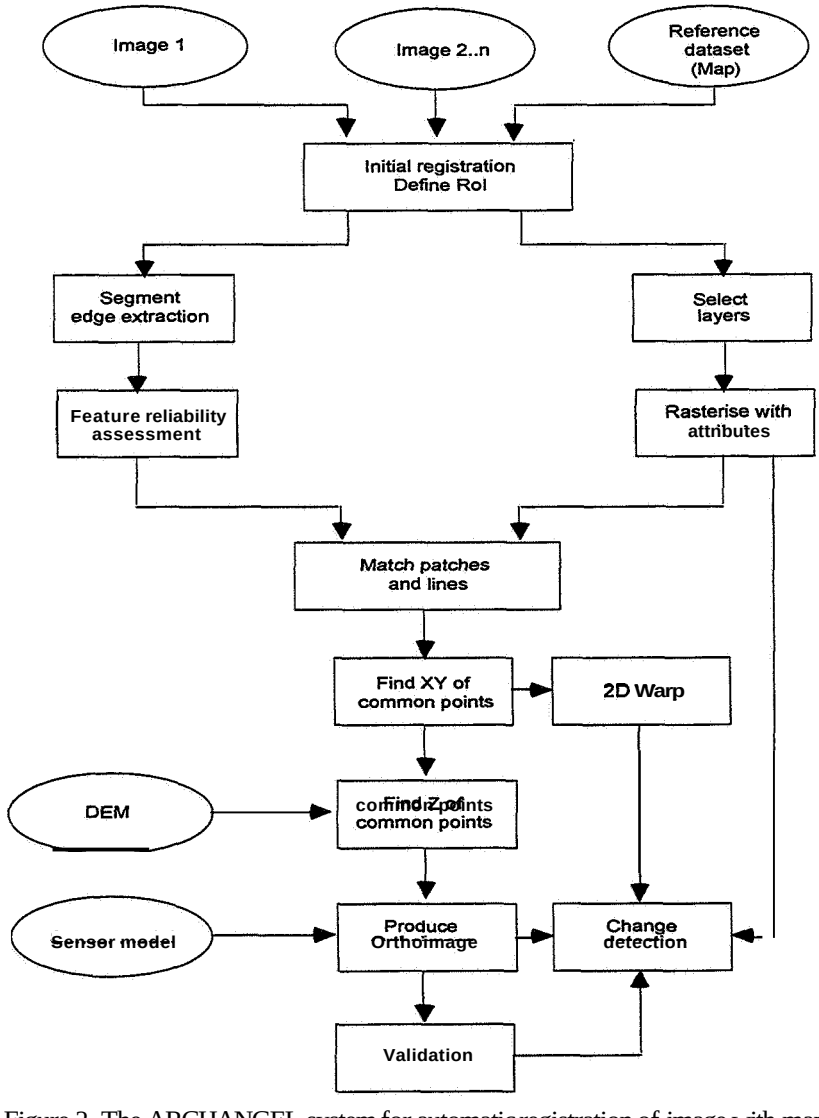


Figure 2 The ARCHANGEL system for automatic registration of image with maps.

The remainder of this section describes two of the ARCHANGEL procedures and algorithms in more detail: feature extraction using image segmentation, and the registrability assessment.

Feature extraction using image segmentation

In ARCHANGEL features are extracted from optical images using a region based segmentation algorithm that is designed to efficiently segment a range of different image types (Ruskoné and Dowman, **1997).** It was tested using LANDSAT TM and SPOT panchromatic data. The algorithm is a classical region growing technique based on the clustering of neighbouring pixels having similar properties. All the agglomerated pixels form a region (or segment). The agglomeration criterion takes into account a grey level difference between a candidate pixel value and the 'average grey level of the region. It is grounded on the comparison with a threshold defined as (Duplaquet and Cubero-Castan, **1994):**

$$Thr = b + m*M_i - v*V_i$$

where:

b is the difference between the average grey level of the region i and the candidate pixel;

m, the influence of the average grey level of the region i (Mi); and

v, the influence of the grey level variance of the region i (Vi),

After the basic agglomeration step, it was noticed that some adjacent regions can still be merged. These segments can take three different aspects:

- regions that a local distortion have split;
- "noise" pixels that form a very small region and that are usually isolated in the middle of a bigger segment; and
- segments in heterogeneous areas. Our algorithm is designed to identify homogeneous regions and a side-effect is the normal oversegmentation of the region that does not respect this criterion (typically, the forest or urban areas).

To remedy this oversegmentation, we apply an iterative merging method that works in the "region space" (which represents a significant compression of the information). The convergence is controlled by two different parameters: a maximal number of iterations and the number of mergings. If the maximal number of iterations is reached or if no merging has been done, the merging process stops.

The first case of oversegmentation is processed by finding out for each region's neighbours the most similar one in terms of radiometric probability density function. For that, we use the usual class separation distance (Castleman **1996).**

$$D = |\mu_1 - \mu_2| / (\sigma_1^2 + \sigma_2^2)^{1/2}$$

where:

μ_1 and μ_2 are the average grey level of regions **1** and **2,** and

σ_1 and σ_2 are the standard deviation of their grey levels.

At each iteration, a threshold is used to assess the class similarity to avoid a merging "chain reaction" The threshold is linearly decreasing to be null at the maximal iteration.

The second type of oversegmentation is solved by analysing the ratio between the longest boundary separating the current region from its neighbours and its area. Again a threshold is applied to decide whether the two regions should be merged. This threshold is fixed to ensure that a region isolated inside a bigger one is merged.

The third type of oversegmentation involves the merging of two neighbouring regions whose shape is clearly jagged. The shape measure is defmed as.

$$g = A/d^2$$

where:

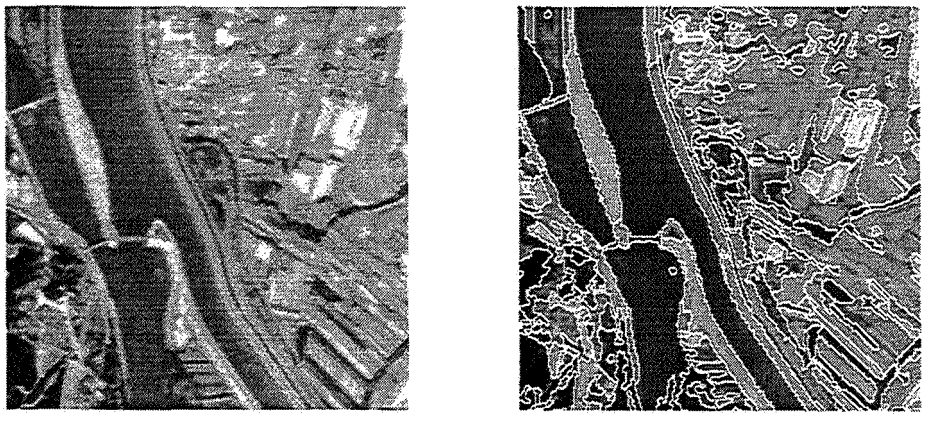
d is the average minimal distance of any point of the region to the boundary; and

A is the area of the region.

The distance is computed by the Chamfer algorithm (Borgefors, **1986)** which generates a distance image in a straightforward way with a minimal (and known) error This shape measurement has been reported as having a better discriminatory power for the more complex shapes compared to the other shape measurements, e.g. the ratio of the squared perimeter on the region area (Castleman, **1986).**

An example of a SPOT image that has been segmented is shown in figure **3.**

A similar segmentation algorithm has also been developed to segment SAR images. It works in much the same way, by region growing, but the difference is that seed points for the regions are determined using the characteristics of the image histogram. In addition, there is no subsequent merging of the regions.



SPOT panchromatic subscene | Segmented region edges on original image

Figure **3:** *An* example of SPOT segmentation

Assessment of the "registrability" of extracted features

Given that registration processing is an extremely expensive task (in terms of complexity and time), it would be useful to find a way to reduce the cost of this step. This can be done by reducing the number of features extracted from the image that are more likely to produce false matches. The matching of erroneous regions is computationally very expensive because of the error propagation and especialiy because of the correction difficulty Correcting a registration mistake requires:

- identification of the mistake (which usually involves very high level knowledge); and
- correction of it (which is a difficult task given that the registration is often based on a global optimisation on the whole data set).

To resolve the problem of erroneous matches, a registrability assessment is under development which filters out features which are likely to be mismatched. It is assumed that a real world object correctly identified in an image has a greater chance of being correctly matched. Therefore if the extracted features are compared with a topographic database it should be possible to decide which ones represent actual features on the ground. The main criteria used in this filtering process are shape of the feature and grey level variance within the feature. In the topographic data set, the shape may be expressed as the number of right angles in the vectorized regions, whereas in the segmented image the shape can be expressed as the amount of "rectilinearity" in the edges that form a region. Therefore, to assess the reliability of the extracted features, four steps are required

1. registration the image is registered manually to the topographic data.
2. distance image computation: done by the Chamfer algorithm, this gives the distance to the nearest object for every pixel of the image.
3. segmentation using the segmentation algorithm described above.
4. validation this step involves the vectorization of the boundaries of the segmented regions, their projection onto the distance image and the decision whether or not a region (or a part of it) has been correctly identified.

This procedure, shown in figure 4, was tested using a SPOT panchromatic image and a corresponding topographic data set. Initial results of the registrability assessment are shown in figure 5.

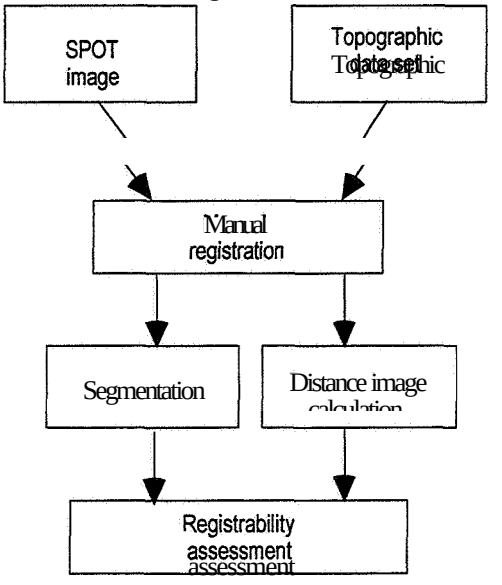


Figure 4 Reliability assessment

In figure 5, the first image shows the boundaries of the regions after segmentation overlaid on the original scene. The second one shows the set of regions after filtering based on variance and geometry criteria. One can see that many ambiguous segmented edges in the forest or urban areas have been removed and that the remaining regions are rather well identified.

To improve the registrability assessment several geometric measurements will be added to help assess the accuracy of a region. We are especially considering the use of fractal measurements as a way to distinguish between natural objects and man-made ones, the latter being easier to match. Another interesting point would be to study the influence of the nature of the object on the matching. We intend to demonstrate that certain objects, such as fields, are the objects that give the best segmentation results and,

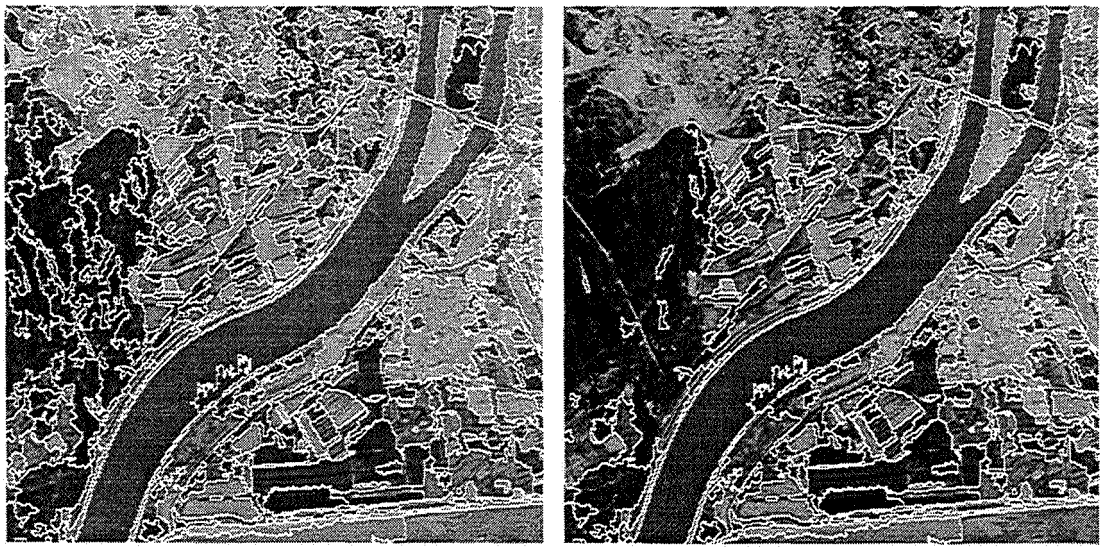


Figure 5: Example of registrability assessment filtering

in future, to use this knowledge to chose the set of objects to be matched.

IV. DISCUSSION

This paper has introduced two projects with similar goals: to develop an end user system for the automatic registration of multisensor/ multisource data sets. The PAIRS system has been shown to successfully register any pair of overlapping SPOT images fully automatically, irrespective of the content of the images. However, since this system was only a prototype system, it does have its drawbacks. For example, there is no correction for terrain, and no account is made for cloud cover Therefore, following on from the success of the PAIRS project, the ARCHANGEL project was initiated. This project is far more wide-ranging than its predecessor, and aims to achieve far greater goals. These include:

- detection and elimination of cloud cover in the images before feature extraction takes place;
- correction for terrain effects, where terrain models are available; and
- introduction of topographic data.

Although some of the PAIRS and ARCHANGEL algorithms are similar, ARCHANGEL endeavours at being more robust and efficient, To achieve this, new concepts have been implemented, such as a registrability assessment for analysing the suitability of patches for matching, before time is wasted on matching erroneous patches.

This paper has described the development of the registrability assessment to date, but more work is still required on this particular aspect of the ARCHANGEL project. When fully complete, it is hoped that the registrability assessment will be able to effectively eliminate all matching primitives (whether they are points or lines) before the matching takes place. This will save an enormous amount of time spent on correcting registration parameters for erroneous matches. In future, it is envisaged that the registrability assessment will be incorporated into an iterative procedure which not only removes ambiguous matching entities, but also ensures there are enough match points across the whole image to ensure an accurate registration. This paper has also described the other main area of research for the ARCHANGEL project at University College London which is concerned with feature extraction based on image segmentation.

The other main components of the project are being researched at other establishments across Europe University of Stuttgart has developed tools for extracting significant features from topographic databases which can be used as matching entities. University of Stockholm has assembled a set of tools for detecting linear features and line intersections in different types of imagery. University of Oporto has been researching an improved push broom scanner camera model which will require less ground control. This has been tested with SPOT data. Some work has already been carried out on integrating all of the algorithms into one architecture, but this will not be completed until the project reaches its final stages, when the validation will also take place.

When the project is finished, the result will be a fully tested and validated end user software package which will be able to automatically register images to images and images to topographic data. Although at this stage of the project it may be difficult to evaluate the robustness of the final product, if the result is similar to the PAIRS project, it should be highly successful.

V. REFERENCES

Abbasi-Dezfouli, M. and Freeman, T., "Patch matching in stereo images based on shape," International Archives of Photogrammetry and Remote Sensing, Vol. 30(3), pp. 1-8, 1994.

Borgefors, G., "Distance Transformations in Arbitrary Dimensions", Comput. Vis., Graphics and Im. Proc. 27, pp. 321-371, 1986.

Castleman, K. R., Digital Image Processing, Prentice Hall International Editions, 1996.

Dare, P M. and Dowman, I. J., "An automated procedure for registering SAR and optical imagery based on feature matching," In Microwave Sensing and Synthetic Aperture Radar, G. Franceschetti et al, Eds., SPIE Proc. Vol. 2958, pp. 140-151, 1996.

Dowman I., Boardman, C. and Newton W., "A system for automatic registration of two SPOT images", Remote Sensing Science and Industry, Proc. of 22nd Annual Conference of the Remote Sensing Society, pp. 169-175, 1996a.

Dowman, I J., Morgado, A. and Vohra, V., "Automatic registration of images with maps using polygonal features," International Archives of Photogrammetry and Remote Sensing; 31(3), pp. 139-145, 1996b.

Duplaquet, M.-L. and Cubero-Castan, E., "Updating cartographic models by SPOT images interpretation", Proc. SPIE vol. 2315, pp. 299-307, Image and Signal Processing for Remote Sensing, Jacky Desachy; Ed., 1994.

Fonseca, L. and Munjunath, B., "Registration techniques for multisensor remotely sensed imagery," Photogrammetric Engineering and Remote Sensing 62(9) pp. 1049-1056, 1996.

Forstner, W and Giilch, E., "A fast operator for detection and precise location of distinct points, comers and centres of circular features," ISPRS Intercommission Workshop on Fast Processing of Photogrammetric Data, pp. 281-305, Interlaken 2-4 June 1987

Haala, B., Hahn, M. and Schmidt, D., "Quality and performance analysis of automatic relative orientation," Proc. SPIE, vol. 1944, pp. 140-152, 1993.

Newton, W., Gurney, C., Sloggett, D. and Dowman, I., "An approach to the automated identification of forests and forest change in remotely sensed images," International Archives of Photogrammetry and Remote Sensing, 30(3), pp. 607-614, 1994.

Ruskoné, R. and Dowman, I., "Segmentation design for an automatic multisource registration," Proc. SPIE vol. 3072, pp. 307-317, 1997

Toutin, T., "Intégration des données multisource comparison de méthodes géométriques et radiométriques," International Journal of Remote Sensing 16(5), pp. 2795-2811, 1995.

Session IV

Correlation Methods

Automated and Robust Image Geometry Measurement Techniques with Application to Meteorological Satellite Imaging

James L. Carr
Carr Astronautics Corp.
1509 Corcoran St. NW
Washington, DC 20009
l00022.3043@compuserve.com

Marc Mangolini and Bertrand Pourcelot
Aerospatiale
BP 99 100 Bd. du Midi
F-06322 Cannes La Bocca Cedex
France

Andrew W. Baucom
Swales Aerospace
5050 Powder Mill Rd.
Beltsville, MD 20785

ABSTRACT

Algorithms for measuring the absolute geometric position of landmarks and independently the stability of their positions are presented. For these purposes, landmarks are small image neighborhoods containing land-water boundaries. Absolute geometric position is measured by matching the edges in the landmark neighborhood with a vector-graphic representation of the land-water boundary derived from a digital cartographic database, such as the World Vector Shoreline (WVS). Landmark position stability is measured independently using a cross-correlation between landmark neighborhoods from successive images. Both algorithms provide measurements accurate to a fraction of a pixel. These two algorithms have found application to meteorological image processing systems. In the United States, they have been applied in the GOES Landmark Analysis System (GLAS), and in Europe, they are being applied in the Image Quality Ground Support Equipment (IQGSE) to be used for Meteosat Second Generation (MSG). The IQGSE implements automatic cloud cover detection to provide added robustness against cloud-cover artifacts.

Introduction

Image geometric quality, comprising "absolute accuracy", "coregistration" and "relative accuracy", is important for users of meteorological satellite images. Absolute accuracy is the accuracy of the knowledge of the relationship between image coordinates (row, column) and geographic coordinates (latitude, longitude). Absolute accuracy is important for localizing severe weather. Coregistration, referring to the relative alignment between spectral bands, is important for meteorological data products derived from multiple spectral bands. Relative accuracy, referring to the geometric stability from image to image, is particularly important for determination of wind fields derived from cloud displacements observed over several consecutive images.

This paper describes algorithms for measuring image geometric quality that have been implemented in three different systems. These algorithms are suitable for both visible (VIS) and atmospheric window-channel infrared (IR) satellite imagery A cross-correlation technique is used for relative error measurement, and edge matching with a map is used for absolute error and coregistration

measurements. The first system is the Aerospatiale Meteosat Image Processing System (AMPS), supporting the Meteosat Operational Program (MOP)' Taking advantage of the AMPS heritage, Aerospatiale and Carr Astronautics are developing the Image Quality Ground Support Equipment (IQGSE) for the European Space Agency (ESA) and EUMETSAT[2] The IQGSE is a state of the art computer system for qualifying Meteosat Second Generation (MSG) geometric image quality prior to launch, and for verifying the MSG geometric image quality, in orbit, during the MSG commissioning phase. The Image Rectification Software (IRS) of the IQGSE is responsible for rectifying MSG images in real-time, and for performing concurrently, the processing of up to 1000 landmarks in the eight window channels every 15 minutes. Such a large throughput requires a fully automatic landmark processing. An important feature of which is automatic cloud cover detection to be used for determining which landmarks are not obscured by cloud cover, and for masking clouds in partially cloudy scenes.

For the GOES program, Carr Astronautics and Swales Aerospace have developed the GOES Landmark Analysis System (GLAS). The GLAS software works with GOES rebroadcast GVAR telemetry This allows GLAS to perform processing on the same GOES data that is broadcast to the end-users. This form of interface to the GOES data allows for independent landmark analysis anywhere that GVAR telemetry is available. In order for GLAS to use the GVAR telemetry, a daemon called GVARD splits the raw GOES data stream into individual files for the VIS, IR2, IR4 and IR5 channels. These "flat bitmap" channel files are easily read and manipulated by GLAS. GVARD and GLAS have been designed to run in a cooperative, autonomous manner. When GVARD has successfully processed an image, it triggers GLAS to begm landmark processmg. When GLAS has completed its landmark analysis, it waits for the next image. The GVARD program senses when GLAS has finished processing and can then orderly purge disk files before moving on to the next image. Both GLAS and GVARD function in a user-driven manual mode, as well as, an automatic mode requiring no human intervention. Both applications employ a simple, user-friendly "point and click" interface.

Measurement of Absolute Error

The absolute geometric position of a landmark found in an image neighborhood can be determined by comparing that neighborhood with a map. To perform this measurement automatically requires some sort of matching algorithm, of which cross-correlation is but one example[3]'' With cross-correlation, a two-dimensional correlation surface $\rho(\Delta x,\Delta y)$ = (neighborhood)**(chip) is formed using a smaller geolocated image called a ''chip'' The vector shift $(\Delta x,\Delta y)$ of the chip with respect to the neighborhood for which $\rho(\Delta x,\Delta y)$ is maximized is the measurement of geometric ''error''. More generally, the relationship between the chip and the neighborhood could involve rotation, stretch, shear, or high-frequency intra-image distortions; however, in most imaging systems, these may be ignored locally if the neighborhood size is a small fraction of the total image size.

The chip could be synthesized from a map, with its radiometric levels assigned thematically' Without significant modeling effort and detailed knowledge of local conditions, such a chip might be a poor representation of the scene radiometric content. Alternatively, the chip could be obtained from previous imagery and retained in a chip library In this latter approach, there must be at least one comparison against a map (whether performed automatically or manually) to geolocate the chip- Any error in the initial geolocation becomes a systematic error that is reintroduced with each use of the chip. Moreover, for meteorological applications, one finds tremendous diversity in scene radiometry, suggesting that a chip library needs to represent the landmark under many different conditions.

Motivated by these considerations, AMIPS, GLAS, and the IQGSE all use an algorithm that compares only neighborhood edge features with a map. This is desirable since a map represents boundaries well and radiometry poorly The landmark features that are used contain land-water boundaries, for which the World Vector Shoreline (WVS)[6] is a suitable digital cartographic database. Such features are

adapted to both visible and IR imagery, due to the typically large differences in albedo and temperature existing between land and water

Figure 1 shows the measurement process as it is implemented in GLAS. The first steps are to remap the database shorelines using nominal imaging geometry, and to create an edge image by applying a Sobel operator[7] and a contrast stretch. The geometric error is measured by shifting the shorelines relative to the edge image, and computing a similarity measure $\zeta(\Delta x,\Delta y)$ by summing the edge values along the shifted shoreline. Since shoreline coordinate pairs $(x_n+\Delta x, y_n+\Delta y)$ are generally pairs of floating-point numbers, an interpolation is performed in the edge image using the nearest neighbors at each referenced site. The geometric error, $(\Delta x,\Delta y)$ such that $\zeta(\Delta x,\Delta y)$ is maximized, is found with subpixel precision by the method of successive bisection in two-dimensions[8] A normalization for the peak similarity measure places it within the range 0 to 1, indicating goodness-of-fit. Figure 2 shows an example of how the algorithm behaves with a GOES-8 visible scene.

Figure 1. GLAS Absolute Error Measurement Process

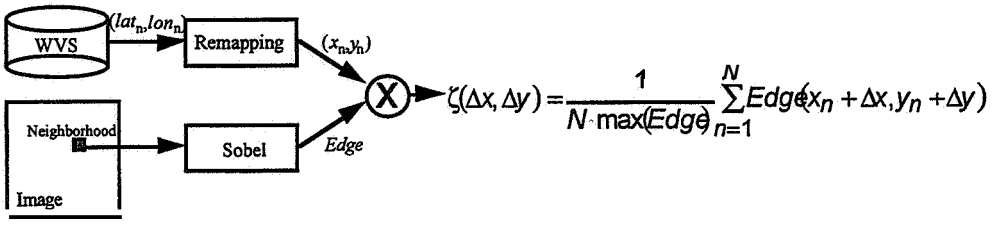
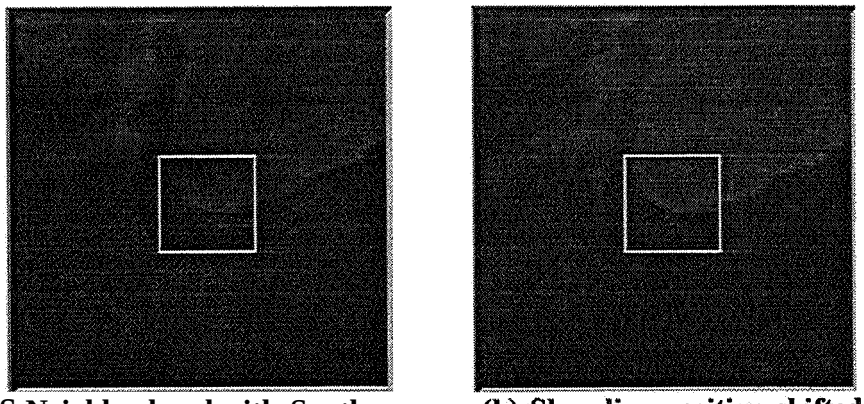
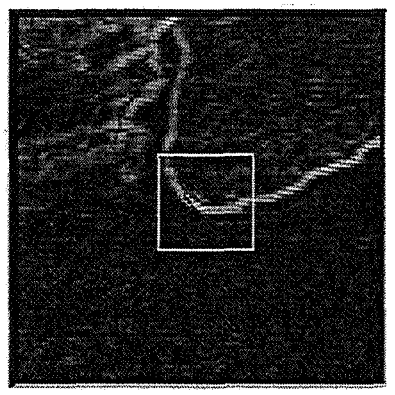


Figure 2. Edge Image and Matching from GLAS

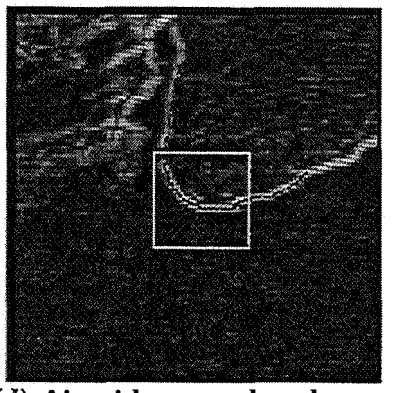


(a) VIS Neighborhood with Southern best tip of Baja peninsula and shoreline at its nominal position.

(b) Shoreline position shifted for match to image shoreline. The shift measures the geometric error.



(c) Algorithm applies a Sobel edge detection operator. Shoreline and cloud edges are apparent.



(d) Aigorithm matches the map shorelines to the edge image within the box shown.

To demonstrate that the algorithm is capable of subpixel accuracy, a **3-km** resolution **VIS** image was created from a 1-km resolution GOES VIS image with **known** random fractional pixel (but an integer number of GOES pixels) offsets in each direction. Anti-aliasing filtering was performed before sampling to the 3-km scale, simulating the image acquisition process at the larger scale. The measurement errors observed over a set of Monte-Carlo trails, where the simulated image noise-to-signal ratio was varied, are summarized in Table 1

This simulation study would seem to indicate that a goodness-of-fit value above **0.4** would be a discriminator indicating whether the algorithm has obtained a meaningful result; however, in practice, it is found that this is not a robust test, and if the threshold is elevated, too many valid measurements are discarded. Robust automatic operation requires either prescreening for cloud cover, good illumination and contrast, or statistical data editing using a model for the expected system behavior For the GLAS application, the latter approach **has** been emphasized, whereas, radiometric cloud cover detection will be implemented in the IQGSE. Figure 3 shows an edited GLAS data set for GOES-8 produced from a sequence of 5 images during about 1 hour. The data are not plotted with respect to time, but rather with respect to number in chronological order, The apparent discontinuities in the absolute error (Figure 3a) occur at image boundaries, and the profiles of the errors within an image reflect both temporal and spatial dependence. The data editing model invoked assumes that the coregistration between spectral bands is constant. Measurements of coregistration are obtained by differencing the absolute errors for each spectral channel pair-wise at each landmark site. Figures 3b-3d shows these measurements after application of recursive n-σ outlier editing. **The** best measurement is that between the channels **IR4** and **IR5** which are spectrally similar, so that scene related systematic errors cancel.

Table 1. Accuracy Simulation Study Results

Noise/Signal	EW RMS Error (Pixels)	NS RMS Error (Pixels)	Goodness-Of-Fit (GOF)
0	0.04	0.08	>0.64
0.1	0.06	0.11	>0.57
0.25	0.09	0.15	>0.44
0.5	0.15	0.21	>0.30
1	Algorithm Fails		<0.38

Besides errors due to radiometric noise, other errors may be present in the measurements related to relief and illumination, temperature gradients across shallow water and coastal zones, tides, and map errors.

Measurement of Relative Error

In geostationary satellite meteorology, landmark neighborhoods are reacquired frequently enough that, for clear-sky conditions, the radiometric content changes slowly from image to image. This permits the cross-correlation algorithm to measure accurately the image-to-image relative error (stability of the absolute error). In many cases, this direct measurement will be a more accurate estimate of relative error than that obtained by simply differencing absolute errors. In the **AMIPS, GLAS,** and **IQGSE** implementations, a chip (C) extracted and saved from a previous landmark neighborhood, is correlated against the current landmark neighborhood (N), forming a normalized correlation surface sampled at integer pixel shifts:

$$\rho(\Delta x, \Delta y) = \frac{(N - mean(N)) ** (C - mean(C))}{\sqrt{var(N)var(C)}}$$

Under the premise that we are sampling a continuous correlation surface, the 3x3 neighborhood of the peak is modeled as a biquadratic surface for the purpose of interpolating to find the relative error with subpixel precision:

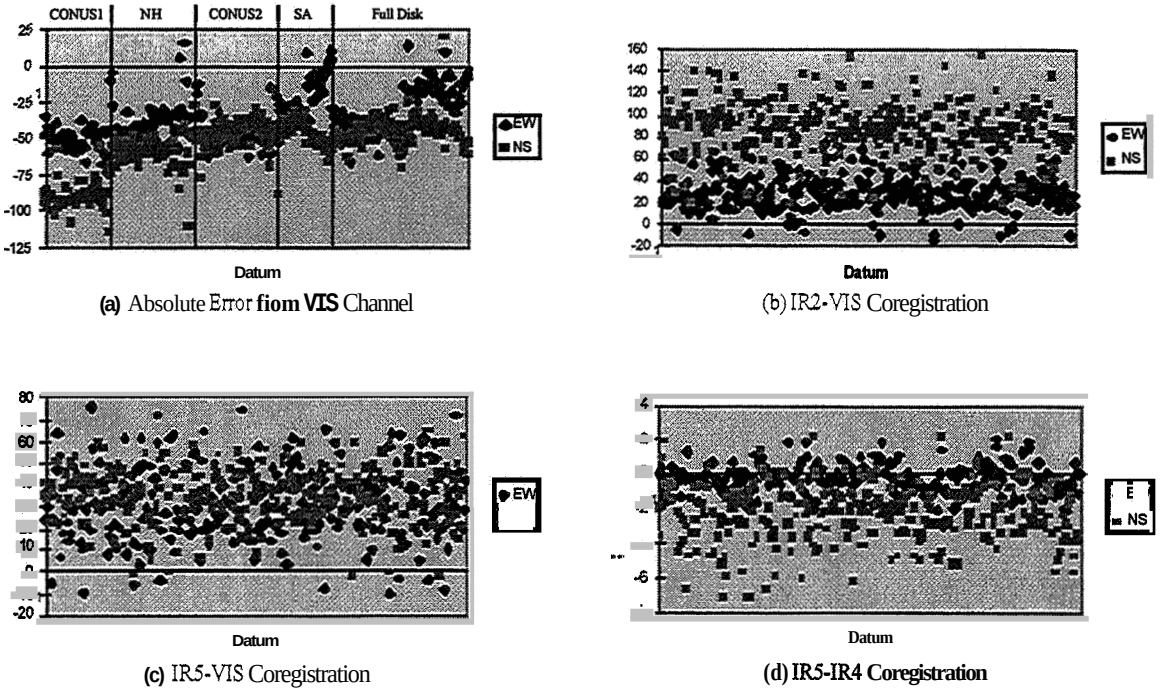
$$\rho(\Delta x, \Delta y) = \rho_0 - \frac{1}{2}(\Delta x - \Delta x_0 \quad \Delta y - \Delta y_0) \cdot A \begin{pmatrix} \Delta x - \Delta x_0 \\ \Delta y - \Delta y_0 \end{pmatrix}$$

The position of the interpolated peak $(\Delta x_0, \Delta y_0)$ gives the interpolated relative error, the peak value (ρ_0, typically >0.95) indicates the quality of the correlation, and a symmetric positive definite matrix **(A)** describes the curvature of the correlation surface near its peak. The matrix **A** can be used to construct a measurement error ellipse, with axes given by its eigenvectors, each normalized in proportion to the inverse of the square-root of its eigenvalue. **A** similar procedure can be used to construct error ellipses for the absolute error measurement.

Figure **4** shows the relative errors measured by GLAS simultaneously with the data set of Figure 3 The relative error at a landmark site is always with respect to the most recent previous acquisition of that landmark.

Using **SPOT** images to simulate 1 **km** resolution VIS scenes shows that this algorithm is capable of better than 0.1 pixel accuracy, which is in agreement with studies conducted for the Landsat program using a similar algorithm[9]

Figure 3. GLAS Noon Data Set for GOES-8
(Continental U.S.-Northern Hemisphere-Continental U.S.-South America-Full Disk)



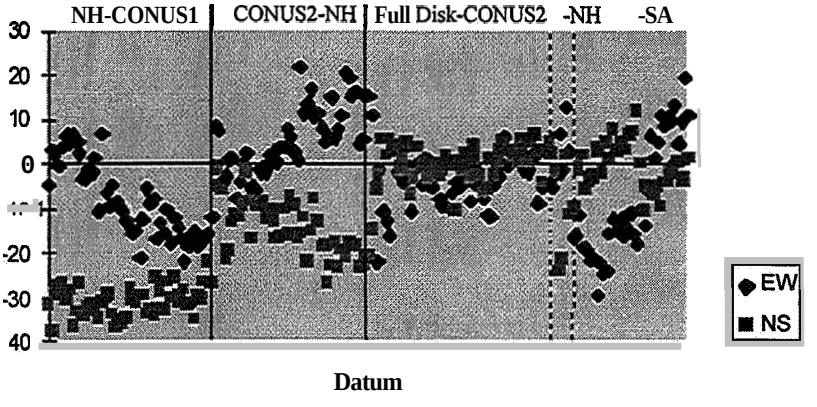
(a) Absolute Error fiom VIS Channel

(b) IR2-VIS Coregistration

(c) IR5-VIS Coregistration

(d) IR5-IR4 Coregistration

Figure 4. GLAS VIS Relative Error Data Set



Integration with Automatic Cloud Detection

Fully automatic landmark processing must deal with the problem of cloud coverage to insure robust automatic operation. In the **IQGSE,** automatic cloud detection will supplement screening based on goodness-of-fit and correlation coefficient. In practice, some particular cloud configurations within the landmark neighborhood lead to acceptable goodness-of-fit and correlation values but erroneous

displacement measurements, both absolute and relative. This situation frequently occurs with coastal clouds, over which cloud edges are mistaken for the shorelines due to their similar morphologies.

An automatic cloud detection scheme has been designed for the IQGSE to process each landmark neighborhood, creating a cloud mask indicating cloud-contaminated pixels, as well as information about the type of detected cloud for each pixel of the neighborhood. This cloud detection scheme has been adapted and tested using the NOAA-AVHRR, which is a polar orbiting sensor featuring five channels with a spatial resolution of 1 1 km. The IQGSE will work with the MSG-SEVIRI instrument which includes the AVHRR channels plus seven other channels with a spatial resolution of 3 km (except for the visible broadband that provides a 1 km resolution). The SEVIRI will allow more robust cloud detection performance not only because of its richer spectral information, but also because of its geostationary platform. successive images are acquired under the same viewing conditions with a very short repeat-cycle period (15 minutes for MSG), and they can be compared on a pixel basis to detect changes in the cloud cover

Concrete tests using visible and thermal IR images can be developed for a specific sensor using heuristics for cloud detection that are generic to any sensor with both visible and thermal IR channels.

- Clouds are colder (thermal IR channels) and brighter (visible channels) than the clear-sky background,
- Cloudy scenes have a higher spatial variability than clear-sky scenes (for kilometric resolutions),
- Cloudy scenes have a higher temporal variability than clear-sky scenes

These are the basic assumptions on which the cloud detection scheme relies, but many particular cases need to be accounted for to avoid detection errors: some land surface types can present a higher spatial variability than clouds (urban areas, mountains), or a lower temperature than clouds (ice, snow), or high reflectance (sunglint areas, deserts). Furthermore, persistent cloud conditions will not be detected by means of a temporal variability test. These restrictions show that a robust cloud detection scheme needs to be based upon a combination of complementary tests, rather than on a single criterion.

Cloud detection tests can be separated in two categories: temporal analysis that detects moving or developing clouds from a sequence of images, and multispectral analysis which identifies clouds by means of their multispectral signatures. During daytime, both visible and infrared channels can be used temporal analysis detects a new cloudy pixel by the conjunction of its visible reflectance increase and its brightness temperature decrease. During nighttime, only the thermal infrared channels can be exploited. Since processing is performed on a pixel basis, an essential prerequisite for temporal analysis is that the two successive images shall be accurately registered. Similarly, the spectral channels which are compared pixel-to-pixel in the multispectral analysis need to be co-registered.

Cloud detection tests developed for the IQGSE make use of the MSG atmospheric window channels: HRV (broadband), VIS0.6, VIS0.8, NIR1.6, IR3.8, IR8.7, IR10.8, IR12.0. The NIR1.6 channel is used in a pre-processing test dedicated to the discrimination between snow/ice backgrounds and clouds. Since the snow/ice response in this channel is much lower than the cloud response, as compared to visible channels'', a Normalized Difference Snow Index (NDSI) is defined for this purpose:

NDSI= (VIS0.6 - NIR1.6) / (VIS0.6 + NIR1.6)

All other window channels are exploited by the following tests[11,12,13,14]

Temporal analysis

[IR10.8(t-Δt) - **IR10.8(t)**] > Thresh_IR (day and night)

[VIS0.8(t) - **VIS0.8(t-Δt)**] > Thresh–VIS (day only)

where VIS0.8 denotes the reflectance in the **0.8** μm channel and IR10.8 the brightness temperature in the 10.8 μm channel. Thresh_IR and Thresh–VIS are background dependant (land or sea).

Multispectral analysis

- Cold cloud test
  IR10.8 < Thresh_cold(land/sea)
- Bright cloud test
  VIS0.6 > Thresh–bright–land over land
  VIS0.8 > Thresh–bright–sea over sea
- Reflectance ratio test
  Thresh_ratio_low < VIS0.8NIS0.6 < Thresh–ratio–high
- High cloud test
  IR10.8 - IR12.0 > Thresh–high–cloud
- Cirrus test
  IR8.7 - IR10.8 > Thresh_cirrus
- Thin cirrus test
  IR3.8 - IR12.0 > Thresh–thin
- Low cloud & fog test
  IR10.8 - IR3.8 > Thresh–fog
- Thermal uniformity test
  Max{IR10.8}$_{3x3\ array}$ - Array_center(IR10.8) > Thresh_tut(land/sea)
- Reflectance uniformity test
  Array–center(VIS0.8) - **Min{VIS0.8}$_{3x3\ array}$** > Thresh–rut–sea

It is very difficult to define fixed and global thresholds for cloud detection tests because of their sensitivity to many parameters: scene illumination, climate, season, atmosphere constituents, *etc.* Two solutions have been envisaged to determine thresholds adapted to the area of interest. The first one is to use ancillary databases containing geographical and climatological information. Geographical

databases provide a landsea mask as well as environmental data (*e.g.*, the Olson map of ecosystems[15]). Climatological databases provide global surface temperature models that help predict clear-sky brightness temperatures, which are of prime importance for cloud detection. The second solution is to estimate dynamic thresholds from the data themselves, by making assumptions on their spatial or temporal distributions with and without clouds.

Figure 5 presents the results of multispectral analysis applied to an AVHRR landmark neighborhood located on the West coast of Saudi Arabia during daytime, with a partial cloud cover over both land and sea. The mosaic is composed of 15 windows representing the source images and the cloud detection results. The five windows in the first row correspond to the five AVHRR channels (resp. 0.63, 0.87, 3 74, 10.8, 12.0 μm). The middle window in the left column is the land/sea mask, where land is coded as light grey and sea as black. The bottom left window is the fmal cloud mask, defined as the union of all elementary cloud masks. Finally, the eight remaining windows show the elementary cloud masks generated by the multispectral tests. The fog & low cloud mask and the thin cirrus cloud mask are blank because these tests cannot be used during daytime. All other tests detected some of the cloudy pixels, and the union mask is consistent with what an operator would manually provide. One can notice that no clouds were detected along the shoreline by the ratio test, the reflectance uniformity test, or the thermal uniformity test. These tests are suppressed adjacent to the shoreline as a precautionary measure to avoid false alarms caused by individual mixed land/sea pixels and the natural nonuniformity in a mixed set of land and sea pixels. The high cloud test detects the temperature difference measured by the two long-wave IR channels. It is well adapted to finding semi-tranparent clouds and cloud edges, but it does not respond well to the very coldest cloud tops in a completely opaque cloudy pixel. This latter case is easily detected by the complementary cold cloud test.

Figure 5. Cloud Detection - next page

The generated cloud mask is used by the IQGSE landmark processing module. Based on the fraction of cloudy pixels in the cloud mask, a landmark cloud status is derived to indicate whether the landmark is acceptable or not for further processing. If the landmark is acceptable, the absolute and relative geometric error measurements will be performed using the cloud mask to mask cloudy pixels from the Sobel edge detection and the cross-correlation computation.

Conclusion

The landmark processing algorithms presented in this paper have been targeted for meteorological satellite applications, but should in principle be adaptable to other types of applications. Theoretical studies and practical experience have shown that they are capable of achieving subpixel accuracy Robust automatic operation requires that some allowance for cloud cover be taken. This can either be in the form of statistical post-processing to identify data violating a model of system behavior (*e.g.*, stable coregistration), or in the form of automatic cloud cover recognition algorithms. The cloud cover recognition problem is complex, and the IQGSE scheme relies on a suite of complementary algorithms to achieve a high degree of reliability The cloud cover detection algorithms selected for the IQGSE are based on heuristics that can be tuned and adapted to other sensor systems possessing visible and thermal IR channels.

Channel 1
Channel 2
Channel 3
Channel 4
Channel 5
Land/sea mask
Bright clouds
Ratio test
Cold clouds
High clouds
Final cloud mask
Reflect. uniformity
Thermal uniformity
Fog & low clouds
Thin cirrus clouds

3

[1] Blancke, B., J. L. Carr, E. Lairy, F Pomport, B. Pourcelot "The AEROSPATIALE METEOSAT Image Processing System", First International Symposium AAAF, Scientific Imagery and Image Processing, Cannes, France, April 4-6, 1995

[2] Blancke, B., J. L. Carr, M Mangolini, F Pomport, B. Pourcelot, D. Rombaut, "The MSG Image Quality Ground Support Equipment", The 1997 EUMETSAT Meteorological Satellite Data Users' Conference, Brussels, Belgium, September, 1997

[3] Bemstein, R., C. Colby, S. W Murphrey, J P Snyder, "Image Geometry and Rectification" in ***Manual of Remote Sensing,*** R. N. Colwell (ed.), 2-nd Ed., American Society of Photogrammetry, 1983.

[4] Eppler, W., D. Paglieroni, M. Louie, J Hanson, "GOES Landmark Positioning System", in GOES-8 ***and Beyond,*** E. R. Washwell (ed.), SPIE, Denver, CO, 1996. Discusses the merits of several algorithms for the satellite meteorological application.

[5] Adamson, J., G. W Kerr and G. H. P Jacobs, "Rectification Quality Assessment of METEOSAT Images", ESA Journal, **12,** 1988,467-482.

[6] Available on CD-ROM fiom the National Geophysical Data Center, **NOAA, U.S.** Dept. of Commerce, 325 Broadway, Boulder CO 80303-3328. Military Specification MIL-W-89012.

[7] Niblack, W., ***An Introduction to Digital Image Processing,*** Prentice-Hall International, 1986, 74.

[8] Press, W H., W T Vetterling, S. A Teukolsky, B. P Flannery, ***Numerical Recipes,*** 2-nd Ed., Cambridge, 1992, 397

[9] Bernstein, R., ***et. al., op. cit.,*** 1983.

[10] Bunting J T., R.P d'Entremont, "Improved cloud detection utilizing Defense Meteorological Satellite" Program near infrared measurements. Air Force Geophysics Laboratory report AFGL-TR-82-0027, 1982, 91

"Gustafson G.B., R.G. Isaacs, R.P d'Entremont, J.M. Sparrow, T.M. Hamill, C. Grassoti, D.W Johnson, C.P Sarkisian, D.C Peduzzi, B.T Pearson, V.D.Jjakabhazy, J.S. Belfiore, A.S. Lisa, "Support of environmental requirements for cloud analysis and archive (SERCAA): algorithm descriptions", Scientific report No. 2, ***Atmospheric and Environmental Research,*** Inc. 840 Memorial Drive, Cambridge, MA 02139, 1994.

[12] Saunders R.W., K.T Kriebel, "An improved method for detecting clear sky and cloudy radiances fi-om AVHRR data", ***International*** Journal ***of Remote Sensing,*** vol. 9, No. 1, 123-150, 1988.

[13] Stowe L.L., E.P McClain, R. Carey, P Pellegrino, G.G. Gutman, P Davis, C. Long, ***S.*** Hart, "Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data", Advances ***in Space Research,*** vol. 11, No. 3, 1991, 351-354.

[14] Strabala K.I., S.A. Ackerman, W.P Menzel, "Cloud properties inferred from 8-12 μm data", ***Journal of Applied Meteorology,*** vol. 33, No. 2, 1994, 212-229.

[15] Olson, J. ***S.,*** "Earth's Vegetation and Atmospheric Carbon Dioxide", in, ***Carbon Dioxide Review. 1982,*** W C Clark (ed.), Oxford University Press, New York, 1982,388-398;
Olson, J. ***S.,*** J A. Watts, and L.J. Allison, "Carbon in Live Vegetation of Major Ecosystems", Report ORNL-5862, ***Oak*** Ridge National Laboratory, *Oak* Ridge, Tennessee, 1983

Techniques for multiresolution image registration in the presence of occlusions

Morgan McGuire
Harold S. Stone
morganm@mit.edu,hstone@research.nj.nec.com
NEC Research Institute
4 Independence **Way**
Princeton, NJ **08540**

Abstract

This paper describes techniques for registering two images with respect to each other when one or both are partially occluded To achieve high-speed registration, we propose to register low-resolution versions of the images, and then progressively increase the resolution to obtain more **ac**curate registration. In earlier work Stone and Shamoon showed how to use occlusion masks to remove the effects of clouds on full-resolution images. Invalid pixels do not participate in the registration computation. Stone extended the occlusion masks ***to*** low-resolution wavelet representations, continuing the use of binary valued masks at all resolution levels. **In** this paper we examine fractional-valued occlusion masks for coarse pixels whose component **fine** pixels include some that are invalid. This captures information that is lost when binary masks are used.

Tests on partially occluded satellite images show that the registration algorithm with binary occlusion masks produces accurate registrations for cases that cannot be registered with a maskless version of the same algorithm. Fractional masks produce sharper registration peaks than binary masks produce because the information retained tends to reduce false matches at the expense of slightly lowering the correlation peak at the correct registration position. At higher resolutions, the difference between the ***two*** masked-based algorithms diminishes, and in the limit of **full** resolution the fractional and binary masks versions are identical.

1 Introduction

Image-registration algorithms attempt to align two images of the same scene so that the pixels of one image fall on the corresponding pixels of the other image. One of the primary **uses** of image registration is to identify how the two images differ, which can be done by subtracting one from the other after they have been registered with respect to ***each*** other This is a key idea underlying the use of satellite images for environmental studies and the use of sequences of radiology images to study the progression of a medical pathology.

Various image-registration techniques have been used successfully for registering cloud-free (unoccluded) satellite images. The techniques tend to fall into three main areas:

1. feature-based registration with image warping [7, 8, 9],
2. correlation-based algorithms in the pixel domain or in the frequency domain [2, 19, 4, 5, 10], and
3. phase-correlation techniques in the frequency domain [1, 11, 14].

All of these techniques tend to fail in the presence of occlusions because they do not distinguish between a valid pixel and an occluded pixel. They rely on matches produced by comparing valid pixels in one image to corresponding valid pixels in the other image. The quality of these matches must be sufficiently high to compensate for discrepancies produced from the comparisons of valid pixels in one image to invalid **ones** in the other image, or to comparisons of invalid pixels in both images.

To overcome the problems of occlusions, Ching, Toh, and Er [6] proposed a scheme that incorporates occlusion detection and registration and bears some similarity with the general approach of this paper Their scheme calculates a correlation based **on** all data in an image, and then selectively removes regions from the correlation when such regions are poorly correlated with the pattern template. Another technique called ***expansion matching*** **[3]** attempts to restore signals in regions where the signals are occluded.

We describe a new technique based on an idea by Stone and Shamoon [18] and extended by Stone [15]. The common thread in all of these is to use occlusion masks to identify the invalid pixels. With these masks, the registration algorithms produce registration criteria that depend only on comparisons of valid pixels to valid pixels. For

high-speed registration, the registration algorithms operate on low-resolution versions of the images, and progressively improve the accuracy of the registration by moving to higher resolution representations. Wavelet representations for this purpose were explored by Le Moigne [12] and by Le Moigne, Campbell, and Cromp [13] without occlusion masks. A factor of N reduction in resolution speeds up the correlation operations by a factor of N^2 Stone [15] incorporates inclusion masking into this framework.

This paper explores the effectiveness of occlusion masks for low-resolution image registration. The original algorithm of [15] adopts a very conservative criterion for handling partially obscured coarse pixels. A coarse pixel is declared invalid if any of its constituent fine pixels is invalid. This decision correctly eliminates the effects of invalid data, but it also ignores the effects of valid data, which possibly could be used to improve the accuracy of the registration. To capture the valid data, we show how to extend the binary occlusion masks to fractional masks that weigh partially occluded pixels in proportion to the number of their valid components. The results of image-registration tests show that occlusion masking is very effective for registering partially occluded images. At low resolution, fractional masks produce sharper registration peaks than do binary masks. However, the peaks at the correct registration points produced by fractional masks may not be as high those produced with binary masks, so it is possible for a fractional-mask mask algorithm to miss a correct registration that is discovered with a binary-mask algorithm.

The next section of the paper reviews binary occlusion masks. Section 3 extends occlusion masks to fractional masks for wavelets. Section 4 shows the results of several image-registration operations that compare fractional masks, binary masks, and unmasked algorithms. The final section contains a summary and suggestions for future research.

2 Occlusion Masks

The idea of fractional occlusional masks evolved from a natural progression of ideas based on masking and correlations. We review briefly the correlation operation and registration criteria based on correlations and occlusion masks.

For the review we work with one-dimensional vectors, and note that the extension to two-dimensional images is straightforward. Let $\mathbf{x} = (x_0, x_1, \ldots, x_{N-1})$ and $\mathbf{y} = (y_0, y_1, \ldots, y_{M-1})$ be two vectors and let M be small compared to N We assume that vector $\mathbf{y}$ has been generated by extracting M contiguous components of $\mathbf{x}$ starting at position j, and adding a small amount of noise to the result. To register $\mathbf{y}$ with s in one dimension, one seeks the position j in $\mathbf{x}$ for which the M successive components of $\mathbf{x}$ at that point are most like $\mathbf{y}$

The registration computation appears to require $O(NM)$ operations because $O(M)$ operations are required to check each position, and there are $O(N)$ positions to check By reducing comparison operations to circular correlations, we can do the computations in the frequency domain at a cost of $O(N \log N)$ operations, which is considerably lower when $\log N$ is small compared to M

The *circular correlation* of $\mathbf{x}$ and $\mathbf{y}$, written as $\mathbf{x} \odot \mathbf{y}$, is the N vector whose jth component is

$$(\mathbf{x} \odot \mathbf{y})_j = \sum_{i=0}^{M-1} x_{(i+j) \bmod N}\, y_i, \quad \text{for } j = 0, 1, \ldots, N-1 \tag{1}$$

Note that the subscripts are taken modulo N, and thus the correlation is end-around or circular This holds in all derivations that follow, so that we do not explicitly show the modulo N dependence on the subscripts from this point on.

The Convolution Theorem produces a fast correlation algorithm in the following way, Extend y to length N by appending $N - M$ zeros. Let $\mathbf{F}$ be an $N \times N$ Fourier transform matrix and let $\hat{\mathbf{F}}$ be the complex conjugate of $\mathbf{F}$ The correlation form of the Convolution Theorem states:

$$\mathbf{x} \odot \mathbf{y} = \mathbf{F}^{-1}\left(\mathbf{F}\mathbf{x} * \hat{\mathbf{F}}\mathbf{y}\right) \tag{2}$$

where the operation $*$ denotes point-by-point multiplication. The operations on the right-hand side are two Fourier transforms, a point-by-point multiplication of length N, and an inverse Fourier transform. The multiplication takes only N complex multiplies, and the Fourier transforms require $O(N \log N)$ operations so that the right hand side of Eq. (2) requires only $O(N \log N)$ operations overall as compared to $O(NM)$ operations to evaluate Eq. (1) directly.

Correlation is a core computation for various criteria that can be used to measure the "closeness" of two images for registration purposes. Among the possible criteria are the sum of square differences and the normalized correlation coefficient Both of these can be written as operations on circular correlations which lead to computational efficiencies. Stone and Li [17] made the following observation.

Consider the sum of square differences criterion. In the pixel domain this is the function

$$\min_{j \in \{0,\ldots,N-M+1\}} \sum_{i=0}^{M-1} (x_{i+j} - y_i)^2$$

$$= \min_{j \in \{0,\ldots,N-M+1\}} \sum_{i=0}^{M-1} x_{i+j}^2 - 2x_{i+j}y_i + y_i^2 \tag{3}$$

The middle term of Eq. (3) is $-2\mathbf{x} \odot \mathbf{y}$ and the last term is a constant. The first term reduces to a circular correlation of a vector of squares of $\mathbf{x}$ and a mask for $\mathbf{y}$ Specifically, let h be a pattern *mask* for $\mathbf{y}$ with 1s in the first M components where y has valid values and 0s in the next $N - M$ components where the extension of $\mathbf{y}$ has 0s. Let $\mathbf{x}^{(2)}$ be the vector $(x_0^2, x_1^2, \ldots, x_{N-1}^2)$. Then the first term is $\mathbf{x}^{(2)} \odot \mathbf{h}$ and the sum of squares criterion becomes

$$\min_{j \in \{0,\ldots,N-M+1\}} (\mathbf{x}^{(2)} \odot \mathbf{h})_j - 2(\mathbf{x} \odot \mathbf{y})_j + \sum_{i=0}^{M-1} y_i^2 \tag{4}$$

$$\min_j \frac{1}{(\mathbf{m} \odot \mathbf{h})_j} \left((\widetilde{\mathbf{x}^{(2)}} \odot \mathbf{h})_j - 2(\tilde{\mathbf{x}} \odot \tilde{\mathbf{y}})_j + (\mathbf{m} \odot \widetilde{\mathbf{y}^{(2)}})_j \right), j = 0, 1, \quad , \mathbf{N} - \mathbf{M} + \mathbf{1} \tag{5}$$

$$\mathbf{C}(\mathbf{x},\mathbf{y})_j = \frac{\sum_{i=0}^{M-1} x_{j+k} y_i - \left(\frac{1}{M}\right) \left(\sum_{i=0}^{M-1} x_{j+i}\right) \left(\sum_{i=0}^{M-1} y_i\right)}{\sqrt{\left(\sum_{i=0}^{M-1} x_{j+i}^2 - \left(\frac{1}{M}\right) \left(\sum_{i=0}^{M-1} x_{j+i}\right)^2\right) \left(\sum_{i=0}^{L-1} y_i^2 - \left(\frac{1}{M}\right) \left(\sum_{i=0}^{M-1} y_i\right)^2\right)}} \tag{6}$$

for $j = 0, 1, \quad , N - M + 1$.

$$\mathbf{C}(\mathbf{x},\mathbf{y})_j = \frac{(\tilde{\mathbf{x}} \odot \tilde{\mathbf{y}})_j - \left(\frac{1}{(\mathbf{m} \odot \mathbf{h})_j}\right) (\tilde{\mathbf{x}} \odot \mathbf{h})_j (\mathbf{m} \odot \tilde{\mathbf{y}})_j}{\sqrt{\left(\left(\widetilde{\mathbf{x}^{(2)}} \odot \mathbf{h}\right)_j - \left(\frac{1}{(\mathbf{m} \odot \mathbf{h})_j}\right) (\tilde{\mathbf{x}} \odot \mathbf{h})_j^2 \right) \left(\left(\mathbf{m} \odot \widetilde{\mathbf{y}^{(2)}}\right)_j - \left(\frac{1}{(\mathbf{m} \odot \mathbf{h})_j}\right) (\mathbf{m} \odot \tilde{\mathbf{y}})_j^2 \right)}} \tag{7}$$

In Eq. (4) the vector $\mathbf{y}$ acts as if it were actually of length $\mathbf{M}$ with its last $\mathbf{N} - M$ components occluded by a one-dimensional cloud. The extension to occlusion masking of both $\mathbf{x}$ and y requires the x vector to have a corresponding mask denoted as $\mathbf{m}$. In this way, both x and y can have clouds occluding some of their pixels. The summation in Eq. (3) for this situation must be over just those terms for which both the pixel x_{i+j} and the pixel y_i are valid The number of such terms varies with the relative positions of x and y in the circular correlations.

In modifying Eq. (4) to produce an equation that allows both x and y to be occluded, we use masked vectors $\tilde{\mathbf{x}} = \mathbf{x} * \mathbf{m}$ and $\tilde{\mathbf{y}} = \mathbf{y} * \mathbf{h}$, which have 0s where their respective occlusion masks indicate that the positions are invalid [18]. Also, let $\widetilde{\mathbf{x}^{(2)}}$ denote $\mathbf{x}^{(2)} * \mathbf{m}$ The location of invalid pixels is arbitrary for both x and y, but we still assume that the valid pixels of y are among the first M components of the extension of y to length N

The criterion for the sum of square differences for occlusion masking is given by Eq. (5). The coefficient $1/(\mathbf{m} \odot \mathbf{h})$ normalizes the criterion because for each j its value is the number of valid products in the summation. If for some j there are no valid terms in the summation, we define the value of criterion to be infinite.

Another popular criterion for image registration is the normalized correlation coefficient defined by Eq. (6).

Stone [16] shows that this reduces to the form shown in Eq. (7) The important point is that all vector operations reduce to correlations, and thus can be done efficiently in the Fourier domain. The normalized correlation coefficient is ideally suited for registration in the case that two images differ by f f i e transformations of the intensity map. That is, intensity u in x maps to intensity $\alpha u + \beta$ in $\mathbf{y}$ for some constants α and β. This type of transformation occurs when two images differ only in lighting intensity. If the source of the lighting is from a different direction, the image differences cannot be modeled by affine translations of the intensity map alone, and the correlation coefficient is less useful for registration

This completes the review of occlusion masking. The next section shows the extension of occlusion masking to fractional occlusion masks.

3 Fractional Occlusion Masks

A major goal of image registration algorithms is speed. Recall from Section 2 that registration operations in the pixel domain have a complexity that grows as $O(N^2)$ and as $O(\mathbf{N} \log \mathbf{N})$ in the Fourier domain for images with $\mathbf{N}$ pixels. Individual computations are quite time consuming for large images, and some applications require myriad image registrations.

To reduce the cost of a search, one can perform correlations on reduced-resolution versions of the images, and progressively refine the registration process by using local searches on higher resolution versions of the images [12]. Occlusion masks work with low-resolution images in the same way that they work with high resolution images. That is, each coarse pixel has an associated mask pixel. A mask pixel of 1 designates the coarse pixel as valid and a mask pixel of 0 designates the coarse pixel as invalid. Stone [15] suggests that the mask pixels be created by making a mask pixel equal to 1 if all of the corresponding fine pixels are valid and otherwise the mask pixel is 0. The sum of square differences and normalized correlation coefficients can then be computed according to Eqs. (5) and Eqs. (7) Note that these equations rely on the use of $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ whose invalid coarse pixels are 0, while those of $\mathbf{x}$ and y are not necessarily 0 because the invalid coarse pixels may contain a mixture of valid and invalid fine pixels.

An interesting problem arises at low resolution as illustrated in Fig. 1. Fig. 1 shows the same image at three resolutions. A program processed the original image for clouds, and marked pixels as invalid depending on reflectivity, temperature, and the intensity of pixels in the neighborhood. Invalid pixels are colored deep black in the figure. Notice how the relative size of the black region greatly enlarges as resolution diminishes. This occurs because the scattered clouds that cause the occlusions tend to touch a large fraction of coarse pixels, whereas they touch only a small fraction of fine pixels. Clearly, image registration operations based on the low resolution image

in Fig. 1 will be using only a small fraction of the information available for registration.

The decision to invalidate a coarse pixel if any constituent fine pixel is invalid is a conservative decision that eliminates the contribution of noise in an invalid coarse pixel. This guarantees that the correlation is based only on data known to be free of occlusion contamination, and will generally produce a higher correlation peak in the correct position than will be produced by an algorithm that uses noisy data in the invalid pixel positions. However, it also is more likely to produce high correlation peaks at false positions because it uses fewer data points in the correlation.

We seek a means to obtain higher discrimination by using more data in the correlations than is used in the scheme described in Section 2. We propose to use fractional masks in a way that makes use of the good data in the partially occluded coarse pixels.

In wavelet analysis, each *coarse* pixel is a weighted sum of a set of fine pixels. For the Haar wavelet, for example, each coarse pixel is the average of four pixels from the next higher level of resolution. Let Level 0 be the highest resolution level, and consider how to generate the pixels and mask for Level 1 from the pixels and mask for Level 0. Clearly if all fine pixels are valid, the coarse pixel at level 1 should be the average of the fine pixels, and the mask at level 1 should be 1. Likewise, if all fine pixels are invalid, the coarse pixel should be 0 and its corresponding mask should be 0. What should be the values in other cases?

First we examine the calculation of the mask coefficients, and then derive a formula for the wavelet coefficients. The fractional mask approximation stems from the fact that the mask acts as a measure of area in Eqs. (5) and (7). Consider, for example, the term $\widetilde{\mathbf{x}^{(2)}} \odot \mathbf{h}$ in Eq. (5) The h mask is the mask for y. A 1 in a mask position of h indicates that the corresponding pixel of y is valid. This 1 incorporates a pixel from $\widetilde{\mathbf{x}^{(2)}}$ into the correlation for each cyclic shift of h. For each cyclic shift of h, the mask as a whole selects an area of $\widetilde{\mathbf{x}^{(2)}}$ to participate in a Correlation sum, because this area aligns with valid pixels of y in the selected shift position.

If a pixel of y is a partially occluded coarse pixel, its actual area depends on the he-pixel occlusion structure. Specifically, the valid area of a coarse pixel is the fraction of its full-resolution pixels that are unoccluded. We can calculate this value recursively by the following process. Let m be a coarse-mask coefficient and $n_i, i = 0, 1, \ldots, R-1$ be the fine-mask coefficients one level of resolution finer. Then we compute m by the equation

$$m = (1/R) \sum_{i=0}^{i=R-1} n_i \tag{8}$$

As an example, consider a two-dimensional Haar transform. The fine-resolution level contains mask coefficients that are all 1s and 0s. After applying Eq. (8) to the next level of resolution, we obtain masks whose values are multiples of 1/4. The multiple reflects the number of unoccluded fine pixels among its components. At the second level, Eq. (8) produces values that are multiples of 1/16. The multiple again indicates the number of Level 0 unoccluded pixels in each coarse pixel.

Now we treat the computation of the wavelet coefficients. Consider a partially occluded image at full resolution and examine the wavelet representation at the next lower level of resolution. Each coarse pixel should have a value that represents the value of its fine pixel components. The coarse pixel value that best represents the unoccluded components is a weighted average of the unoccluded components. To find the weights to use, consider the wavelet equation for the coarse component. Let the equation for the coarse pixel q be the filter polynomial

$$q = \sum_{i=0}^{i=R-1} a_i p_i \tag{9}$$

where the coefficients of the R-term filter are a_i and the values of the fine pixels are p_i. Let n_i be the fractional mask values for pixels p_i. The following equation has the properties that we seek for q.

$$\begin{aligned} q &= \left(\frac{\sum_{i=0}^{i=R-1} a_i}{\sum_{i=0}^{i=R-1} n_i a_i} \right) \sum_{i=1}^{i=R-1} n_i a_i p_i, \\ &\quad \text{if } \sum_{i=0}^{i=R-1} n_i a_i \neq 0 \\ &= 0, \text{ otherwise} \end{aligned} \tag{10}$$

The filter terms that contribute are those for which the fine components are unoccluded. The multiplier $(\sum_i a_i)/(\sum_i n_i a_i)$ normalizes the filter polynomial so that it occupies the full dynamic range of the original polynomial in the absence of the occluded terms. If there are no occluded pixels in the summation, Eq. (10) produces the same answer as Eq. (9). To compute wavelet coefficients at various levels, apply Eq. (10) recursively.

As an example, consider a two-dimensional Haar transform. In this case, the coefficients a_i are equal to 1/4, and R = 4. Eq. (10) reduces to the average of the unoccluded components at the first-level, and for each finer level of resolution is a weighted average where the weights are the mask coefficients multiplied by a normalizing factor.

The effect of using the fractional masks is illustrated in Fig. 2. Figs. 2(a-b) correspond to Figs. 1(b-c). Note the difference in the black areas. The effects of scattered clouds has been diminished greatly in the visible image. Because the masks are fractional, many of the visible pixels in Fig. 2 have mask pixels less than unity to weigh them in proportion to their information content. This sharpens the correlation peak as we show in the example in the next section.

4 Example

In this section we look at the use of the fractional masking technique for image registration of partially occluded images. The image pattern appears in Fig. 3. It is the rectangular area of an image of the western coast of South

Africa. The coastline is quite irregular with a structure that lends itself well to image registration

A sample data base of nine images of the same area taken over a nine-day period appears in Fig. 4. These images are taken over a period of eight days, and are among the most cloudfree of that region in a three-month period. Even these images have substantial cloud coverage.

The cloud-detection algorithm places narrow clouds along the northwest coastline in all images. This is the coastline in the registration pattern in Fig. 3. Coastlines usually make excellent features for registration, but in this set of nine images, the coastline itself is of virtually no use because the cloud-detection algorithm believes that it is occluded in this image set.

Fig. 5 shows the results of an image-registration search for the rectangular pattern of Fig. 3. The registration used Level 4 wavelets ($1/256th$ resolution). To the right of each image is a query result indicated by a small white rectangular area containing small regions of gray-scale data. Each pixel in the query result corresponds to a position of the pattern with respect to the associated image. That pixel is colored white if the match is poor, black if the match is excellent, and grey if the match is something in between. The data displays the value of the normalized correlation coefficient of Eq. (7) after the coefficient has been subjected to a threshold. The threshold between black and white is set to reveal the peaks of the correlation, and a common threshold applies to all of the results. The threshold is set to 0.71 for this study, which is low enough to display a peak at a correct registration point in all cases for which the correct peak is the highest peak. No peak is shown when the correct registration is not the highest peak.

The results shown in Fig. 5 are for a search with fractional occlusion masks. The highest peak in the query result is indicated by a circle, and the corresponding rectangular position is marked on the associated image.

Fig. 6 shows comparable results for a search with binary occlusion masks. The results of a search with no compensation for occlusions appears in Fig. 7 For this nine-image data base, the binary mask and the fractional mask produced the highest correlation peak in the correct place in all cases. The uncompensated registration algorithm had the correct location in 8 of the 9 cases. However, there is a distinction in the quality of the correlations returned. The dark areas of Fig. 6 that are not close to the correct registration peak are false matches. Note that neither the fractional mask algorithm nor the uncompensated registration algorithm produce as many false matches as does the binary algorithm This is balanced in part by the darker and broader peak obtained for the binary-mask algorithm which is an indication that the correct correlation produces a higher peak for the binary-mask algorithm than for the other two.

To obtain some idea of the signal to noise ratio in the correlation coefficients, Figs. 8, 9 and 10 show plots of the correlation coefficient through the horizontal line that contains the correct peak correlation. It also gives the (x, y) coordinates for the peak correlation in *each* image. There is no obvious difference in the signal-to-noise ratio of Figs. 8 and 9. For some images, the binary mask has higher minor peaks and for some images fractional masking produces higher minor peaks. The main difference is that the peaks of Fig. 8 are sharper and lead to more accurate registration than do the slightly higher and broader peaks of Fig. 9. The case that produces the largest difference between the two algorithms is in Row 2, Column 3 where the peak correlation is 0.88 for the binary-mask algorithm and 0.83 for the fractional-mask algorithm.

In comparing the results in Fig. 10, the registration without occlusion masking, it is quite obvious that the correlation peaks are degraded by the occlusions so that peaks at incorrect positions may be higher than at the correct position. The overall signal-to-noise level is degraded as well because the correlation peaks are significantly lower and noise peaks are comparable with the peaks produced by the other *two* registration techniques. Fig. 10 shows an incorrect registration for the case in Row 2, Column 3 mentioned above. The correletion peak at the correct registration point in this case is only about 0.70 as compared to a peak at (1,1) of 0.73 for this sample image. Correlations at the correct registration point are 0.83 and 0.88 for the other two algorithms applied to this sample image. Clearly both of the mask-based algorithms give better discrimination than does the uncompensated algorithm.

5 Open Research

We have demonstrated that occlusion masks produce better search results than searches that ignore the effects of clouds. Moreover, the experimental results suggest that it may be better to retain partial information in low-resolution images than to discard that information to avoid errors caused by occlusions. This information is useful in producing a narrow registration peak. However, the results are not conclusive in that binary masks and fractional masks did not produce distinctly different results for the test data. There is a trade-off between the increased precision of the fractional-mask algorithm and the slightly better peak detection of the binary-mask algorithm

The idea of removing occlusions from images has applicability beyond the context here, which is correlation based. For example, feature-based searches can use occlusion information to make registration faster and more robust. If a region is known to be invalid, such algorithms would not attempt to find and identify features that lie within such regions.

Additional speedups and accuracy over binary masks may be possible by using the fractional masks with low-resolution transforms. Conservative binary masks increase the relative size of occluded regions as resolution is reduced, while fractional masks maintain the proportion of invalid to valid weighted data, as can be observed by visually comparing Fig. 1(c) to Fig. 3. At low resolutions, a binary mask can grow large enough to occlude all significant features. This problem does not occur when using fractional masks, so they enable some low-resolution searches to succeed, leading to faster search In practice, however, some relevant features may fall below the detection threshold in low-resolution images, making registration difficult, especially in the presence of noise from

other factors. The loss of accuracy diminishes the value of the faster processing available with low-resolution images. Because preprocessing with transformations like rotation and dilation also tend to increase the fraction of invalid pixels when using binary masks, fractional masks may increase the accuracy of searches of images sets that have undergone such transformations. It remains open as to whether this will have a significant effect on the search results.

We believe that there is no simple way to incorporate occlusion masking into phase-coherence algorithms. These algorithms rely on frequency domain representations of images. It is very costly to remove effects of invalid pixels in a frequency domain representation, which forces the algorithms to return to the pixel domain to remove those effects. However, once in the pixel domain, in order to remove an invalid pixel from the same position in two different images to make their transforms comparable, the images should be registered with respect to each other But the objective of the algorithm is to register the images, so that it does not seem feasible to prepare the images for registration without registering them first

Other approaches that may eliminate or reduce the effects of clouds may guess the values of occluded pixels [3]. Guesses may be fairly accurate when invalid pixels are near neighbors of visible pixels. Also, visible linear and edge features may be extended beneath clouds with high accuracy near the edge of the clouds and through narrow or scattered clouds.

Acknowledgments

The collection of images used in this paper has been provided by NASA through Dr J Le Moigne. Interested parties should contact the authors to obtain copies of the images. The authors thank Bo Tao of Princeton University and Louis Wang of Columbia University for their contributions to the program that produced the results in the paper

References

[1] Alliney, S., G. Cortelazzo, and G. A. Mian, "On the registrations of an object translating on a static background," *Pattern Recognition,* vol. 29, no. 1, pp. 131-141, January 1996.

[2] Anuta, P E., "Spatial registration of multispectral and multitemporal digital imagery using fast-Fourier Transform techniques," *IEEE Trans. on Geoscience Electronics,* vol. 8, no. 4, pp. 353-368, October, 1970.

[3] Ben-Arie, J., and K R. Rao, "On the recognition of occluded shapes and generic faces using multiple-templat expansion and matching,'' *Proc. IEEE 1993 CVPR,* New York, pp. 214-219, June 1993.

[4] Casasent, D., R. Schaefer, and R. Sturgill, "Optical hit-miss morphological transform," *Applied Optzcs,* vol. 31, no. 29, pp. 6255-6263, 10 Oct, 1992.

[5] Casasent, D., J S. Smokelin, and R. Schaefer, "Optical correlation filter fusion for object detection," *Optical Engineering,* vol. 33, no. 6, pp. 1757-1766, June 1994.

[6] Ching, W.-S., P.-S. Toh, and M.-H. Er, "Recognition of partially occluded object," *Proc, IEEE TENCON'93,* vol. 2, Computers, Communication, Control, and Power Engineering, Beijing, China, pp. 930-933, Oct. 1993.

[7] Cracknell, A. P., and K. Paithoonwattanakij, "Pixel and sub-pixel accuracy in geometrical correction of AVHRR imagery," *International Journal of Remote Sensing,* vol. 10, nos. 4 and 5, pp. 661-667, 1989.

[8] Devereux, B. J., R. M Fuller, L. Carter, and R. J. Parsell, "Geometric correction of airborne scanner imagery by matching Delaunay triangles," *International Journal of Remote Sensing,* vol. 11, nos. 12, pp. 2237-2251, 1990.

[9] Fiore, P D., "Image registration using both distance and angle information," *Proc. ICIP 1995,* IEEE, Washington DC, October, 1995.

[10] Khosravi, M., and R. Schaefer, "Template matching based on a grayscale hit-or-miss transform," *IEEE Trans. on Image Processing,* vol. 5, no. 6, pp. 1060-1066, June 1996.

[11] Kuglin, C. D, and D. C. Hines, "The phase correlation image alignment method." *IEEE 1975 Conference on Cybernetics and Society,* pp. 163-165, September, 1975.

[12] Le Moigne, J., "Parallel registration of multi-sensor remotely sensed imagery using wavelet coefficients," *Proc. SPIE O/E Aerospace Sensing, Wavelet Applications,* Orlando, pp. 432-443, April, 1994.

[13] Le Moigne, J., W J Campbell, and R. F Cromp, "An automated parallel image registration technique of multiple source remote sensing data," to appear in *IEEE Trans. on Geoscience and Remote Sensing,* 1998.

[14] Reddy, B. S., and B. N Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. on Image Processing,* vol. 3, no. 8, pp. 1266-1270, August 1996.

[15] Stone, H. S., "Fourier-Wavelet techniques in image searching," *Proceedings of 1997 IEEE International Symposium on Circuits and Systems,* Vol. 11, Circuit Theory and Power Systems, Communications and Multimedia, Hong Kong, pp. 1472-1475, June, 1997

[16] Stone, H. S., "Progressive wavelet correlation using Fourier methods," submitted to *IEEE Transactions on Signal Processing,* November, 1996

[17] Stone, H.S., and C. S. Li, "Image Matching by Means of Intensity and Texture Matching in the Fourier Domain," *Proceedings of SPIE Conference on Image and Video DataBases*, vol. 2670, Storage and Retrieval for Image and Video Databases IV, San Jose, CA, pp. 337-349, Feb., 1996

[18] Stone, H.S., and T Shamoon, "The use of image content to control image retrieval and image processing," to appear in Journal of Digital Libraries, 1998. An early version appeared in *Multimedia Computing, Proc. of the Sixth NEC Research Symposium*, SIAM, pp. 43-54, June, 1995.

[19] Wong, 3. W., and E. L. Hall, "Scene matching with invariant moments," *Computer Graphics and Image Processing*, vol. 8, pp. 16-24, 1978.

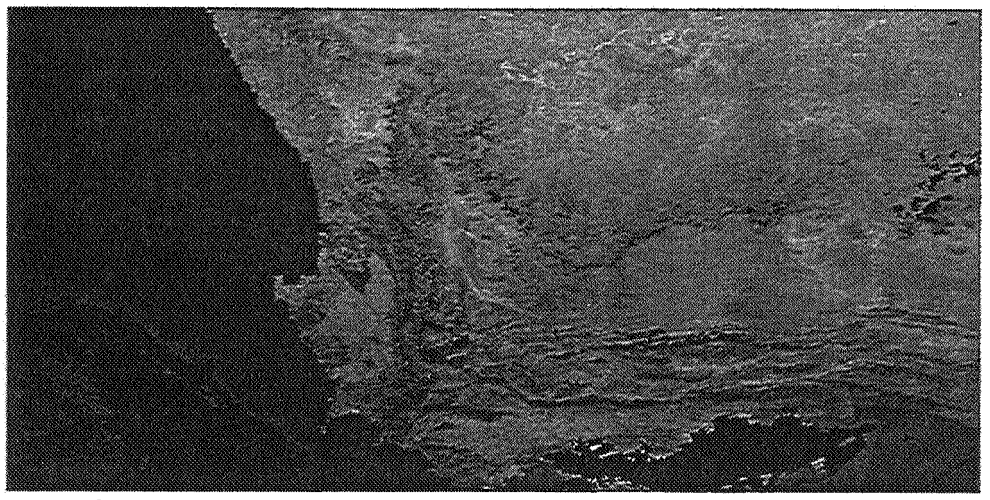


Fig. 1(a): Original image of South Africa, 512 by 1024 pixels, binary mask.

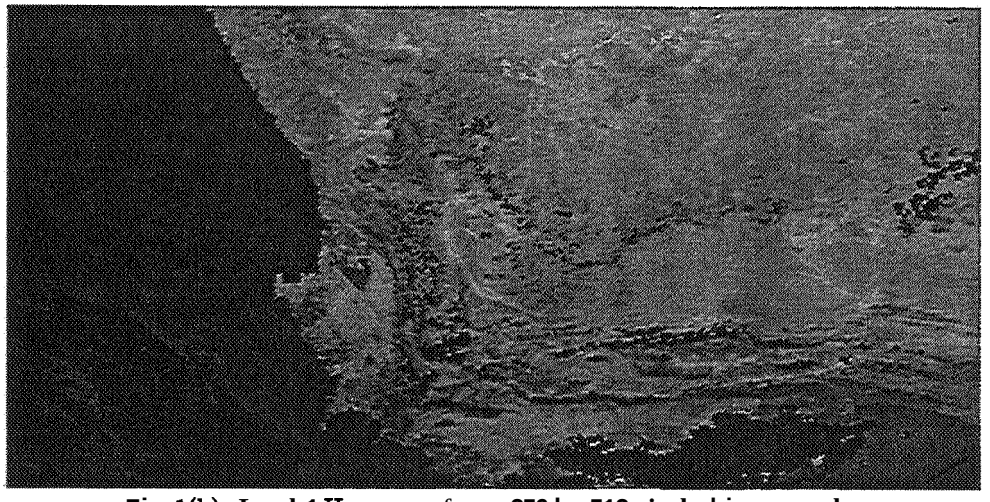


Fig. 1(b): Level 1 Haar transform, 256 by 512 pixels, binary mask.

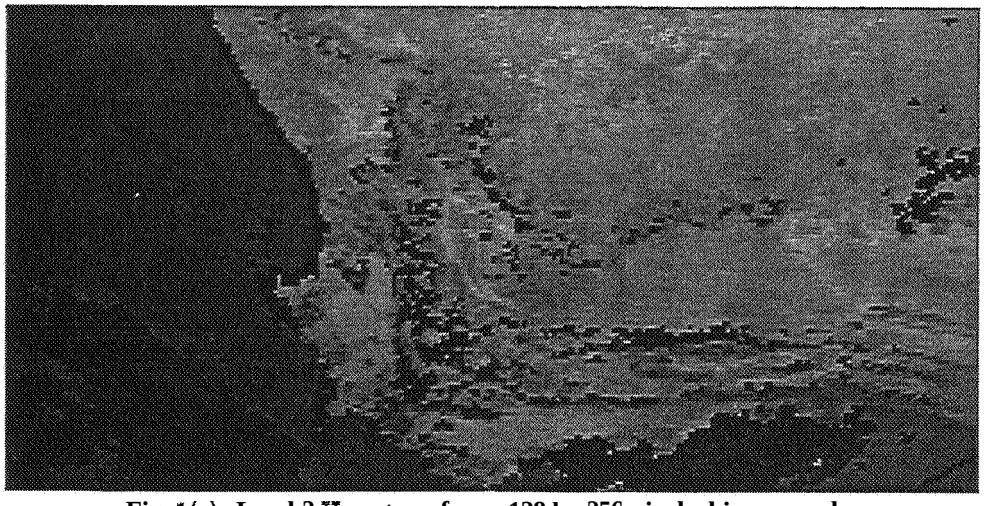


Fig. 1(c): Level 2 Haar transform, 128 by 256 pixels, binary mask,

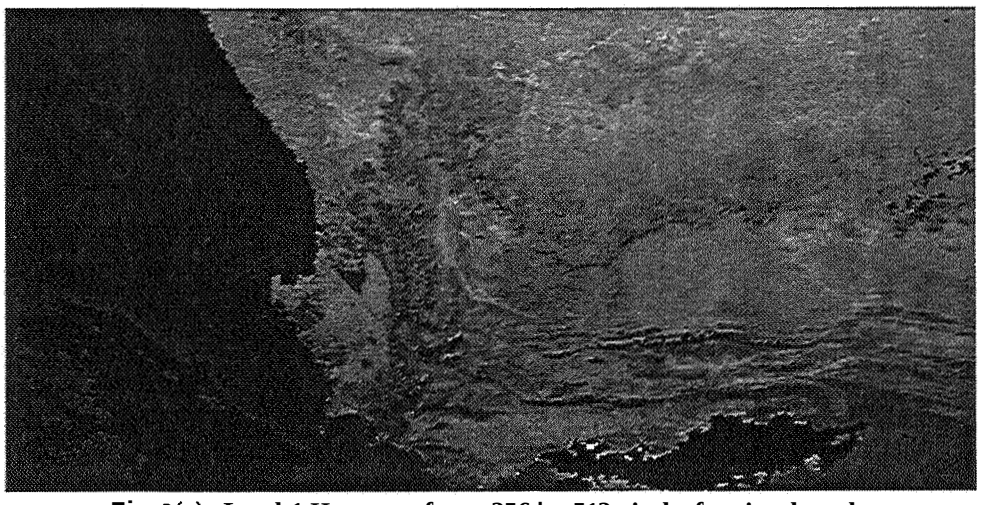


Fig. 2(a): Level 1 Haar transform, 256 by 512 pixels, fractional mask.

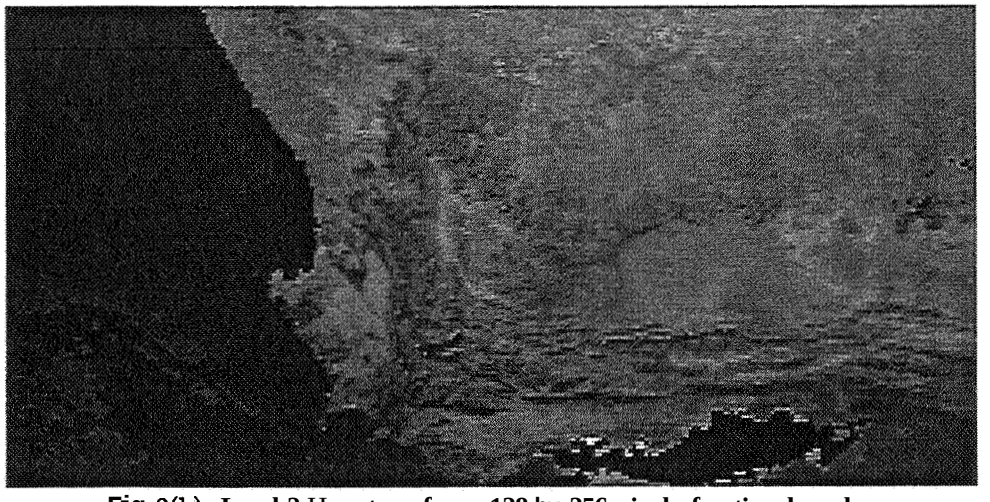


Fig. 2(b): Level 2 Haar transform, 128 by 256 pixels, fractional mask

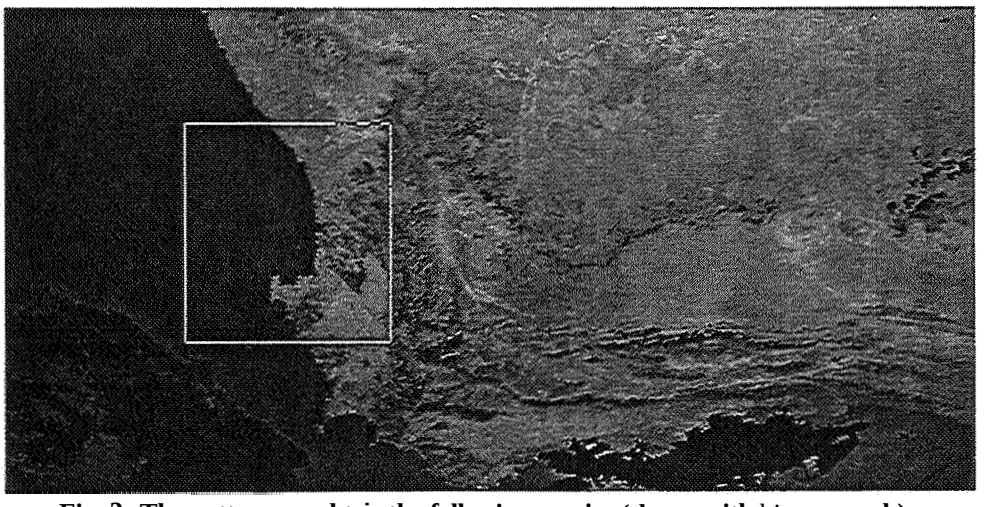


Fig. 3: The pattern sought in the following queries (shown with binary mask).

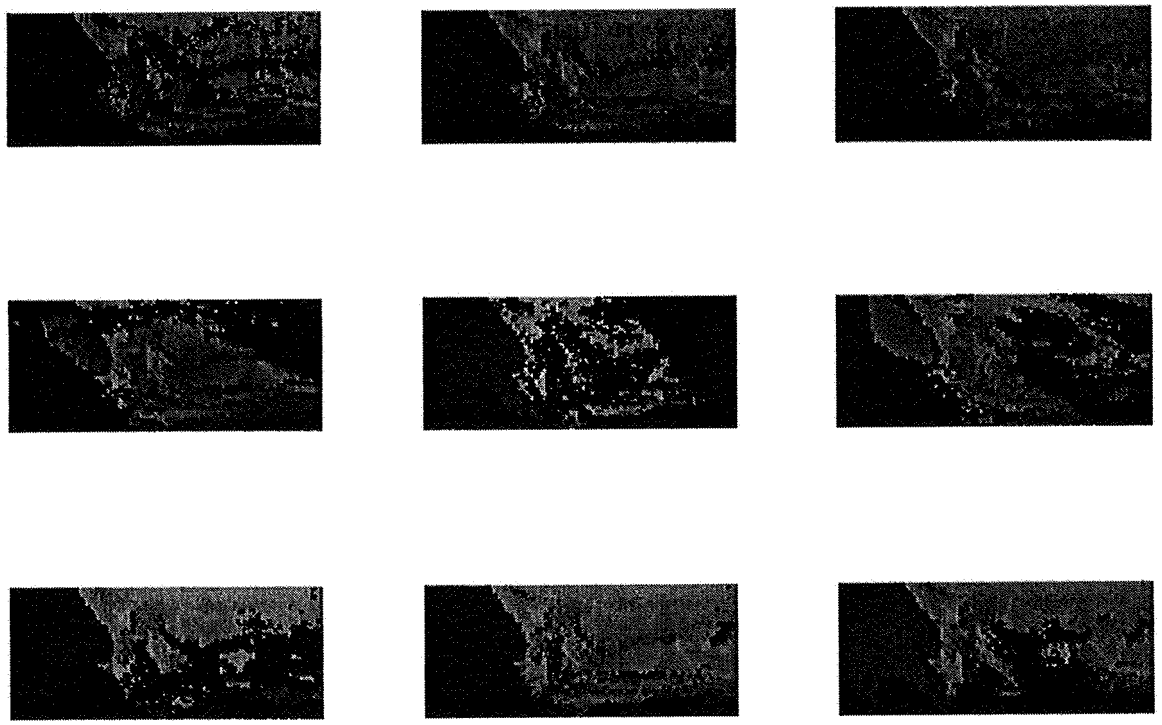


Fig. 4: Nine-image data base. Level 3 Haar transform, 64 by 128 pixels, binary mask.

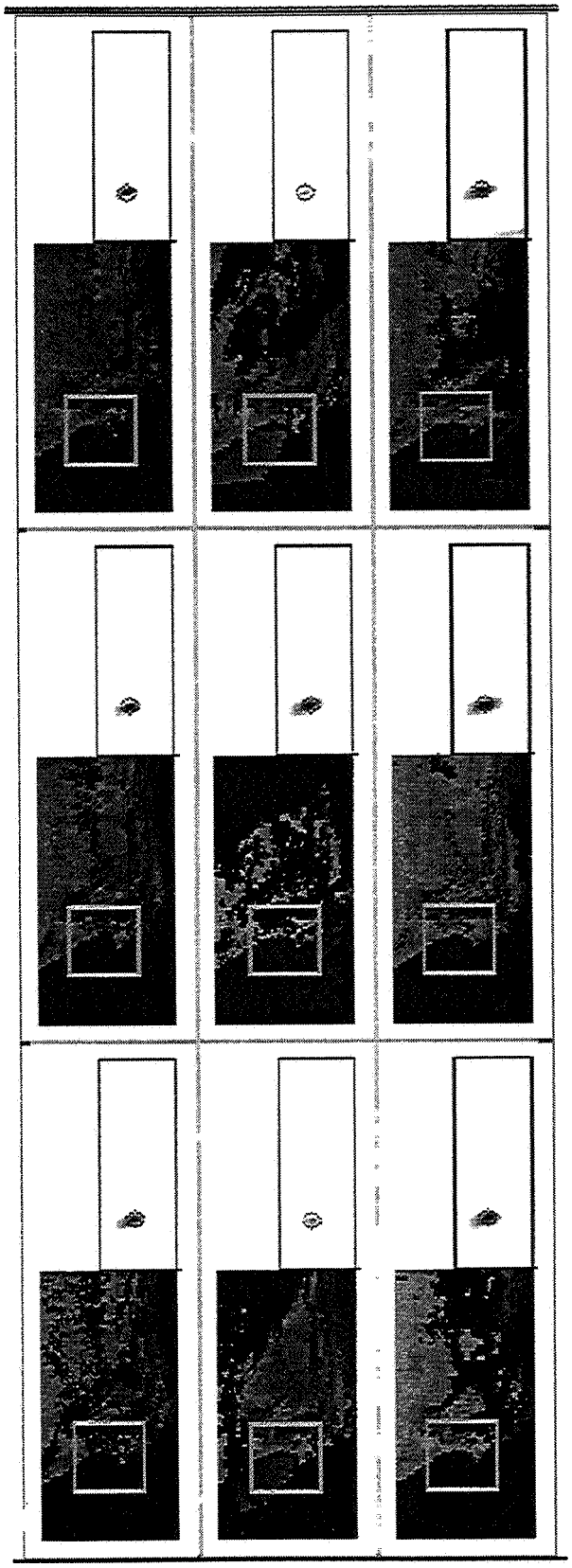


Fig. 5: Results of fractional-mask registration.

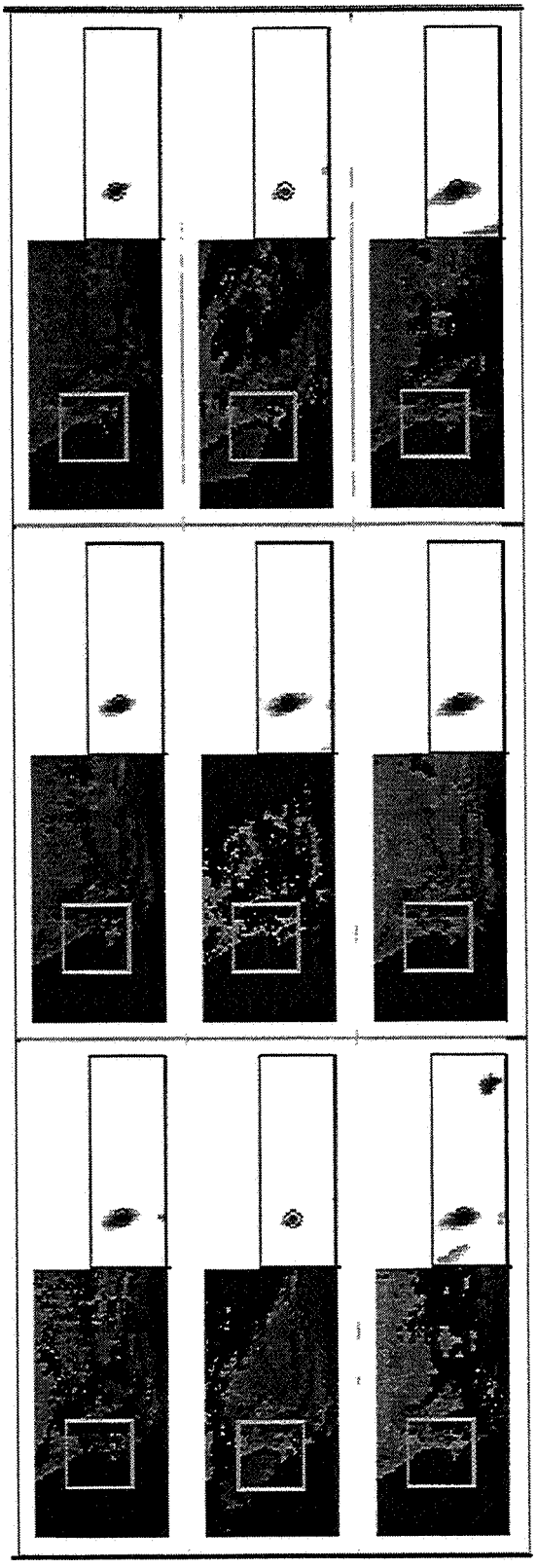


Fig. 6: Results of binary-mask registration.

Fig. 7: Query results for registration with no occlusion masks.

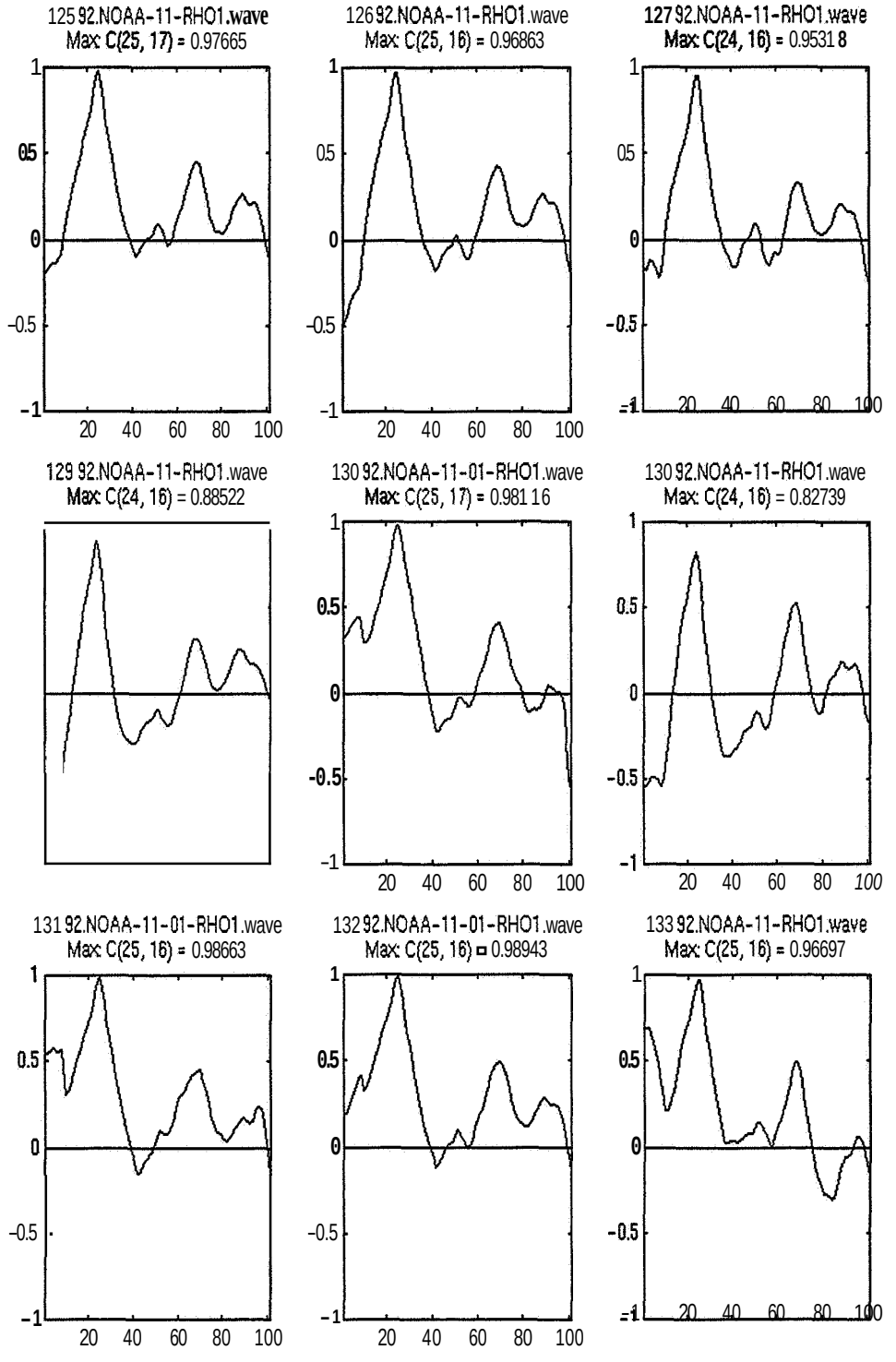


Fig. 8: Fractional-mask algorithm correlations for ROW 16.

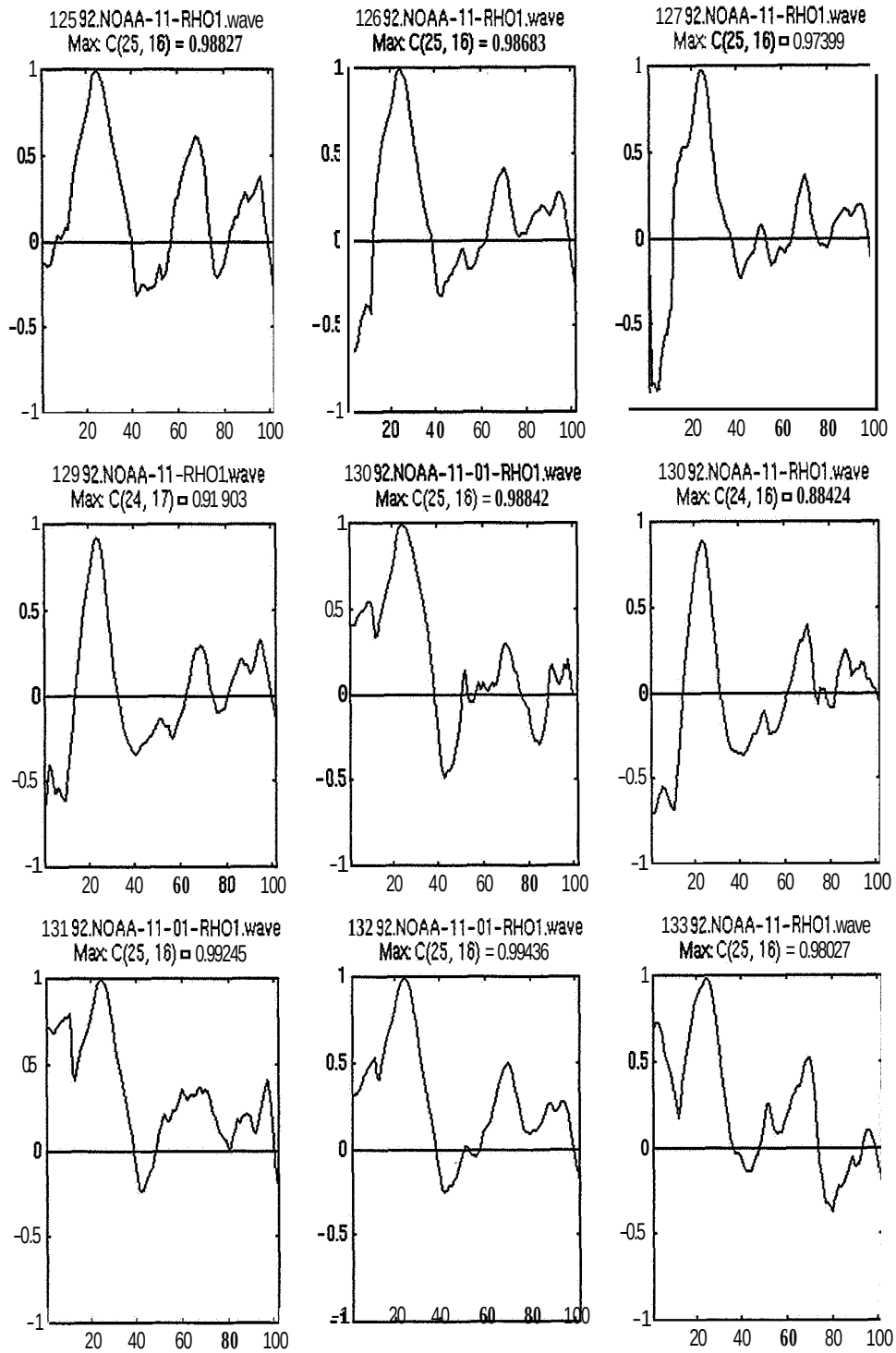


Fig. 9: Binary-mask algorithm correlations for Row 16.

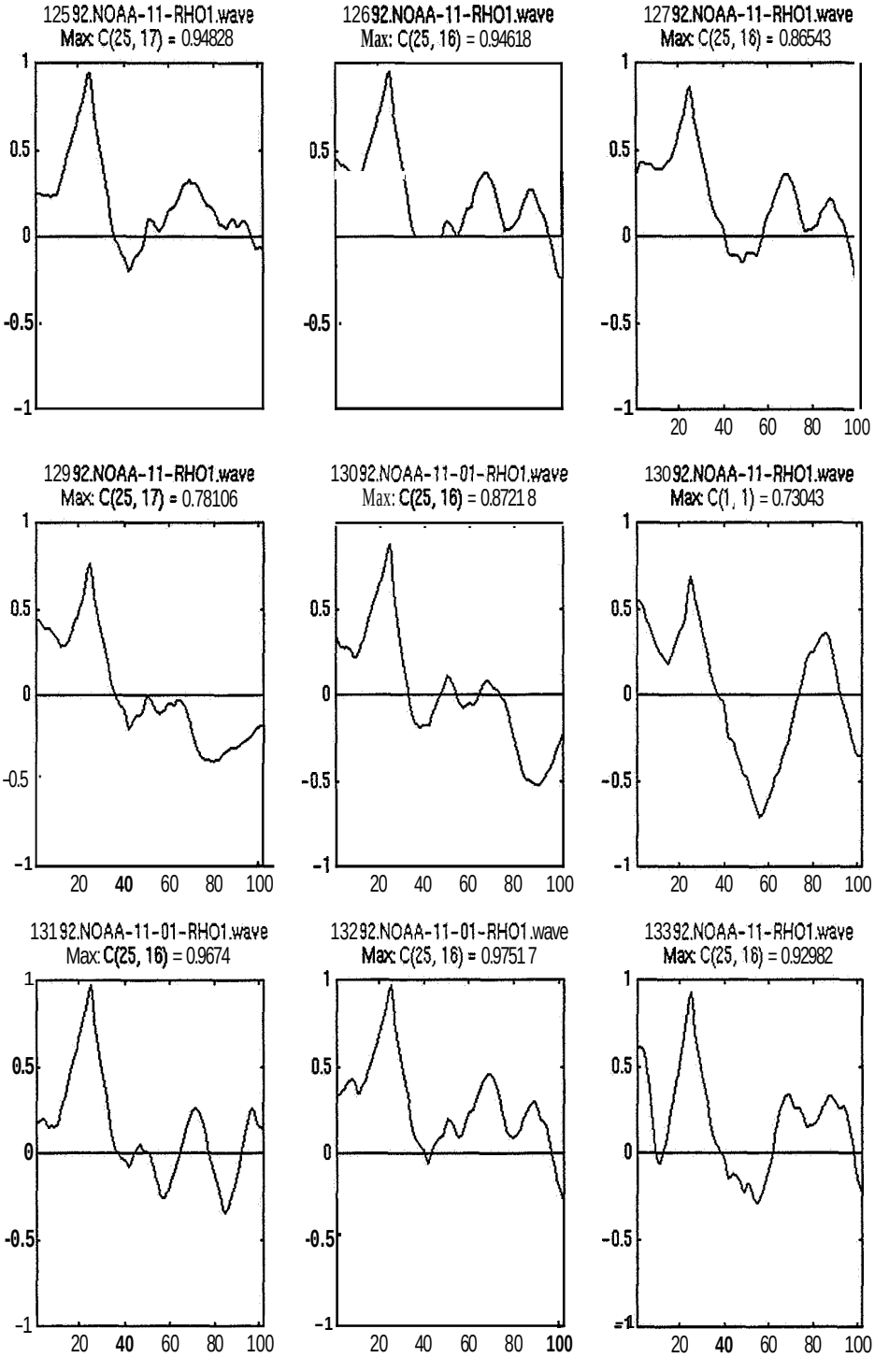


Fig. 10: Uncompensated algorithm correlations for Row 16.

Registration Assisted Mosaic Generation

Jean J Lorre
Thomas H. Handley, Jr
Jet Propulsion Laboratory
Pasadena, California

Abstract

This paper presents a general strategy for assembling mosaics from numerous individual images where uncertainty exists in the position and orientation of those images. Both of the presented applications relate to remotely operated camera platforms, the first being the Galileo solid state imaging (SSI) camera presently in orbit around Jupiter, and the second being the Imager for Mars Pathfinder (IMP) stereo camera on Mars.

A basic strategy in both applications is to determine the correct relative camera pointing followed by direct map projection of the images. It is assumed that approximate camera pointing exists sufficient to locate adjacent images and to place initial tiepoints within reach of the correlator

Spatial correlation is used to fix tiepoints whose initial locations are predicted by the camera pointing. We use either an fast fourier transform (fft) algorithm or a variant of Gruen's scheme permitting limited image rotation and skew

The Gruen correlator has three hierarchical modes:

1. A classical spatial least squares correlation on integral pixel boundaries used when rotation is small.
2. An annealing non-deterministic search used when rotations are unknown.
3. A simplex deterministic search used for the end game.

The correlation operation can be performed either interactively or autonomously.

The final camera pointing solution relies upon a simplex downhill search in 2n or 3n dimensions where n is the number of images comprising the mosaic and the objective function to be minimized is the disagreement between tiepoint locations predicted from the camera pointing with those observed by the correlator For Galileo the 3n unknowns are euler angles defining camera pointing in planet coordinates, and for Mars Pathfinder they are 2n unknowns representing commanded azimuth and elevation in the Lander coordinate system.

Introduction

Solid State Imager for Galileo The scientific objectives of the solid-state imaging (SSI) camera investigations have a wide scope: a comparative study of satellite surfaces, a study of the Jovian atmosphere, characterization of Jovian and satellite auroral phenomena, and an assessment of the rings of Jupiter For the Galilean satellites Io, Europa, Ganymede, and Callisto, the imaging investigators hope to map a large portion of each surface to a resolution of 1 kilometer or better In a few areas, features smaller than 10 meters will be distinguished. In addition, variations in color and albedo (reflectivity) will be mapped at a scale of about 2 kilometers. Scientists will look for changes on the surfaces since Voyager. The shape and the location of the spin axis of each Galilean satellite will also be measured.

The other smaller satellites will be studied throughout the orbital tour Studies will also be made of Jupiter's rings. Small, new satellites may be found in or near the rings.

The SSI will be used to determine structure, motions, and radiative properties of the atmosphere of Jupiter. It will measure wind profiles by tracking how fast clouds move at various altitudes. Radiative properties of the atmosphere, which are important for understanding energy management, will be determined by measuring the scattering of light from specific features at various wavelengths and at various angles of illumination, observations of auroral phenomena will be correlated with fields and particles measurements done with other instruments.

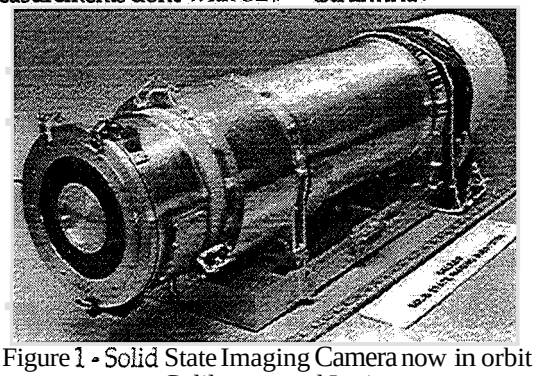


Figure 1 - Solid State Imaging Camera now in orbit on Galileo around Jupiter

The SSI is an 800- by 800-pixel solid-state camera consisting of an array of silicon sensors called a "charge-coupled device' (CCD). The optical portion of the camera is built as a Cassegrain (reflecting) telescope. Light is collected by the primary mirror and directed to a smaller secondary mirror that channels it through a hole in the center of the primary mirror and onto the CCD. The CCD sensor is shielded from radiation, a particular problem within the harsh Jovian magnetosphere. The shielding is accomplished by means of a 1-centimeter-thick layer of tantalum that surrounds the CCD except, of course, where the light enters the system.

An eight-position filter wheel is used to obtain images of scenes through different filters. The images may then be combined electronically on Earth to produce color images.

The spectral response of the SSI ranges from about 0.4 to 1.1 micrometers. (A micrometer is one millionth of a meter.) Visible light has a wavelength covering the band of 0.4 to 0.7 micrometers.

- The filter wheel has eight filters, of the interference type, centered at 611, 404, 559, 671, 734,756,887, and 986 nm.
- The SSI detector is a virtual phase, buried channel, thick, frontside illuminated, 800 x 800 line CCD.

- The SSI has a pre-flash system to eliminate residual images after each exposure. This system bathes the CCD in near-infrared light (~930 nm) several times and reads the CCD several times at highspeed.
- The SSI has five transmissive elements plus two mirrors.
- All SSI optical surfaces have anti-reflective coatings with a transmission factor of 62% at 576 nm.

  The front optical element is coated to minimize radiated heat loss and heaters provide thermal stability across the front aperture and within the body of the telescope.
- The long wavelength sensitivity limit (~1100 nm) is set by the detector and the short wavelength limit (~375 nm) by the anti-reflection coatings.

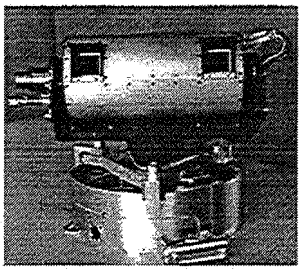
The SSI pointing uncertainty is about 20 pixels. Figure 3B illustrates the steps in mosaicking SSI images.

Imager For Mars Pathfinder (IMP) The Imager For Mars Pathfinder is a stereo imaging system with color capability provided by a set of selectable filters for each of the two camera channels. It has been developed by a team lead by the University Of Arizona with contributions from the Lockheed Martin Group, Max Planck Institute For Aeronomy in Lindau, Germany, the Technical University Of Braunschweig in Germany and the Ørsted Laboratory, Niels Bohr Institute for Astronomy, Physics and Geophysics in Copenhagen, Denmark. It consists of three physical subassemblies: (1) camera head (with stereo optics, filter wheel, CCD and pre-amp, mechanisms and stepper motors); (2) extendible mast with electronic cabling; and (3) two plug-in electronics cards (CCD data card and power supply/motor drive card) which plug into slots in the Warm Electronics Box within the lander.

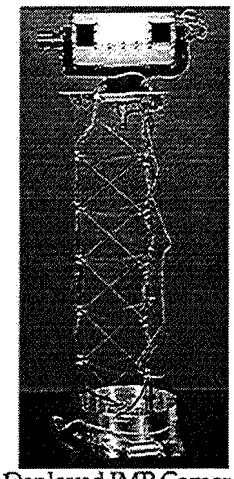
Azimuth and elevation drives for the camera head are provided by stepper motors with gear heads, providing a field of regard of ± 180 degrees in azimuth and +83 degrees to -72 degrees in elevation, relative to lander coordinates. The camera system is mounted at the top of a deployable mast, a continuous longeron, open-lattice type provided by Able Manufacturing, Inc. When deployed, the mast provides an elevation of 1.0 m above the lander mounting surface.

The focal plane consists of a CCD mounted at the foci of two optical paths where it is bonded to a small printed wiring board, which in tum is attached by a short flex cable to the preamplifier board The CCD is a front-illuminated frame transfer array with 23 micrometer square pixels. Its image section is divided into two square frames, one for each half of the stereo FOV's. Each has 256x256 active elements. A 256x512 storage section (identical to the imaging section) is located under a metal mask. The IMP focal plane and electronics are nearly identical copies of the comparable subsystem employed in the Huygens Probe Descent Imaging Spectroradiometer (DISR), using the Loral 512X512 CCD.

The stereoscopic imager includes two imaging triplets, two fold mirrors separated by 150 mm for stereo viewing, a 12-space filter wheel in each path, and a fold prism to place the images side-by-side on the CCD focal plane. Fused silica windows at each path entrance prevent dust intrusion. the optical triplets are an f/10 design, stopped down to f/18 with 23-mm effective focal lengths and a 14.4 degree field of view. The pixel instantaneous field of view is one milliradian. The filter wheel four pairs of atmospheric filters, two pairs of stereo filters, eleven individual geologic filters (which, when combined with the two pairs of stereo filters, result in thirteen distinct geologic filters) and one diopter or close-up lens, designed to acquire images of magnetic, wind-blown dust which adheres to a small magnet located on the IMP tip plate.



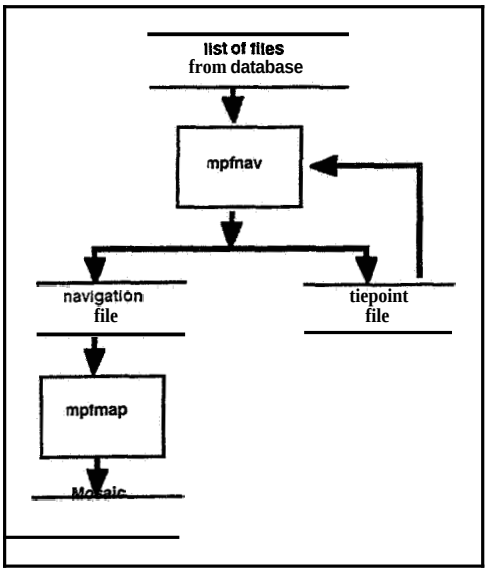
Stowed or Un-deployed IMP Camera
Figure 2 A IMP Camera System



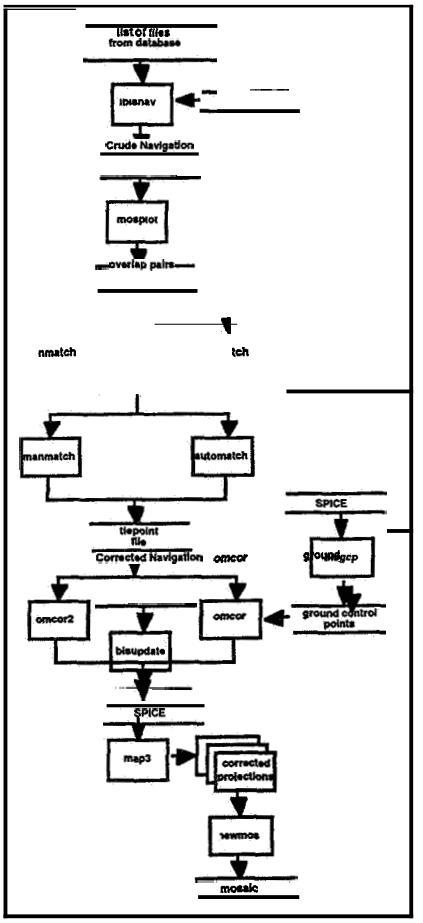
Deployed IMP Camera
Figure 2B IMP Camera System

Full panoramas of the landing site are acquired during the mission using the stereo baseline provided by the camera optics. Additionally, monoscopic panoramas are acquired both prior and subsequent to the mast deployment, yielding vertically displaced stereo pairs with approximately 80 an baseline. Images of a substantial portion of the visible surface are acquired in multispectral images with as many as eight spectral bands.

IMP pointing uncertainty can be as large as 0.8 degrees or 14 pixels. Figure 3A illustrates the steps in mosaicking IMP images.



Mars Pathfinder Mosaic Procedure
Figure 3a - Mosaicking Procedures



Galileo SSI Mosaic Procedure
Figure 3B Mosaicking Procedures

Method of Solution

The general strategy adopted at MIPL for assembling mosaics is to determine the true camera pointing for each image by analyzing a set of tiepoints connecting every combination of overlapping image pairs. Once the pointing is determined then a camera model can be used to relate pixels to the scene. A map projection program in concert with a mosaicking program can finally warp the individual image into a common rendering of the scene. This process reduces to knowledge of a Navigation Model, a Camera Model, a Correlator, an Objective Function which, when minimized, produces the desired navigation solution, and finally a Projection and Mosaicking program. We will discuss these in turn.

Navigation Model and Camera Model

One must be able to relate the pointing of the camera to the scene in a scene specific coordinate system. For Mars Pathfinder there are two parameters, namely the azimuth and elevation of the camera head in Lander spacecraft coordinates. For spacebome sensors such as Galileo there are six parameters, namely the x,y,z location of the spacecraft (which is known beforehand), and the three euler angles $\alpha_1, \alpha_2, \alpha_3$ which define the rotation from planet coordinates into camera coordinates. In order to assemble a mosaic from numerous images it is necessary to determine the pointing parameters (two for a lander and three for an orbiter) for each image in the mosaic.

For an orbiter viewing a planet a surface coordinate xyz is expressed in latitude ϕ and longitude λ as

$$\overline{op} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r\cos\phi\cos\lambda \\ r\cos\phi\sin\lambda \\ r\sin\phi \end{pmatrix}$$

where $\overline{op}$ is the vector from planet center to the surface and r is the radius. The vector $\overline{cp}$ from the spacecraft at $(x_s y_s z_s)$ to the point xyz on the planet surface is

$$\overline{cp} = \overline{op} - \overline{oc} = \begin{pmatrix} x - x_s \\ y - y_s \\ z - z_s \end{pmatrix}$$

where $\overline{oc}$ is the location of the spacecraft in planet coordinates and is known from the state vector Vector cp is rotated into camera coordinates by matrix M

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathrm{M}\ \overline{cp}$$

For Mars Pathfinder the azimuth and elevation define M whereas for Galileo the three euler angles describe M It is these angles we wish to determine for each image. Finally the image Coordinates $x''y''$ in the camera focal plane are related to $x'y'z'$

$$\begin{pmatrix} x'' \\ y'' \\ f \end{pmatrix} = \left(\frac{f}{z}\right)\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

and

$$l = l_0 + y''T$$

$$s - s_0 + x''T$$

where l and s are lines and sample, f is the focal length and T is the scale in pixels/mm.

Orbiter spacecraft navigation is modeled as above (reference 1). We have found it more expeditious to use an alternate mode (reference 2) for the lander camera because of the method of calibration.

The Mars Pathfinder Camera Model

The camera model adopted for Mars Pathfinder is based upon a vector model using four vectors $CAHV$ Vector ?locates the camera entrance pupil. Vector $\overline{H}$ lies parallel to images lines l and vector $\overline{V}$ lies parallel to image columns s Vector $\overline{A}$ points down the optical axis towards the scene. For an object point $\overline{P}$ its image lies at location s, l on the image plane

$$s = \frac{(\overline{P} - \overline{C}) \bullet \overline{H}}{(\overline{P} - \overline{C}) \bullet \overline{A}} \quad l = \frac{(\overline{P} - \overline{C}) \bullet \overline{V}}{(\overline{P} - \overline{C}) \bullet \overline{A}}$$

Calibrating a camera consists of extracting the four vectors $CAHV$ from an image of a field with known targets which have been surveyed in the coordinate system of choice. This calibration image was acquired at a particular commanded azimuth and elevation. To transform the camera model to another commanded position only requires rotation of $\mathbf{A}, H$ and $\overline{V}$ from the calibrated position to the new pointing. Similarly, the camera itself can be translated to a new location by changing vector $\overline{C}$ Both of these operations are required for Mars Pathfinder since the azimuth rotation point does not coincide with the optical axis. As with an orbiter camera, we are searching for the correct rotation matrix M which converts from calibrated to commanded pointing.

The Correlator

The mechanism adopted to obtain corrected camera pointing relies upon the acquisition of tiepoints between adjacent image pairs. We rely upon the provided navigation to determine which images overlap and to place a tiepoint closely enough that an autonomous correlator can function. The correlator of choice is that developed by Gruen (reference 3) because it provides great accuracy and it is tolerant of modest amounts of rotation. Our programs acquire tiepoints automatically and have an interactive option to assist in locating points if the initial camera pointing is too inaccurate.

The correlator has five options, each of which attempts to minimize a cost function which will be described later We will call these options 1-5. The objective is to locate a template selected in one image within a search area in another image.

Option 1 In this case the template is kept on integer pixel boundaries and is matched to each possible pixel location in the search area. That integer location which *minimizes* the cost function is selected. In order to return a sub-pixel location, the correlation values of points on a column and a row through the best match point are fitted to a quadratic and the peak of the quadratic is selected and returned in both sample and line. Since the correlation is never actually performed at the sub-pixel location the returned values are only estimates. The returned correlation value is still that determined at the best integer location. If the peak correlation occurs at pixel i and has an amplitude $f(i)$, then the location of the interpolated peak is computed from

$$x = i - \frac{1}{2}\frac{f(i-1) - f(i+1)}{2f(i) - f(i-1) - f(i+1)}$$

and similarly for y

Option 1 is the mode of choice when one knows:

1. That rotation and scale differences between the template and the search area are minimal.
2. That the location of the best correlation could be anywhere in the search area.
3. That accuracy is never required to exceed 1/10 of a pixel.
4. That time is of the essence.

Option 2 This option makes use of simulated annealing (reference 4) to amve at the best correlation location. This is the slowest of all options.

The method consists of guessing the six polynomial coefficients to an afine transformation which maps the template onto the search area. Any mapping is acceptable provided it remains within the search area. Each *guess* consists of adding to the last location six values obtained from a random number generator constrained to remains within a certain range, or temperature. Gradually the temperature is reduced so that *guesses* remains more localized, The heart of the algorithm is to compute at each step the Boltsman probability of *transitioning* from the previous correlation to the current one. If the current correlation is higher than the last then we adopt the new afine position. If it is lower we compute the probability of that transition and compare it with chance. If the coin comes up we accept it, and if the coin comes down we reject it and try again. The essence of annealing is that it gives us a way of escaping from local false minima in the solution space. Thus, it is a non-deterministic method because the next move is not constrained entirely by the last move. Unlike option 1, the algorithm does not systematically search the solution space for all combinations of mappings. It starts at an initial estimate and bounces about trying all sorts of combinations of afine mappings while remembering the best location visited. Repeatedly it is forced to revisit the best correlation location. Gradually the range of guesses is reduced until it freezes near the best cost function minimum. If the number of iterations is kept small, it will freeze at the wrong location. If the number of iterations is kept

large, the best correlation location will always be found but at the expense of time.

Option 2 is the mode of choice when one knows:
1 That there is an unknown amount of rotation, scale, or distortion between the template and the search area.
2. That the location of the best correlation could be anywhere in the search area.
3. That if the images are distorted this distortion is to be compensated for
4. That accuracy is never required to exceed 1/30 of a pixel.
5. That time is unimportant in exchange for a tiepoint.

Option 3 This mode uses the simplex downhill search strategy (reference 5). A simplex is a tetrahedron with one comer greater than *the* dimension of the problem or surface it resides on. In this case the surface is the cost functions which is a function of six dimensions, those of the six afine mapping coefficients. These coefficients map the template to the search area We seek the six coefficients where the cost function is a minimum. The simplex stands on this surface. There are four rules describing permitted changes in shape for the simplex as it seeks to walk down the surface towards a minimum. Eventually, it will find the bottom of the correlation surface and will compress itself down to the desired precision. This is a deterministic method because the next move depends entirely upon the last. Deterministic schemes have the drawback that if they start in the wrong minimum they have no means of escape. Therefore, the initial estimate for the afine mapping polynomial must be within the correct minimum for option 3 to function correctly The user can control the starting location for the search but this only sets two of six afine coefficients. In most cases the driving program provides initial estimates for a unity mapping, and this is adequate if the initial tiepoint is within the correlation distance. If data were strongly distorted, however, it might not suffice.

Option 3 is the mode of choice when one knows:
1 That if there is rotation, scale, or distortion between the template and the search area and that initial mapping polynomial coefficients are available to begin a search within the correct minimum.
2 That the initial tiepoint location is within a few pixels of the true one.
3 That as much accuracy is desired as possible.
4. That if the images are distorted this distortion is to be compensated for
5. That time is important but subordinate to accuracy.

Option 4 This is a hybrid mode In this case option 1 is first used to determine the tiepoint location. This location is then passed on to option 3 along with a unity mapping transformation.

Option 4 is the mode of choice when one knows:
1 That the location of the best correlation could be anywhere in the search area.
2 That the amount of rotation, scale, and distortion between the template and the search area is slight
3 That accuracy is essential.
4. That time is important.

Option 5 This is also a hybrid mode. In this case option 2 is used first to determine the mapping coefficients. These coefficients are then passed to option 3 which refines the solution. The combination provides great precision along with the minimum of *a priority* knowledge on the part of the user.

Option 5 is the mode of choice when one knows:
1 That the location of the best correlation could be anywhere in the search area.
2. That the amount of rotation, scale, and distortion between the template and the search area is substantial or unknown.
3. That accuracy is essential.
4. That time is unimportant

The Cost Function

Each of the five correlation mode discussed above is really a means of determining the location in an image where some quantity, which we call correlation value, is a maximum. This quantity is computed in the *same* fashion for all modes and is itself mode independent. The correlation value is a least squares cost function called the coefficient of determination (r) It measures the quality of a least squares linear fit made between the intensity values of the template and the corresponding intensities in the search area as determined by the afine mapping polynomial. This function value lies between 0.0 (no correlation at all) to 1.0 (perfect correlation- or anti- correlation). The correlation quality is computed from

$$r^2 = \frac{(\sum xy - \frac{\sum x \sum y}{mn})^2}{(\sum x^2 - \frac{(\sum x)^2}{mn})(\sum y^2 - \frac{(\sum y)^2}{mn})}$$

where x and y are the intensity values in the template and the search area respectively The cost function is $1-r$ Note, that because the measure is a least squares determination, correlation quality is indifferent to intensity differences between the template and the search area which are of the nature of scale, offset, or complement. Anti-correlations are just as valid as correlation's since both imply non randomness.

The Afine or Mapping Polynomial

All modes except option 1 permit the template to suffer a distortion before it is compared with the search area. The nature of the distortion is anything that a first order polynomial or afine transformation can do. There are six coefficients involved $(c_1 - c_6)$, three for sample and three for line. The options 2 through 5 are concerned with determining what these six coefficients are. By varying the coefficients one can simulate changes in offset, scale, rotation, skew, transpose, or flip. Since we are really interested in the tiepoint location we only want the offset term in the sample and the line equations, however, we need to compute all the terms in order to extract the two we want. The polynomial terms are of the form:

$$l_s = l_t c_1 + s_t c_2 + c_3$$
$$s_s = l_t c_4 + s_t c_5 + c_6$$

where subscripts refer to s for search area or t for template. s is sample and l is line.

The Objective Function

In order to determine the pointing parameters for each image we minimize an objective function which returns a single scalar, The scalar is the sum of the disagreements between the tiepoints predicted by the camera model and those selected by the software or user Each tiepoint is transformed into scene coordinates such as (latitude, longitude) or (x, y) on a plane and is then transformed into image coordinates in the adjacent image. The disagreement δx and δy between this predicted location based upon an assumed camera pointing and the actual correlated tiepoint is used to determine the validity of the assumed camera model.

$$\text{Cost function} = \sum_{n}^{N} (\delta x_n^2 + \delta y_n^2)$$

where N is the total number of overlapping image pairs and n is the pair number We use a numerical minimization routine based upon the simplex method (reference 5). For Mars Pathfinder there are $N \times 2$ unknowns to be optimized and for orbiters there are $M \times 3$ unknowns, where M is the number of images contained within a mosaic.

The largest mosaic navigated for Mars Pathfinder consist of about 180 images.

Projection and Mosaicking Program

Once the camera pointing is known it remains to decalibrate the images radiometrically and to transform the pixel values to the surface of a map or to a system of azimuth and elevation. Each mosaic comprises a single spectral band in radiance units. For orbiters the mosaics are collections of map projections in conformal or authalic projection and for landers like Mars Pathfinder they are either in Mars Local coordinates or in the format which a wider field of view camera would provide.

Figure 4 illustrates mosaicking of eighteen images from the Galileo SSI camera. Figures 5 and 6 illustrate the "before and after" images mosaicking of IMP images

Acknowledgments

The authors wish to thank JPL's Multimission Image Processing Laboratory (MIPL) development and operations team (Galileo and Mars Pathfinder) for their collective and individual contributions to the understanding and refinement of these results.

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

If you have comments, questions, or would like to discuss the algorithms, findings, or conclusions, the authors may be reached by

jean.lorre@jpl.nasa.gov
1.818.354.2057
tom.handley@jpl.nasa.gov
1.818.354.7009

References

1 Duxbury, T.C. "Television Investigations of Phobos" Genry A. Avanesov, Editor Nauka Publishing House, USSR Academy of Sciences, Moscow USSR

2. Yakimovsky, Y., Cunningham, R., "A System for Extracting Three-Dimensional Measurements from a Stereo Pair of TV Cameras", Computer Graphics and Image Processing, 7,195-210 (1978)

3. Gruen, A. W., Baltsavias, E. P., "Adaptive Least Squares Correlation with Geometrical Constraints", SPIE Computer Vision for Robots, Vol. 595, 1985 Cannes.

4. Lorre, J. "Function Minimization With Partially Correct Data via Simulated Annealing", SPIE, Applications of Digital Image Processing XI, August 1888, Vol. 974, pp. 398-403.

5. Press, W H, et al, Numercial Recipes, section 10.4, Cambridge University Press, Cambridge, 1986.

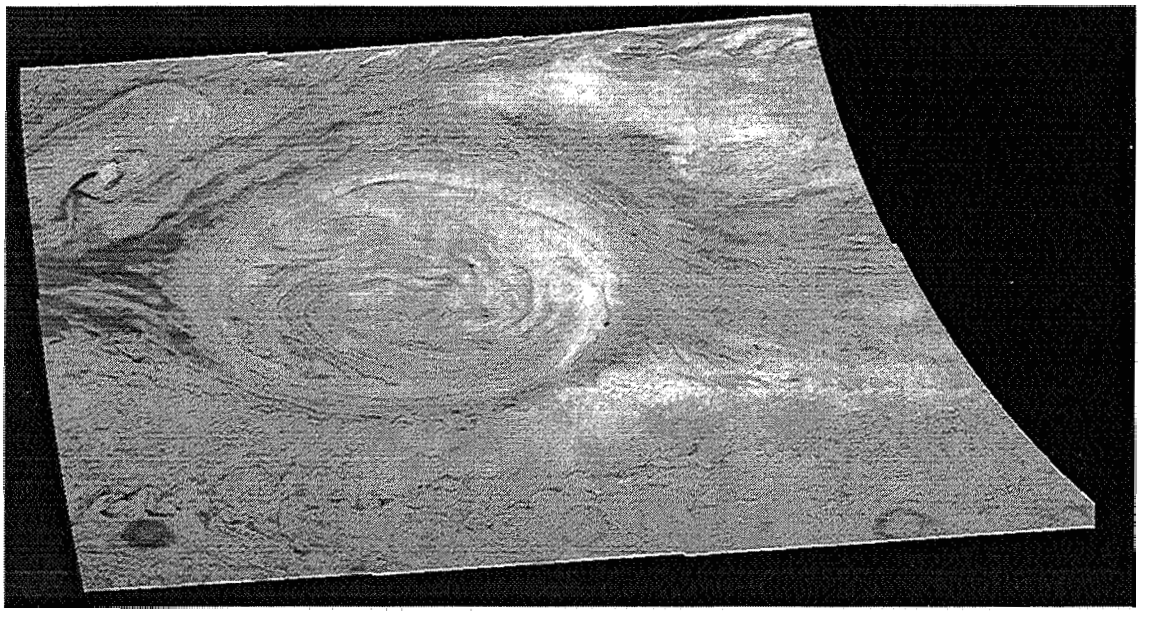


Figure 4 A Galileo mosaic Jupiter's red spot consisting of 18 image

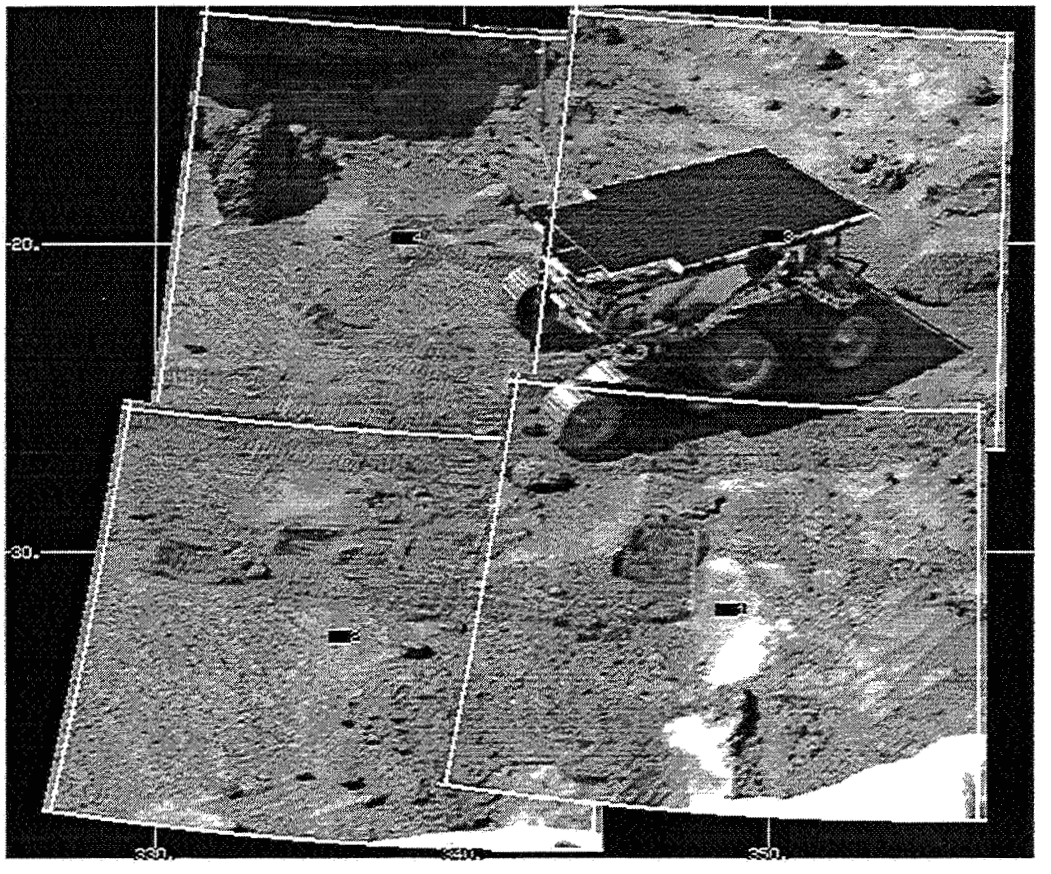


Figure 5 A Fragment of an uncorrected mosaic of four images taken by Mars Pathfinder

Figure 6 The same mosaic as figure 4 but corrected for camera pointing based upon common tiepoints.

COMPARISON OF REGISTRATION TECHNIQUES FOR GOES VISIBLE IMAGERY DATA

James C. Tilton
NASA's GSFC
Applied Information Sciences Branch
Greenbelt, MD **20771**

Abstract

This paper briefly describes and then compares the effectiveness of five image registration approaches for GOES visible band imagery. The techniques compared are (i) NOAA "point" manual landmarking, (ii) manual landmarking on extant feature (whole island, lake, etc.), (iii) automatic phase correlation, (iv) automatic spatial correlation on edge features, *and* ***(v) automatic spatial correlation of region boundaries derived from image segmentation.***

1 Introduction

Starting the GOES-8 satellite, the GOES pointing stabilization went from a spin-stabilized system to a gyroscope stabilized system. Built in to the new gyroscope stabilized system is a feedback mechanism for correcting for pointing inaccuracies due to mainly to gravity anomalies and platform motion caused mechanical motion of systems on the platform. This feedback mechanism requires the uploading from Earth of pointing correction factors. NOAA had originally hoped that an automatic phase correlation approach implemented in the NOAA Product Monitor would be sufficient. Early experimentation, however, showed that the automatic phase correlation was not robust enough for routine use (primarily due to the confounding effects of even a minor amount of clouds). NOAA then had to used its fall-back option of "point" manual landmarking (described below) in order to provide the required correction factors.

This paper briefly describes and then compares the effectiveness of five image registration approaches for GOES visible band imagery The techniques compared are (i) NOAA "point" manual landmarking, (ii) manual landmarking on extant feature (whole island, lake, etc.) - which is considered to be the "true" registration, and (iii) a modified version of the NOAA automatic phase correlation. Two new approaches are also described and compared to the earlier approaches: **(i)** automatic spatial correlation on edge features, and (ii) automatic spatial correlation of region boundaries derived from image segmentation.

Comparative results are given for varied sets of GOES visible imagery

2 Registration Techniques

NOAA "point" manual landmarking: The NOAA operator displays a 128x128 pixel section of the GOES-8 imagery data on the Product Monitor and compares the image to a Landmark map. The operator finds the line and pixel location of a particular land/water feature (peninsula, bay, etc.) and records the ***difference*** between the expected line and pixel location and the line a pixel location actually found.

Manual landmarking on extant feature (whole island, lake, etc.): Performed manual registration of GOES-8 imagery data from a particular day and Landmark by flickering between successive images using the Khoros function "animate." This result was used as a baseline "best result" for comparison to the other techniques. This version of manual registration differs from the NOAA landmarking in that is uses spatial extant features (e.g. whole island, large section of coastline, whole lake, etc.).

Automatic phase correlation: Phase correlation is a mathematical technique that was developed to register images to one another, It works best in cases in which the misregistration is only a translation (which is the case for GOES data). The technique can be described as follows (from [1]):

> Given a reference image, g_R and a sensed image g_S, with two-dimensional Fourier transforms G_R and G_S, respectively, the cross-power spectrum of the two images is defined as $G_R G_S^*$ and the phase of that spectrum as

$$e^{j\phi} = G_R G_S^* / |G_R G_S^*|$$

The phase correlation function, d, is then given by

$$d = F^{-1}\{e^{j\phi}\}$$

where F^{-1} denotes the inverse Fourier transform.

The spatial location of the peak value of the phase correlation function, d, corresponds to the translation misregistration between g_R and g_S.

The innovation in the implementation demonstrated here is in the finding of the peak of the correlation function, d. Instead to looking for an interpolated peak of d (as in the implementation installed - and not used - on the NOAA Product Monitor), the center of mass of the peak of d is found. We have found that this gives a more robust result than searching explicitly for the peak.

Automatic spatial correlation on edge features: Automatic edge detection is followed by spatial correlation (multiply reference image times the shifted input image). Automatic edge detection was performed using the "vdrf" function from Khoros - which is an implementation of an edge detection algorithm developed by Shen and Castan [2,3].

Automatic spatial correlation of region boundaries derived from image segmentation: Automatic image segmentation is followed by spatial correlation (multiply reference image times the shifted input image). Automatic image segmentation is performed by Iterative Parallel Region Growing developed by Tilton [4].

One problem was encountered when spatially correlating the reference image and input image. Mainly due to varying lighting conditions, the input image region outline obtained may be expanded or shrunk as compared to the reference image region outline. In tests reported here, this was compensated by slightly smearing the reference image region outline (region boundary pixels were given the value "3," while pixels touching the region boundary pixels were given the value "2," and the pixels touching the pixels touching the region boundary pixels were given the value "1." Results using both the unsmeared and smeared reference image region outlines are given.

3 Results

Registration results were obtained for three widely different GOES 8 visible scenes, examples of which are given in Figures 1, 2 and 3. Complete results for the scene over Isla Ángel de la Guarda off of Baja California, Mexico (Figure 1) are given in Table 1, using the non-smeared reference image for the region boundary registration case. *Summary* results for all three images are given in Table 2. The NOAA point manual landmarking shows an overall bias shift versus the other techniques. This is probably due to a problem with the navigation program used to locate the image segments - it is probably not an indication of an overall bias in the NOAA point manual landmarking itself. The standard deviation results indicate that the spatial Correlation on edge features approach was most consistent in matching the manual landmarking on extant features (assumed as reference). The results also show that "smearing" the reference edge map did improve the results from the spatial correlation on region boundaries approach for the BAJ case, but not enough to match the results produced by the spatial correlation on edge features approach.

References

[1] C. D. Kuglin, A. F Blumenthal, J. J Pearson, "Map-matching techniques for terminal guidance using Fourier phase information," SPIE Vol. 186 Digital Processing of Aerial Images, pp. 21-29, May 22-24,1979.

[2] J Shen and S. Castan, *"An* optimal linear operator for edge detection," Proceedings of CVPR '86, Miami Beach, FL, pp. 109-114, 1986.

[3] J. Shen and S. Castan, *"An* optimal linear operator for step edge detection," Graphical Models and Image Processing, Vol. 54, No. 2, pp. 112-133,1992.

[4] J. C. Tilton, "A refinement of Iterative Parallel Region Growing and application to remotely sensed imagery data," to be submitted to IEEE Transactions on Geoscience and Remote Sensing.

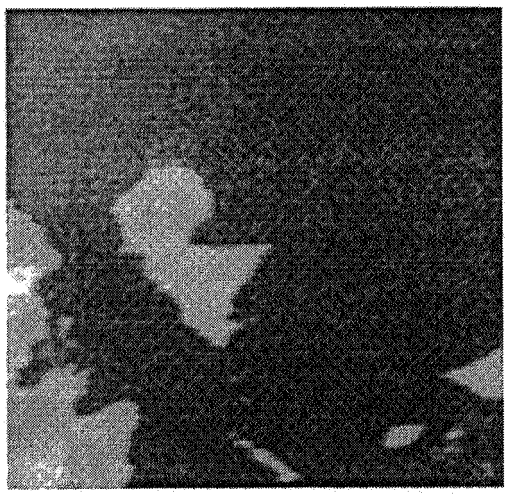


Figure 1 GOES 8 visible image from 1815 UTC on June 12,1996, from over Isla Angel de la Guarda off of Baja California, Mexico. Selected as reference image.

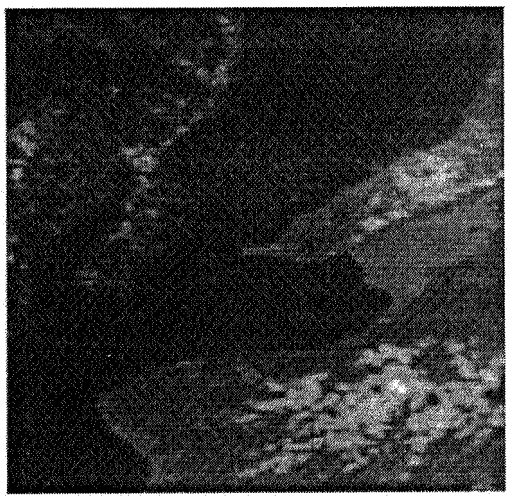


Figure 2. GOES 8 visible image from 1915 UTC on June 13, 1996, from over Puerto Vallarta, Mexico. Selected as reference image.

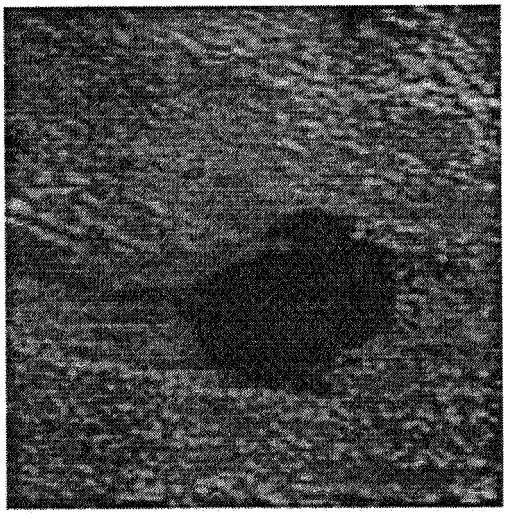


Figure 3. GOES 8 visible image from 1445 UTC on June 14, 1996, from over Lake Okeechobee, Florida, U.S.A.. Selected as reference image.

Table 1 Shifts west and east from the manual registration result for Phase Correlation, Edge Feature Correlation, Region Boundary Correlation, and NOAA Point Manual Landmarking for the GOES 8 scene over Isla Angel de la Guarda off of Baja California, Mexico (Figure 1).

	Phase	Corr.	Edge	Corr.	Region	Boundary	NOAA	Landmark
UTC	shift west	shift east	shift west	shift east	shift west	shift east	shift west	shift east
1345	-0.05	-0.31	-0.25	0.50	-0.75	0.50	-2.00	-3.00
1402	-0.48	-0.34	0.25	0.50	0.25	0.50	-2.25	-3.25
1432	0.16	0.19	-0.25	0.00	0.00	0.75	-2.75	-3.00
1602	-0.01	0.96	-0.25	0.00	0.00	1.00	-2.25	-3.50
1615	0.12	-0.22	-0.25	0.25	0.25	0.50	-2.00	-3.00
1632	0.34	-0.27	0.00	0.50	0.25	0.50	-2.50	-3.00
1645	0.35	0.48	0.00	0.00	0.00	0.25	-1.75	-2.75
1702	0.00	-0.35	0.25	0.50	0.50	0.50	-2.00	-3.00
1715	0.64	1.00	0.50	0.50	0.00	2.00	-2.25	-3.25
1732	0.26	-0.18	0.25	0.25	0.50	0.50	-1.00	-3.00
1745	0.20	-0.73	0.25	0.25	0.50	0.50	-0.50	-2.75
1815	0.00	-0.01	0.00	0.00	0.00	0.00	-1.00	-3.00
1845	0.42	-0.29	0.50	0.50	0.50	0.50	-0.75	-3.00
1902	0.34	-0.22	0.50	0.25	0.25	2.00	-1.75	-3.00
1915	0.64	-0.24	0.50	0.25	0.75	2.00	-1.50	-3.00
1932	0.25	-0.40	0.75	0.50	0.50	2.25	-0.75	-3.00
1945	0.12	-0.45	0.25	0.50	0.25	2.50	-1.75	-3.00
2002	0.50	0.77	0.25	0.25	0.00	0.25	-2.50	-3.50
2015	0.18	-0.27	0.00	0.50	-0.25	1.75	-1.75	-3.00
2032	0.30	-0.37	0.25	0.50	-0.25	0.75	-2.50	-3.00
2045	0.09	-0.09	0.00	0.25	0.00	1.00	-1.75	-3.00
2115	0.16	-0.40	0.00	0.50	0.25	0.50	-1.75	-3.00
2132	0.25	0.57	0.25	-0.25	-0.25	0.00	-3.00	-3.75
2145	0.25	0.37	0.00	0.25	0.00	0.00	-1.50	-2.75
2232	0.27	-0.13	0.75	0.50	0.50	0.50	-3.50	-4.00

Table 2. Mean and standard deviation **of** the **shifts** west and east from the manual registration result for Phase Correlation, Edge Feature Correlation, Region Boundary correlation, and NOAA Point Manual Landmarking for all three GOES 8 image data sets. BAJ corresponds to the data **from** over Isla Angel de la Guarda off of Baja California, Mexico (figure 1). P W corresponds to the data from over Puerto Vallarta, Mexico (figure 2). OKE corresponds to the data from over Lake Okeechobee, Florida, **U.S.A.** (figure 3**).**

	Phase	Corr.	Edge	Corr.	Region	Bound.	Region	Bound.	NOAA	Land.
	shift west	shift east	shift west	shift east	shift west	shift east	shift west (smeared)	shift east smeared)	shift west	shift east
BAJ mean	0.21	-0.04	0.18	0.31	0.15	0.86	0.14	0.57	-1.88	-3.10
BAJ stdv	1.13	2.28	1.44	1.08	1.62	3.69	1.54	2.11	3.56	1.46
PUV mean	-3.64	4.34	-0.24	0.10	0.23	0.56	0.26	0.45	-0.57	-2.48
PUV stdv	43.42	58.20	5.10	2.59	10.21	3.94	10.48	3.40	7.51	3.89
OKE mean	-0.54	0.50	-0.38	0.30	-0.86	0.50	-0.80	0.47	-5.75	-3.03
OKE stdv	3.06	2.15	1.87	0.81	3.45	1.41	3.92	1.62	2.78	1.87

Iterative Edge- and Wavelet-Based Image Registration of AVHRR and GOES Satellite Imagery

Jacqueline Le Moigne [1], *Nazmi El-Saleous* [2], *Eric Vermote* [2]
lemoigne@cesdis.gsfc.nasa.gov

(1) USRA/CESDIS; (2) University of Maryland
Contact. Code 930.5, Goddard Space Flight Center, Greenbelt, MD 20771

Abstract

Most automatic registration methods are either correlation-based, feature-based, or a combination of both. Examples of features which can be utilized for automatic image registration are edges, regions, comers, or wavelet-extracted features. In this paper, we describe two proposed approaches, based on edge or edge-like features, which are very appropriate to highlight regions of interest such as coastlines. The two iterative methods utilize the Normalized Cross-Correlation of edge and wavelet features and are applied to such problems as image-to-map registration, landmarking, and channel-to-channel co-registration, utilizing test data, AVHRR data, as well as GOES image data.

1. Introduction

Digital image registration is very important in many applications of image processing, such as medical imagery, robotics, visual inspection, and remotely sensed data processing. For all of these applications, image registration is defined as the process which determines the most accurate match between two or more images acquired at the same or at different times by different or identical sensors. Registration provides the "relative" orientation of two images (or one image and other sources, e.g., a map), with respect to each other, from which the absolute orientation into an absolute reference system *can* be derived. Image registration is usually motivated by such goals as object recognition, model matching, pose estimation, or change detection. In this paper, we will only refer to remote sensing applications, for which automated image geo-registration has become a highly desirable technique.

Currently, the most common approach to remotely sensed image registration is to extract a few outstanding characteristics of the data, which are called ***control points*** (CP's), ***tie-points,*** or ***reference points,*** The CP's (usually selected interactively) in both images (or image and map) are matched by pair and used to compute the parameters of a geometric transformation. But such a point selection represents a repetitive, labor- and time-intensive task which becomes prohibitive for large amounts of data. Also, too few points, inaccurate points, or ill-distributed points might be chosen thus leading to large registration errors. Recently, a relatively important number of automatic image registration methods have been developed [1,2], and they focus on the automatic selection of the control points, on the speed of processing as well as on the accuracy of the resulting registration. Most automatic registration methods are either correlation-based, feature–based, or a combination of both. Examples of features which are utilized for registration are edges, regions, comers, or wavelet-extracted features

In this paper, we describe two proposed approaches based on edge or edge-like features, which are very appropriate to highlight regions of interest such as coastlines. These two algorithms perform automatic image registration in an iterative manner, first estimating the parameters of the transformation, and then iteratively refining these parameters. Preliminary results are presented utilizing test data, as well as map-to-image registration of AVHRR image data, and landmarking and co-registration of GOES data.

2. Edge- and Wavelet-Based Registrations

Among all potential features to utilize as control points, edges and edge-like features appear as ones of the most promising. Often, a human operator will choose sharp curvatures in coastlines or rivers, the intersections of two roads or the coastline of a lake as control points to perform manual registration. The initial approaches presented in this paper are based on these observations as well as on the conclusions of three previous studies [3,4,5] which compare manual registration, edge matching, and phase correlation. All these converge to conclude that edge or edge-like features are very appropriate to highlight regions of interest such as coastlines. In one of these studies [4], the Normalized Cross-Correlation [6] is rated as one of the best matching measures. Therefore, in this study we present the results obtained by two methods based on the Normalized Cross-Correlation of edge and wavelet features.

2.1 Edge Detection

An edge detection computes the gradient of the original gray levels and highlights the pixels of the

image with higher contrast. Since edge values are less affected by local intensity variations or time-of-the day conditions than original gray level values, edge features should be more reliable than original gray levels. Figure 1 shows on an example of a coastliie how edge detection can be useful to detect the exact location of the edge point A located on the coastline: if we draw a line perpendicular to the coastline at point A, and we consider the function representing the intensity of the image points along this line, we get a function which is about constant over the land, jumps down at the edge and gets constant again over the water If we compute the gradient (or first derivative) of this function, we get a edge function which is about zero everywhere except around the edge point where it reaches a maximum. Locating this maximum is equivalent to finding the edge point, i.e. the coastline point A in this example.

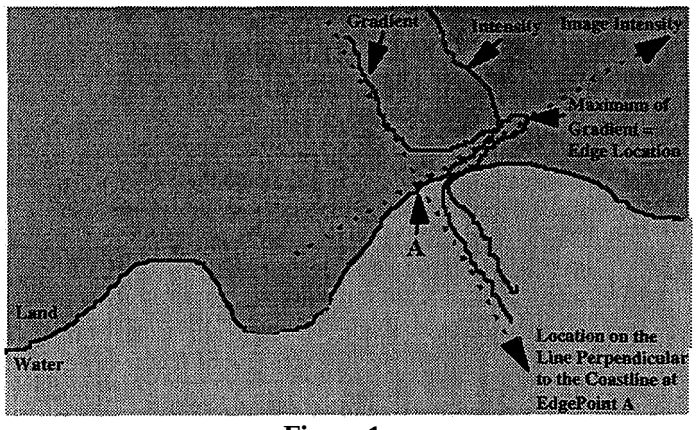


Figure 1
Edge Detection for Coastlines Extraction

In the method presented below, we will be using a Sobel edge detector which computes the gradient (or first derivative) of the 2-D image signal through two filters, an horizontal filter, Fh and a vertical filter, FV

$$Fv = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \qquad Fh = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

The magnitude of the gradient is computed as:

$M = \mathrm{Sqrt}(Fv^2 + Fh^2)$

and the direction at each point is computed as:

$D = \mathrm{Arctg}(Fv/Fh)$.

Higher magnitudes values of M correspond to edge features including coastlines. Other edge detection methods [7,8], computationally more expensive but less sensitive to noise, will be investigated later

2.2. Wavelet Decomposition

Similarly to a Fourier transform, wavelet transforms provide a time-frequency representation of a signal, which can be inverted for later reconstruction. However, the wavelet representation allows a better spatial localization as well as a better division of the time-frequencyplane than a Fourier transform, or than a windowed Fourier transform. In a wavelet representation, the original signal is filtered by the translations and the dilations of a basic function, called the "mother wavelet" For the algorithm described below, only discrete orthonormal basis of wavelets have been considered and have been implemented by filteringthe original image by a high-pass and a low-pass filter, thus in a multi-resolution fashion. At each level of decomposition, four new images are computed; each of these images is half the size of the previous original image and represents the low frequency or high frequency information of the image in the horizontal or/and the vertical directions; images LL (Low/Low), LH (Low/High), HL(High/Low), and HH (High/High) of Table 1 Starting again from the "compressed" image (or image representing the low-frequency information), the process can be iterated, thus building a hierarchy of lower and lower resolution images. Table 1 summarizes the multi-resolution decomposition.

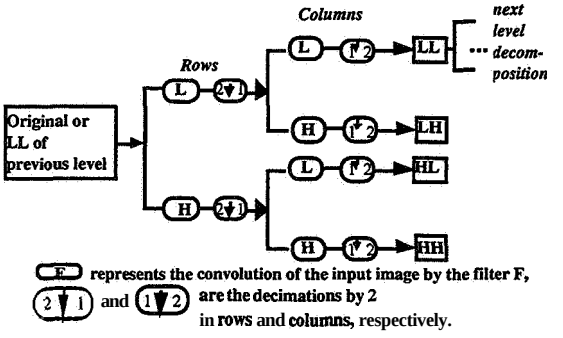


Table 1
Multi-Resolution Wavelet Decomposition

The features provided through this type of wavelet decomposition are of two different types: the low-pass features which provide a compressed version of the original data and some texture information, and the high-pass features which provide detailed information very similar to edge features. The advantages of using a wavelet decomposition are twofold; (a) by considering the low-pass information, one can bring different spatial resolution data to a common spatial resolution without losing any significant features, which is very useful for channel-to-channel eo-registration or registration of multi-resolution data, @) by utilizing high-pass information, one can retrieve significant features which are correlated in the registration process, similarly to edge features. When these features are extracted at a lower resolution, only the strongest features are still present, thus eliminating weak higher resolution features. Furthermore, utilizing wavelet transforms ties wavelet-based registration algorithms to a more general data management framework [9], in which wavelet decomposition could serve the multi-purpose of image compression, reconstruction, and content extraction as well as image registration.

Our wavelet-based registration algorithm [2,10,11] is based on the high-frequency infirmation extracted from the wavelet decomposition (images "LH" and "HL" from Table 1). Only those points whose intensities belong to the top x% of the histograms of these images are kept (x being a parameter of the program whose selection can be automatic); we call these points "maxima of the wavelet coefficients," and these maxima form the feature space.

2.3 Feature Matching

Correlation measurement is the usual similarity metric [6], although it is computationally expensive and noise sensitive when used on original gray level data. Using a pre-processing such as edge detection or a wavelet multi-resolution search strategy enables large reductions in computing time and increases the robustness of the algorithms. If R and I are the reference and the input images, the correlation measure between R and I is defined by:

$$Corr(R,I) = \frac{\Sigma((Rj-R^*).(Ij-I^*))}{\sqrt{\Sigma(Rj-R^*)^2 \ \Sigma(Ij-I^*)^2}}$$

with R^* and I^* are the mean of images R and I, respectively Other similarity metrics are described in [12] and will be evaluated in future work as well as the use of a robust feature matching algorithm which is described in [13].

2.4 Description of the Two Algorithms

In general, we will assume the transformation to be either a rigid or an affine transformation. Although high-order polynomials are superior to isometric transformations when the deformation includes more than a translation and a rotation, isometric transformations are more accurate when a smaller number of points is available and are less sensitive to noise and to largely inconsistent tie-points. Both types of transformations include compositions of translations and rotations; therefore, as a preliminary study, our search space is composed of rigid transformations for the edge-based method and compositions of 2-D rotations and translations for the wavelet-based method, and will be extended later to affine transformations. Both edge- and wavelet-based methods represent a three-step approach to automatic registration of remote sensing imagery The first step involves the edge extraction or the wavelet decomposition ofthe reference and input images to be registered. In the second step, we extract domain independent features from both reference and input images. In the wavelet method, feature extraction is perbrmed at each decomposition level. Finally, we utilize these features to compute the transformation function. Both methods perform the registration in an iterative manner, first estimating the parameters of the deformation transformation, and then iteratively refining these parameters.

For the edge-based implementation, we chose to model the transformation as a combination of a scaling in both directions (dsx, dsy), a rotation (dΘ), and a shift or translation (dtx,dty)in both directions. This algorithm is based on the assumption that rotation angle and scaling parameters are small (within 5 degrees for the rotation angle and within [0.9,1 1] for the scaling parameters). At the first iteration, the translation parameters are first extracted assuming rotation and scaling to be negligible. Then, knowing the translation parameters, the rotation angle is found, assuming the scaling negligible. Then scaling parameters are retrieved. At each next iteration, the five parameters are retrieved simultaneously, by computing the cross-correlations for all successive values of the parameters taken at incremental steps. At each iteration, the *accuracy* on the transformation parameters is divided by two. So, if at the first iteration, the translation parameters are found at a 1 pixel accuracy, the accuracy will be 0.5 pixel at iteration 2, 0.25 at iteration 3, etc.

For the wavelet-based method, the search intervals on the parameters can either be chosen arbitrarily or be reduced by the user utilizing a priori information about the sensor movement and the satellite navigation system. The search strategy is explained below when looking only for rotations and follows the multiresolution provided by the wavelet decomposition. At the deepest level of decomposition (where the image size is the smallest), the search is exhaustive over the whole search space (e.g., [0,90degrees]) but with an accuracy equal to A (e.g., 8 degrees). The first approximation of the best rotation, Rn, is chosen over this search space; then Rn becomes the center of a new search interval of length 2A, [Rn - A , Rn + A], and at the next lower level, the new approximated rotation, Rn-1, is found within this search interval with an accuracy of Δ/2. This process is repeated until the first level of decomposition, where the search interval is [R2 - Δ/2**(n-2), R2 + Δ/2**(n-2)J and the final registration rotation, R1, is found with an accuracy equal to Δ/2**(n-1). In particular, if δ is the desired registration accuracy (e.g., 1 degree), Δ is chosen as 2**(n-1) δ, where n is the number of levels of wavelet decomposition. In the case of an affine transformation, instead of looking for all parameters simultaneously, the search is performed over each set of parameters independently, first looking for shifts, then for rotations, etc. Such a use of successive "subsearches" reduces dramatically the amount of computations. Other solutions will be investigated, such as the use of an optimization technique or of a statistically robust point matching method [13].

Below is a more general description of both registration algorithms using edge or wavelet features for AVHRR and GOES image data:

(R1) ***Preprocessing Step:*** The pre-processing step enhances the contrast of the features which are

utilized to perform the registration. For some channels, gray levels may have to be inverted to consider homogeneous computations.

(R2) *Wavelet Decomposition:* Since images to be registered may have different spatial resolutions, the wavelet decomposition step brings these images to a common spatial resolution without degrading the image quality. Wavelet decomposition is pursued further down if wavelet coefficients are used for the registration.

(R3) *Registration:* This step can be performed by cross-correlating either edge features or wavelet features. With either type of features, the registration is performed by:

(R3.1) Estimating independently the five parameters: rotation, shift in the x-direction, shift in the y-direction, scaling in x, and scaling in y First the rotation and the scalings are assumed to be negligible and the shift in x and y are estimated. Then, taking into account the estimated shift, the two scaling parameters are neglected and the rotation angle is estimated. Then, after applying the previous shift and rotation parameters, the two scaling parameters are computed. These scaling parameters are kept constant for the rest of the search.

(R3.2) The three previous rotation and shift parameters are iteratively refined, at better and better accuracies. For example, if the first step looks at an accuracy of 2 degrees rotation, and 2 pixels shift, four successive iterations will look at the respective accuracies of 1 degree/1 pixel , 0.5 degree/0.5 pixel, 0.25 degree/0.25 pixel.

The two differences between the edge-based registration and the wavelet-based registration reside in the type of features that are considered to perform the registration, edges versus wavelet coefficients, and in the size of the images on which the computations are carried out: for the edge-based registration, the full size images are utilized for every step. For the wavelet-based registration, the initial search is carried out on the lowest level of wavelet decomposition, i.e., the smallest size images, then each refinement is computed on the next size-up, with the fmal refinement being computed on the full size image. This last variation explains the difference in the number of operations needed for each algorithm, about 480 floating point operations per pixel for the edge-based registration versus about 200 floating point operations per pixel for the wavelet-based registration.

3. Results

These two algorithms have been tested with four different types of datasets;

- the first dataset represents *synthetic test patterns* where pattems and transformations are well-controlled.
- the second dataset is representative of *multi-temporal studies* and is firmed with a series of AVHRR/LAC scenes over South Africa. For this application, a map of the coastlines is available so the images can be geo-registered to the map.
- the third dataset is an example of *landmark navigation* with a series of five GOES images of a landmark, "Cape Cod," which are registered to a reference image (or "chip,,) of this landmark.
- the fourth dataset represent examples of *channel-to-channel co-registration (or calibration)* with two series of five-channels GOES images for which channel-to-channelregistration is performed.

3.1. Test Data

Figure 2 represents seven diffit test patterns which have been utilized to test the edge-based registration method. This series of test patterns was created to test the response of our edge-based algorithm to edge directions, local intensity variations as well as texture variations.

A few preliminary experiments were conducted by applying two known transformations to the original image data (in this case, a composition of a rotation and a translation), and then registering the transformed images to the original image using the edge-based automatic registration. Results are shown in Table 3 and indicate average absolute errors of 0.12 degree in rotation and 0.28 pixel in translation. Although these first results are very encouraging, the present size of the dataset is not sufficient to give any conclusions about the accuracy or the precision of the system, but it will be extensively tested in future experiments.

TEST PATTERN	"TRUE" TRANSFORM (Rot: degrees TransIX: pixels TransIY: pixels)			COMPUTED TRANSFORM			ERROR=Itrue-computedI		
	(Rot	TX	TY)	(Rot	TX	TY)	(Rot	TX	TY)
GRID2.2.15.15	(0.5	1.8	0.5)	(0.5	2.2	0.85)	(0	0.4	0.35)
	(1	0.5	1.5)	(1	0.95	1.9)	(0	0.45	0.4)
GRIDG2.2.15.15	(0.5	1.8	0.5)	(0.5	1.85	0.5)	(0	0.05	0)
	(1	0.5	1.5)	(0.7	0.6	0.75)	(0.3	0.1	0.75)
GRID5.5.20.20	(0.5	1.8	0.5)	(0.5	0.75	0.55)	(0	1.05	0.05)
	(1	0.5	1.5)	(1	0.6	1.55)	(0	0.1	0.05)
GRIDG5.5.20.20	(0.5	1.8	0.5)	(0.5	2.2	0.45)	(0	0.4	0.05)
	(1	0.5	1.5)	(0.75	0.55	0.75)	(0.25	0.05	0.75)
RING10.15	(0	0.8	0.2)	(0.15	0.65	0.25)	(0.15	0.15	0.05)
	(0	2	1)	(-0.25	1.75	1.5)	(0.25	0.25	0.5)
RING2.20	(0	0.8	0.2)	(0.3	0.6	0.2)	(0.3	0.2	0)
	(0	2	1)	(-0.45	0.75	1.35)	(0.45	1.25	0.35)
MOSAIC	(0.6	1.5	0.8)	(0.6	1.55	0.8)	(0	0.05	0)
	(1	0.6	1.5)	(1	0.65	1.5)	(0	0.05	0)
	Avenge Error Rotation: 0.12 Degrees								
	Average Error Translation: 0.28 Pixels								

Table 3

Results of the Automatic Edge-based Registration on the Test Patterns of Figure 2

3.2. AVHRR Data (Image to Map)

The second dataset is a series of 13 512 rows by 1024 columns AVHRR/LAC images over South Africa. Raw AVHRR data are navigated and georeferenced to a geographic grid that extends from -

30.20 S, 15.39 E (upper left) -34.79 S, 24.59 E (lower right). The navigation process uses an orbital model developed at the University of Colorado [15] and assumes a mean attitude behavior (roll, pitch and yaw) derived using Ground Control Points [16]. A map of the coastline derived from the Digital Chart of the World (DCW) is generated for the same geographic grid, Figure 3 shows one image of this sequence superimposed with the map of the coastline. Note that in this case, there is a slight misregistration. The coastline map is used to create a mask, i.e. the area of the input images where the edge features are correlated with the map. Figure 4 shows these features in a [-20,+20] pixel interval around the coastline. Affer correlation, the {rotation,shift,scaling} transfirmation is computed and used to correct the input image. Figure 5 shows the corrected image superimposed with the map and Table 4 shows the registration results using the edge-based method.

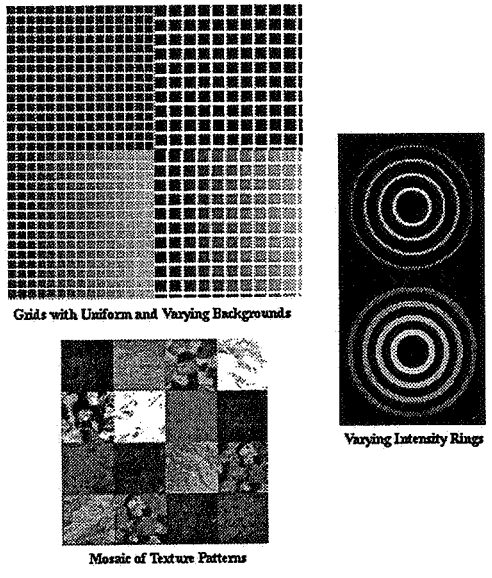


Figure 2
Test Patterns Including Grids, Rings, and a Texture Mosaic

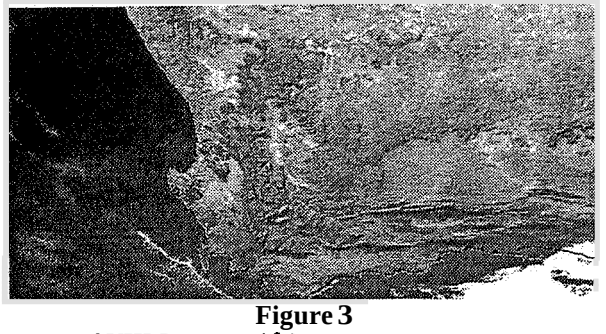


Figure 3
Original AVHRR South Africa Image with Coastline Map

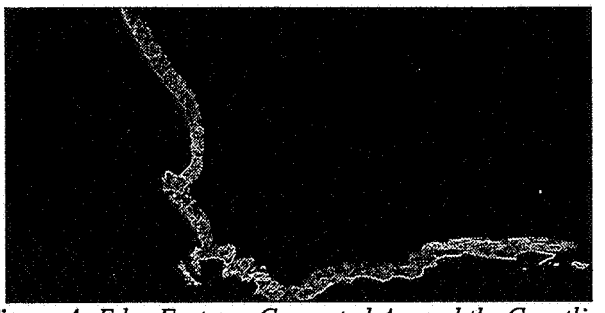


Figure 4 -*Edge Features Computed Around the Coastline*

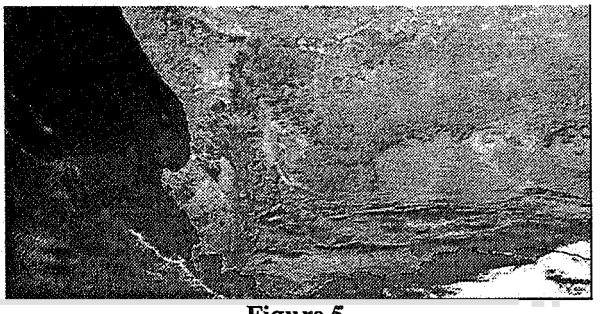


Figure 5
ected AVHRR S Africa Image with
eg s S i =(2, pixels, S s 1)

DATA	sa125	sa126	sa127	sa129	sa1300	sa1311	sa1322	sa133	sa1411	sa143	sa146	sa1488
Iter. Edge Match												
Register to Map	(0,0)	(-1,-1)	(-1,-1)	(-1,-1)	(1,1)	(1,0)	(1,-1)	(0,-1)	(0,-1)	(-1,-3)	(-2,-2)	(2,3)

Table 4
A VHRR Data Registration Results (Edge-Based Method) (All Rotations=0, All Shifts=1 Only Shifts are Indicated)

3.3. GOES Data

3.3.1. Landmarking (Image to Landmark)

Landmark registration has usually two purposes: geo-registering new incoming images, as well as refining the satellite orbit computation and navigation system. Two recent studies dealing with satellite meteorological data show significant contribution in this domain. The first study [3] deals with Meteosat data, while the second one [4] concentrates on GOES data. Both studies consider a shift-only transformation, and obtain sub-pixel accuracy by up-sampling the data. The Meteosat study uses Normalized Cross-Correlation (NCC) on edges of landmarks such as coastlines. The GOES study uses only lakes and islands for landmarks, and evaluates six different matching methods. Among these methods Cross-Correlation (CC) and NCC of enhanced gray levels, as well as Edge Matching are evaluated as performing the best; Edge Matching is the least sensitive to cloud cover, while (N)CC provides a slightly more accurate position estimation. Both studies also utilize a masking of the clouds in order to increase the reliability of the registration. The GOES study also provides a good procedure description and some requirements for the choice of the landmarks.

Our first landmarking tests using our wavelet-based algorithm were performed on a sequence of five successive 128x128 images from the GOES satellite

over the Cape Cod area, "RAW1" to "RAW5", shown in Figure 6. Since we do not currently have a map database available which would be used to perform the landmark registration, we simulated this process by taking as reference "chip" a sub-image extracted from the center of the first image "RAW1" This reference chip is then automatically registered to the other images utilizing the wavelet-based registration algorithm described above.

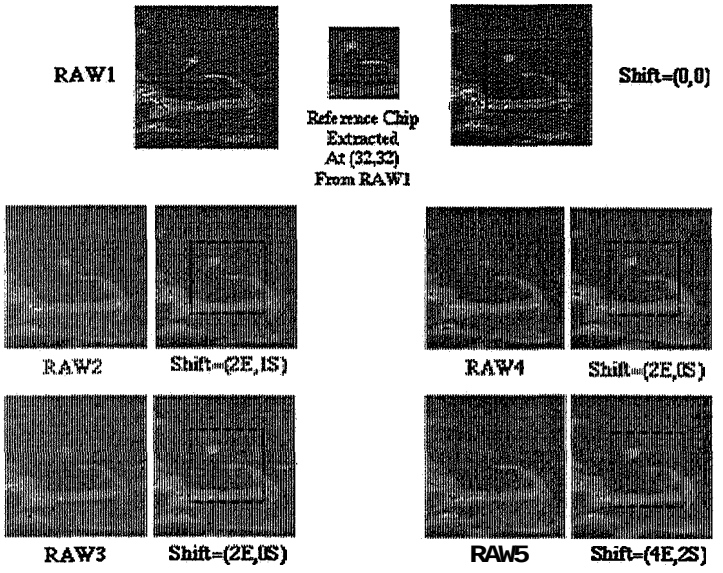


Figure 6
Wavelet-based Registration of a Sequence of Five GOES Images of Cape Cod.

As expected, when the reference chip is registered to the image from which it was extracted, the transformation which is computed is (0E,0S), which means that there is no shift towards the East or South directions. Then, the respective shifts between reference chip and input images are given for each of the input images "RAW2" to "RAW5", and the results are displayed in Figure 1 by a black window located at the shifted position which has been automatically computed by the algorithm. Qualitatively, these results look very satisfactory, and we can notice that even the presence of a small cloud in the reference chip which then disappears in the following images of the sequence does not seem to affect the results. Future work will include the use of a cloud masking as well as determining the maximum percentage of the image which can be covered by clouds without affecting the reliability of the algorithm [14].

3.3.2. Channel-to-Channel Co-Registration (Image-to-Image)

Channel-to-channel co-registration is a calibration-type operation which might not be necessary to perform for every image received by a multispectral instrument. Co-registration might be necessary when an instrument has just been launched, and then will be computed about twice a day (or even. less often): at each of these computations a look-up-table will be updated with the five co-registration parameters (rotation, shift, and scaling). Depending on the number of channels, instead of computing these parameters on every possible pair of channels, a reference channel from each focal plane (or each group of channels) can be selected and co-registration is computed in two steps:
Step 1 reference channels are co-registered to the highest-resolution channel which is visible at that time of the day,
Step 2: every channel of a given group is co-registered to the reference channel of this group.
For example, for the 5 GOES channels, Channels 1 (Visible) and 4 (Long Wave-IR) are taken as reference channels. So the co-registration is performed on the following cascade of pairs of channels:
Step 1 (1,4)
Step 2: (4,2), (4,3), (4,5)

To our knowledge, no systematic study has been attained for the co-registration of remote sensing data, in particular meteorological satellite data. In this case, the issues are somewhat different from the general image registration problem:

- The transformations to be considered are in general smaller. For example, we can assume that:
  - rotations will vary in the interval [-1°,+1°],
  - translation shifts in [-2pixels,+2pixels],
  - scaling factors in [-0.9pixels,+0.9pixels].
- Although the observed areas are about the same in each channel, the visible features can be quite different, which leads to two main differences with the general image registration problem:
  - coastline registration is not always applicable,
  - clouds should be used and not eliminated from the registration process.
- The highest-resolution channel which is utilized as a reference will be different for different times of the day and its spatial resolution will be lower at night.

This set of experiments was performed utilizing multiple channels of a sequence of GOES images taken during 24 hours in two different sectors, "Baja" and "Florida" There are 5 GOES channels with the respective spatial resolutions of 1 km and 4 km. Figure 7a shows the 5 channels of a GOES scene of the "Baja" sector. After basic preprocessing of the data and in order to deal with similar spatial resolution data, a wavelet decomposition of Channel 1 is performed (see Figure 7b). After 2 decomposition levels, the spatial resolution of the decomposed Channel 1 is identical to the resolution of Channels 2 to 5. Then a Sobel edge detection is computed on the compressed Channel 1 and on Channels 2 to 5 (see Figure 7c), and the *edge*-based registration described above is applied to these edge features. We can notice that Channel 3 (Water Vapor Channel) does not present the same original or edge features as the other channels and this remark explains the following results relative to the registration of Channel 3.

We tested 33 five-channel scenes corresponding to the "Baja" sector. According to the two steps described earlier, we chose Channel 1 as the reference channel for the visible wavelength, and Channel 4 as the reference channel for the infra-red wavelength, then the cascade of channels to register is: (1,4), (4,2), (4,3), (4,5). For each of the 33 scenes, three tests of registration were performed:

(E1) All original channels are registered

(E2) Channel 1 was artificially translated by the vector (1,2) pixels, leaving Channels 2 to 5 unchanged,

(E3) Channel 1 was artificially rotated by the angle 0.5 degrees and translated by the vector (0.5,1.2) pixels, while Channels 2 to 5 were unchanged.

Table 5 shows some results of these experiments, detailed for the first three scenes, and then the *accuracy* and the standard deviation (or precision) of the rotation and translation parameters is computed over the 33 scenes for the four pairs of channels. The *accuracy* and the standard deviation of a parameter "a" are computed according to the following formulas:

$$\text{Accuracy (a)} = \frac{1}{33*3} \sum_{i=1}^{i=33*3} (a(i) - true\_a)$$

and

$$\text{StdDev (a)} = \sqrt{\frac{1}{33*3} \sum_{i=1}^{i=33*3} ((a(i) - true\_a)^2) - (\text{Accuracy(a)}^2)}$$

Correlation coefficients, which vary between 0 and 1, are also indicated for each registration. The results show a perfect registration of Channels 4 and 5 with a correlation coefficient close to 1, and a relatively good registration of Channels 4 and 2. As we noticed previously, the features extracted from Channel 3 do not permit a good registration and it is illustrated by very low correlation coefficients (around or below 0.1). In this example, the registration of Channels 1 and 4 seems to present a bias of (0.95,0) in shift that is consistent among the 33 scenes of this sequence. In the absence of ground truth data and with correlation coefficients of average value, no conclusion could be drawn from this result. If additional information was available, such registration could indicate a shift in the geometric calibration of Channel 1 relative to the other channels.

Image	Channels 1/4			Channels 4/2			Channels 4/3			Channels 4/5		
	Rot. (Deg.)	Transl. (Pix.)	Correl.	Rot. (Deg.)	Transl. (Pix.)	Correl.	Rot. (Deg.)	Transl. (Pix.)	Correl.	Rot. (Deg.)	Transl. (Pix.)	Corr.
...101615..												
Original	0.15	0.95,0.35	0.44	0	0.2,0.2	0.65	0.15	-0.85,-0.1	0.12	0	0.0	0.94
R=0 T=(1,2)	0.1	1.9,2.3	0.44									
R=.5 T=(.5,1.2)	0.5	1.55,1.25	0.44									
...101632..												
Original	0.15	0.85,0.1	0.44	0	0.0	0.64	0.35	-1.6,0.65	0.65	0	0,0	0.94
R=0 T=(1,2)	0.65	1.95,2.05	0.44									
R=.5 T=(.5,1.2)	0.5	1.55,1.25	0.44									
...101645..												
Original	0	1.0	0.44	0	0.0	0.65	0.15	-1.1,0.15	0.11	0	0.0	0.94
R=0 T=(1,2)	0	2.2	0.44									
R=.5 T=(.5,1.2)	0.5	1.55,1.25	0.44									
After 33 Experiments:												
Accuracy	-0.06	(-0.72,-0.18)		0	(0.01,0.01)		-0.217	(0.62,-0.19)		0	(0,0)	
Stand. Dev.	0.07	(0.29,0.13)		0	(0.04,0.04)		0.115	(0.85,0.32)		0	(0,0)	

Table 5
Results of Registration Experiments for the "Baja" Sector, Fig. 7a

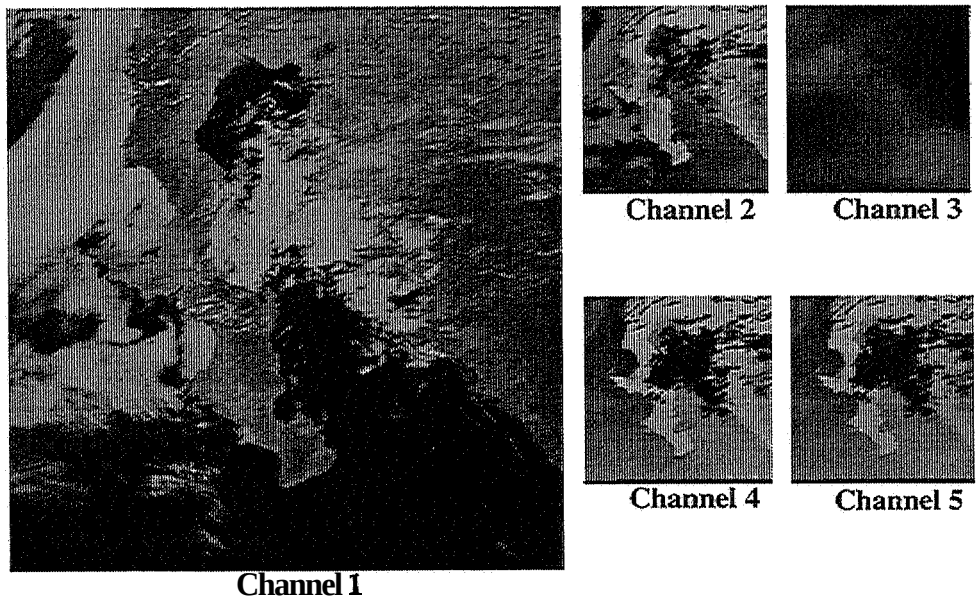


Figure 7a
GOES Scene; "Baja" Sector; Five Channels
(Images reduced for display purposes)

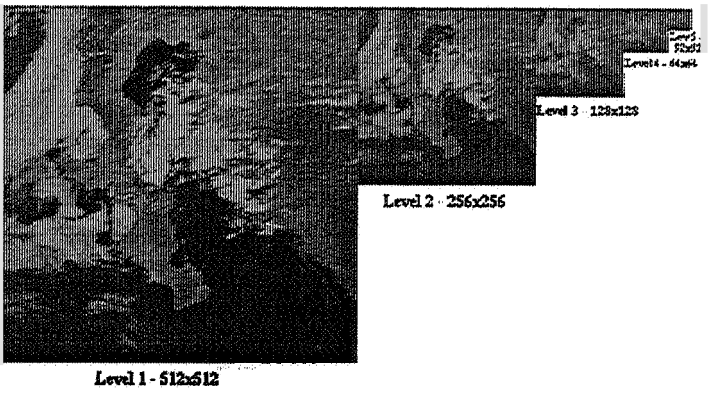


Figure 7b - *Decomposition of Channel 1 shown in Figure 7a*

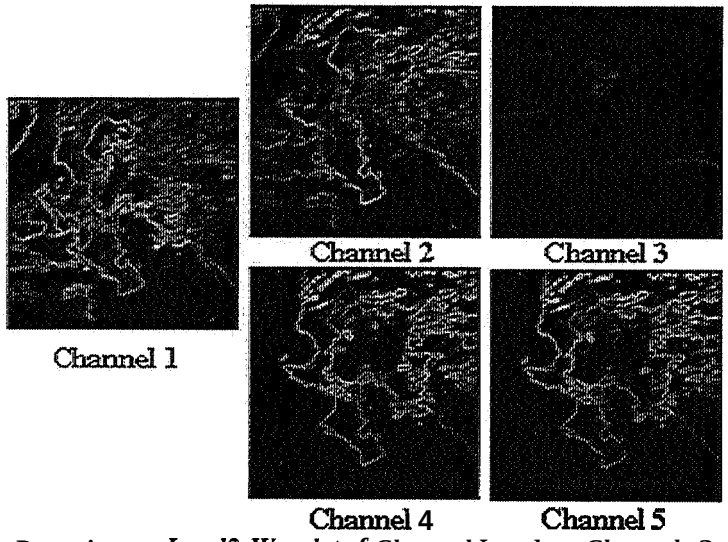


Figure 7c - *Edge Detection on Level2-Wavelet of Channel I and on Channels 2 to 5 of Figure 7a*

Similar experiments were conducted with the successive scenes of the "Florida" sector shown in Figure Sa. In this case, no land features are visible in any of the five channels and this example shows how channel-to-channel co-registration can be performed with only cloud features. Figure 8b shows the edges extracted from the five channels and Table 6 details the results of the registrations performed for the above experiments (El), (E2) and (E3). These few results show good registrations of the five channels

4. Conclusion and Future Work

These preliminary experiments utilizing edge- and wavelet-based image registration are very promising. Still more work needs to be done in the following areas:

- ***Sub-pixel accuracy:*** from the previous GOES and Meteosat studies, we feel that automatic registration schemes based on edge or edge-like features should be able to achieve sub-pixel accuracy Of course a landmark registration at such an accuracy assumes that a map database will be available at a very high resolution (e.g., DMA map) and for a very large number of landmarks. Ifneeded, accuracy can also be improved by interpolating the correlation curve using for example a cubic-spline method.
- ***Robustness of Algorithms*** The robustness of the algorithm will have to be evaluated in hction of variable conditions (such as time of day), especially for the co-registration process, since different channels with a lower spatial resolution will have to be utilized as reference at night. Robustness will also have to be evaluated relative to the issue of cloud occlusion.
- Dependency on ***Initial Conditions:*** The cost of computing the spatial correlation of two images is a function of the number of steps where the correlation is computed. If we perform the correlation of 2 images of size NxM at k different positions, the computational cost is equal to 2*k*N*M floating point operations. So if the initial search for the five parameters is computed in some interval [-n,+n] with a step of 1, the correlation cost is: 2NM*(2*(2n+1)**2+(2n+1)). If the initial search interval increases, i.e., the accuracy of the initial parameters given by the attitude model decreases, then the computational cost of the registration will increase according to the above formula. A larger uncertainty in the initial conditions might also lead to false paths in the search for

registration parameters and therefore decrease the overall registration accuracy

In general, quantitative testing, as well as accuracy and simulation studies will be pursued for the two previous types of registration for the purpose of multi-temporal registration, landmark navigation, as well as channel-to-channel co-registration.

Acknowledgments

The authors would like to thank Bill Campbell and the NASA/ Goddard Space Flight Center's Applied Information Sciences Branch for their support in the development of the algorithms, as well as Jim Tucker, Dennis Chesters, Del Jenstrom, Sanford Hinkal, Rick Lyon, and the GATES project team for providing advise, expertise and data for our experiments.

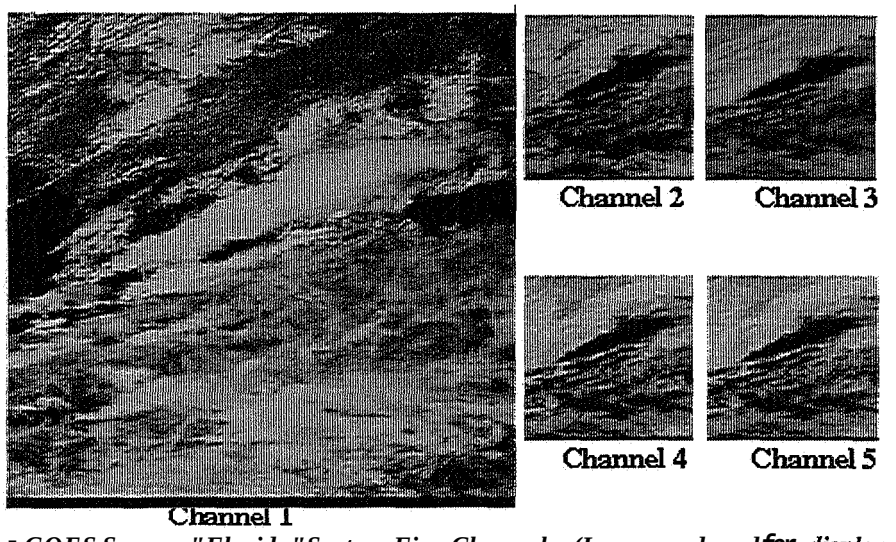


Figure 8a - *GOES Scene; "Florida" Sector; Five Channels (Images reduced for display purposes)*

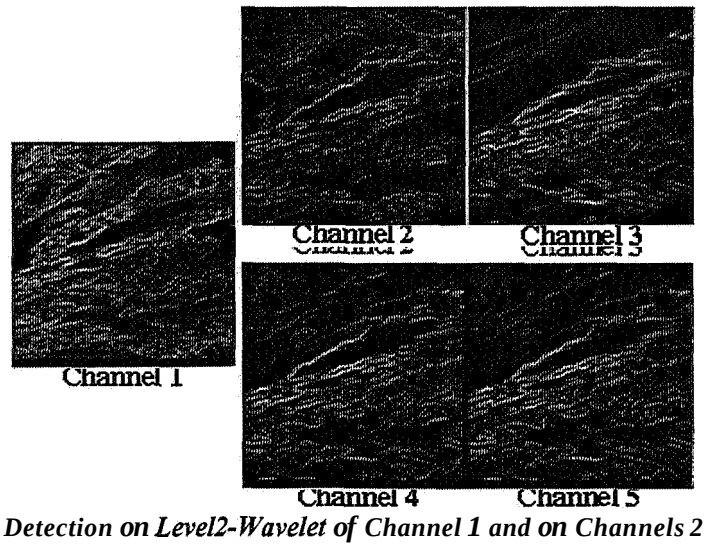


Figure 8b - *Edge Detection on Level2-Wavelet of Channel 1 and on Channels 2 to 5 of Figure 8a*

Image	Channels 1/4			Channels 4/2			Channels 4/3			Channels 4/5		
	Rot. (Deg.)	Transl. (Pix.)	Corr.	Rot. (Deg.)	Transl. (Pix.)	Corr.	Rot. (Deg.)	Transl. (Pix.)	Corr.	Rot. (Deg.)	Transl. (Pix.)	Corr.
Original	0	(0,0)	0.47	0	(-.4,-.4)	0.68	0	(0,0)	0.72	0	(0,0)	0.98
R=0 T=(1,2)	0	(1,2)	0.47									

Table 6 - *Results of Co-registration Experiments for the Scene of the "Florida" Sector (Fig.8a)*

Bibliography

[1] L.Brown, "A Survey of Image Registration Techniques," *ACM Comp.Surv.,* Vol 24, 4, 1992.

[2] J Le Moigne et al, "An Automated Parallel Image Registration Technique of Multiple Source Remote Sensing Data," CESDIS TR-96-182 - submitted to *IEEE Trans.Geosc. and Rem. Sens.*

[3] B Blancke, J.L. Carr, E. Lairy, F Pomport, B. Pourcelot, "The Aerospatiale Meteosat Image Processing System (AMPS)," *I-st Intern. Symp. Scientific Imagery & Im. Proc.,* Cannes, France, Apr 1995.

[4] W Eppler, D. Paglieroni, M. Louie, J. Hanson, "GOES Landmark Positioning System," SPIE Proceedings Vol. 2812, Intern. Symp. on OPtical Science, Engin., and Instrumentation, GOES-8 and Beyond, Denver, CO, August 4-9, 1996.

[5] J.C Tilton, "Comparison of Registration Techniques for GOES Visible Imagery Data," *IRW97,* NASA-GSFC, Oct. 97

[6] W.K. Pratt, "Correlation Techniques of Image Registration," *IEEE Trans. Aerosp. and Electr Systems,* Vol.AES-10, *3,* 1974.

[7] J. Canny, "Computational Approach to Edge Detection," IEEE-PAMI, Vol.8, Nov. 1986.

[8] J. Shen, S Castan, "An optimal linear operator for step edge detection," *Graphical Models and Image Processing,* Vol 54, No. 2, 1992.

[9] R.F Cromp, W J. Campbell, "Data Mining of Multidimensional Remotely Sensed Images," Proc. *2nd Int. Conf. on Information and Knowledge Management,* Wash.DC, Nov 1993.

10] J.LeMoigne,"Parallel Registration of Multisensor Remotely Sensed Images Using Wavelet Coefficients," *SPIE Aerosense,* Apr 1994.

[1 13J Le Moigne, "Towards a Parallel Registration of Multiple Resolution Remote Sensing Data," *IGARSS'95,* Firenze, Italy, July 1995.

[12] D.I. Bamea, and H.F Silverman, "A Class of Algorithms for Fast Digital Registration," *IEEE Trans. on Computers,* Vol. 6-21, 179-186, 1972.

[13] D. Mount, N. Netanyahu, J LeMoigne, "An Efficient Algorithm for Robust Feature Matching", *IRW97,* NASA-GSFC, Oct. 97

[14] M. McGuire, H. Stone, "Techniques for Multi-Resolution Image Registration in the Presence of Occlusions," *IRW97,* NASA-GSFC, Oct. 97

[15] G W Rosborough, D.G. Baldwin, W.J. Emery, "Precise AVHRR Image Navigation," *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 32, No. 3, May 1994.

[16] D. Baldwin, W Emery, "Spacecraft Attitude Variations of NOAA-11 Inferred fiom AVHRR Imagery," Int. J. of Remote Sensing, Vol. 16, No. 3, 531-548, 1995.

Session V

Applications to Satellite Sensors

Automated Construction of Large-Scale Electron Micrograph Mosaics

Robert C. Vogt, John M. Trenkle, Laurel A. Harmon

ERIM International
P.O. Box 134001, Ann Arbor, Michigan, 48113-4001

ABSTRACT

A series of processing steps is described for the automatic acquisition, pre-processing, registration, and mosaicking of large-scale transmission electron micrograph images (> 1 Gigapixel). This work was carried out as part of a final-phase clinical analysis study of a drug for the treatment of diabetic peripheral neuropathy, to determine if the compound under study could reduce or reverse the destruction of myelinated fibers. In order to verify the existence of rare, regenerating nerve fibers, it was necessary to image at high enough magnifications in order to see basement membranes, only 50 nanometers in width.

More than 500 nerve biopsy samples were prepared, digitally imaged, mosacked, processed, and reviewed, For a single sample, more than 1000 electron micrograph frames were typically acquired, each composed of 1.5 megapixels, for a total of between 1 and 2 gigabytes per sample. These frames were automatically registered and mosaicked using a two-stage search process, forming a single virtual image composite that was later used to perform automatic cueing of axon and potential regenerative clusters, as well as review and correction marking by qualified neuroanatomists. Registration was non-trivial because the placement accuracy of the computer-controlled stage allowed for motion errors of up to 100 pixels (1 75 microns) in any direction. In addition, edge artifacts due to the effects of the electron beam scanning process, also impeded accurate registration and required mitigating preprocessing steps to be carried out on each frame.

The first stage of the registration search process was accomplished via a full binary correlation at reduced resolution, while the second stage used a tree-structured search on filtered binary images that emphasized larger, dark structures. This approach guaranteed that registration results could be generated quickly enough to meet the high throughput demands of the project. Overall, the registration effort was successful except for "bland" regions of the samples, lacking any clear structure, where it was nearly impossible to match edges of adjacent frames. The paper discusses potential improvements to the registration process which might reduce the impact of such errors.

The effort described here demonstrated a new, entirely digital capability for doing largescale electron micrograph studies, in which all of the relevant specimen data could be included at high magnification, as opposed to simply taking a random sample of discrete locations. It offers the possibility for new types of studies involving electron microscopy that would not have been possible prior to 1995, due to the convergence of technologies (mass storage, faster processors, accurate computer-controlled stages, and a digital camera), that were used here to achieve the result, including the acquisition of more than 500,000 micrograph frames, and processing of over 800 gigabytes of data, in less than 6 months.

Keywords: cluster counting, computer-aided image analysis, diabetic neuropathy, digital mcroscopy, electron microscopy, image registration, image mosaicking, nerve fiber density, nerve regeneration

1. INTRODUCTION

This paper describes a suite of image processing algorithms which were developed as part of a clinical study, on the effect of a particular drug in the treatment of diabetic peripheral neuropathy. The portion of the study described here

examined the effect of the drug on peripheral nerve tissue, through the use of electron mcroscopy (EM). The purpose of the algorithms was to make it possible to capture images at high magnification, mosaic them into a single, very large composite, and then provide tools to aid neuroanatomists in reviewing this volume of imagery.

The anatomical changes that were being studied here included an increase in the percentage or area density of regenerative axon ***clusters,*** or an increase in the area density of the axon ***fibers*** themselves, particularly the smaller ones (less than 5-6 microns in diameter). To make this assessment, a trained neuroanatomist ***reader*** had to review the entire cross-sectional area of complete nerve fiber bundles, or ***fascicles,*** such as the one shown at low resolution in Figure 1 First, the reader was required to identify all of the axon fibers (donut-like objects) in the facsicle, which could number as high as 1000-2000. Second, he or she was required to identify all of the regenerative clusters, which are groups of two or more small axons that are completely surrounded by a single ***basement membrane.*** These clusters are considered to provide direct anatomical evidence of nerve regeneration, and thus are critical to definitively demonstrating the utility of the drug for diabetic patients. Unfortunately, such clusters are also relatively rare, and difficult to verify, due to the narrow width of the basement membrane which must surround them.

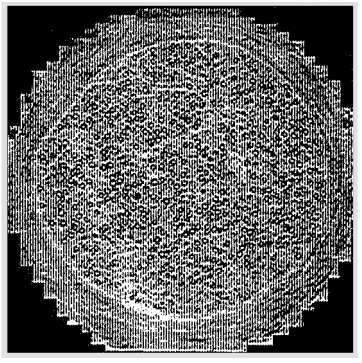


Figure l — Example composite image of a nerve fascicle

1.1 Data volumes

The use of cluster counting to measure nerve regeneration is a relatively new development, and up to now, very few studies have been performed based on this technique. The main difficulty in taking this approach is the sheer volume of data necessary to collect and review. Because the basement membranes are extremely thin (around 50 ***nanometers*** wide), in order to see them reliably, we needed to have 2-4 pixels across their width. This translates to an imaging resolution of around 60 pixels per linear micron. Since the fascicles themselves are typically more than 0.5 millimeters in diameter, then, to image an entire fascicle, we would need more than 30,000 pixels across the diameter, or around 1 gigapixel of rectangular area. In actual practice, because overlapping frames were required for registration purposes, and we had to image some tissue outside of the actual fascicle area, raw datasets for a single sample were typically 1-2 gigabytes (1 byte per pixel), and sometimes as high as 2.5 gigabytes. At a nominal screen resolution of 72 dpi (one pixel per typographic "point"), this corresponds to an image that is between 30 and 60 feet in diameter! Over 500 samples of this type were imaged in the course of this study, for a total of nearly *I **terabyte,*** each, of both raw and processed (mosaicked) data. To acquire all of these samples, more than 500,000 electron micrograph images were collected, making this one of the largest electron microscopy efforts ever attempted.

In the past, small studies of this type were carried out using film-based electron micrograph systems. The films were taken, developed, and then literally ***pasted together*** manually to form a gigantic image, a few yards across, which would be reviewed by an anatomist using a magnifying glass, for counting axons and identifying clusters. For the current study, with so many samples being acquired, and the need to complete all of the work within a year, this approach would not have been feasible. Instead, the imaging team for the study developed an entirely ***digital*** acquisition and processing system. Images were acquired using a digital camera connected to the electron microscope. The microscope stage was automatically controlled so that all of the 1000 or more frames for a given sample could be acquired without human intervention, once the set-up operations were complete. Re-processing, registration, and

mosaicking of the raw EM image frames were done using an entirely automated process, as was cueing of the axons and potential clusters. Review and marking of the sample imagery by the neuroanatomists was also aided by the development of a digital reviewing station, which made it easier for the readers to navigate around the large datasets, and to record their observations.

1.2 Processing stages

In the sections that follow, we will describe some of the processing stages for an individual sample, from the time the nerve biopsy is taken, through image acquisition and processing, until the statistics based on the review by a neuroanatomist are collected (see Figure 2). Our focus in this paper will be primarily on the acquisition, pre-processing, registration, and mosaicking steps.

When the nerve biopsy is taken, there are a number of sample preparations steps that occur, prior to the time that the sample is imaged. Some of these occurred at the clinic or hospital which performed the biopsy, and others were carried out at the University of Michigan, where the study was conducted. Following this preparation, the sample was automatically imaged on the electron microscope, and the data files were transferred to another computer for pre-processing, registration, and mosaicking. When the composite, or "virtual" image product was complete, the data was transferred to a third computer station where automatic axon and cluster cueing was carried out. The outputs from this process, along with the composite image data, were then sent to a sample reviewing station, where a neuroanatomist reader would examine the sample, and mark axons and regenerative clusters. These marks formed the basis of the statistics that were analyzed to evaluate the efficacy of the drug.

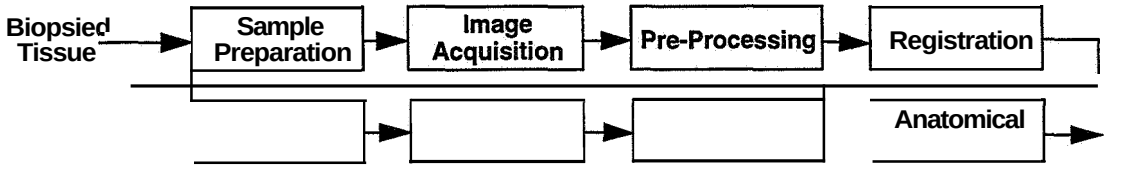


Figure 2 —Block diagram of processing stages

2. IMAGE ACQUISITION

Image acquisition for the study was performed by the University of Michigan (U of M) Department of Anatomy and Cell Biology. The following section briefly describes the manner in which samples for the study were prepared, and how they were imaged on the electron microscope. The digital equipment used is also described, as are the set-up procedures employed to ensure reasonable uniformity within and between the acquired EM datasets. Many tests were carried out during the initial stages of the program, to arrive at a set of procedures and operating parameters that would result in high quality images for the subsequent image processing and neuroanatomist review tasks.

2.1 Sample preparation

Samples for this study were obtained from sural nerve biopsies taken from different legs, before and after a 1-year treatment period. The sample tissue was fixed and embedded in a plastic material at the site of the biopsy, and then shipped to the U of M Diabetes Clinic, for subsequent handling. Here the samples were divided into blocks, and a single fascicle was chosen as being the most acceptable for EM imaging, based on a number of criteria. A thick section of the sample was prepared next, the area of the fascicle was measured using light microscopy, and the section was then sent to the U of M Department of Anatomy and Cell Biology for EM imaging. Here it underwent further preparation steps, including thin-sectioning, staining with heavy metals that would absorb electrons, and carbon-coating of both sides for stability. The resulting thin section was then placed on an open slot grid for imaging within the microscope.

2.2 Imaging Equipment

The transmission electron microscope (TEM) used was a Philips CM100. The stage was a computer-controlled Phil-

ips Gompustage, providing 0.5 microns positional accuracy (standard deviation) for a single move, or 0.1 microns accuracy (mean) over many repeated moves. Images were acquired with a Kodak 1.6Megaplus camera, that was attached to the TEM with lenses in order to image an internal phosphor plate (24x36mm), and also connected to an ITI frame-grabber board mounted in a Pentium-based PC. Data collected during the day was stored on the PC hard disk, and then shipped via FTP to a workstation processing node, over a local network connection.

2.3 Burn-in procedure

One major difficulty with performing a study of this type, where adjacent electron micrographs have to be registered and mosaicked together, is that, unlike a light microscope or light-based camera, the electron beam can significantly interact with the tissue being imaged (particularly with thin sections), causing it to heat up, become slightly distorted, or lose some material via "etching" We found, dunng initial testing, that these effects could make it difficult to register adjacent image pairs. To mitigate these effects, we tested the idea of "pre-burning" the sample by exposing it to the beam at lower magnification, prior to the final acquisition. This step significantly reduced the distortions appearing in the final acquisition, and made it easier to register the image pairs more reliably. During production processing, the burn-in step was performed at 4X lower magnification, to reduce acquisition time.

2.4 Acquisition procedure

For carrying out the actual image acquisitions, a very specific process was defined for setting the intensity levels, the KeV value, and other parameters of the electron microscope, as well as for setting the position of the sample, setting the focus, and obtaimng a background image—all with the purpose of trying to guarantee reasonably consistent imagery over the course of each sample, and over the course of the study as a whole. For example, to attempt to ensure that the illumination on each sample was consistent, a live background image (no sample present) was taken using a fixed exposure time (250ms). The response from this setting was monitored in real time via a histogram, and the illumination was tuned to move the histogram distribution to a particular position on the display. Then, the sample tissue was introduced, and the exposure time was then set to a longer fixed period of 500 ms.

Images were acquired at 4600X magnification within the scope, corresponding to 15,000X to 20,000X at the level of the computer screen, or about 57 pixels per micron on the sample. The spot size setting was 3, and the KeV setting was 60, which represented a compromise between obtaining better image contrast versus longer life from the EM filament and vacuum system, When all of the set-up steps were completed and the acquisition area was defined, stage movement and imaging proceeded entirely automatically, typically over the course of 3 or 4 hours. Successive EM frames were acquired in a raster scan order, starting from the top of the sample and moving down, over an implied grid of equally-spaced rectangles. (We did not actually capture all of the frames within the grid—the software was designed to only acquire those frames within an inscribed circle or ellipse inside the bounding box, saving both disk space and acquisition time). Each frame was 1024 rows by 1500 columns, and successive frames were nominally overlapping by about 180 pixels in all four directions. The frames were acquired in raster order, rather than an S-scan order, so that we could be sure that any sliding or uncontrolled motion in the stage would be consistent from one acquired row to the next. This greatly aided the subsequent registration process. As an additional means of controlling the stage motion and making it as accurate as possible, the stage control program was modified to move the stage 3 times for each single positioning step to acquire the next frame: First, it was moved 1.5 times the desired distance and direction, then back 1.0, then forward 0.5 This was done to correct for backlash in the stage positioning mechanism, and it did result in reducing the accumulated mohon error, based on early tests that we performed.

3. PRE-PROCESSING

Once a complete set of raw image frames had been collected for a sample, the next step was to perform pre-processing operations, both in preparation for the image registration process, as well as to provide the best possible image contrast for later review by the anatomists, and for the automated cueing algorithms. Normally the process here would be to divide each frame by a background frame (taken without the sample present), in order to remove spatial response vanations in the phosphor from the images. Following this division, the frames would be linearly constrast-stretched over 0-255 to enhance visibility of the structures. One of the major difficulties at this stage, however, was caused by the effect of the electron beam on the sample tissue, as mentioned earlier. Because the beam necessarily

covered a larger area than the frame being imaged, it would disturb or etch tissue that had not yet been imaged in the raster sequence. The net effect of this was to cause the lower-right portion of each frame (farthest from the beam in the raster scan) to be darker than the upper-right portion, since the latter was more impacted by the spurious effects of the beam.

3.1 Raw frame averaging

When images acquired in this manner were initially mosaicked together, it gave a "tiled" appearance to the overall composite image (See Figure **3).** In addition, this effect of the beam also interfered with the registration process, since it meant that the right (lower) side of one image in an adjacent pair was not entirely comparable with the left (upper) side of the adjacent image, with which it was supposed to match. This degraded the appearance of the image and also created serious difficulties for the registration process. To remove this problem, we took the step of first *averaging* together all of the raw image frames for a given sample, in order to get a "mean" image that would take into account both the phosphor variations in the background, as well as the etching side effects of the beam. Examination of dozens of these mean images (each derived from hundreds or thousands of different frames for each sample) showed that they were devoid of any obvious structural content, and did accurately represent the combination of the beam effects with the more subtle phosphor variations.

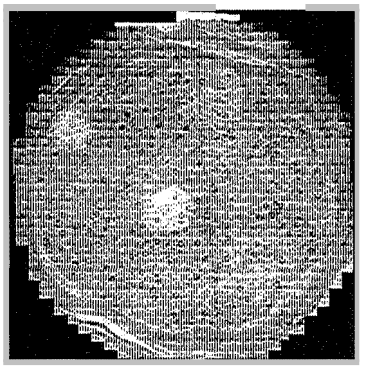


Figure 3—Mosaicked composite showing tiling effects of electron beam

3.2 Global floating point histogram

We next used this mean image as a revised "background image", and divided it into each frame. Instead of generating a new 8-bit frame immediately, however, we saved the floating point division values into a high-precision "floating-point histogram" (retaining fractional values down to .001), for subsequent analysis. This floating-point histogram, collected over all of the frames for the sample (over 1 gigapixel), allowed us to first of all ignore regions where no tissue was present (i.e., with a ratio >=1.0), and then to find the upper and lower 0.5 percentiles of the remsuning points. These percentiles would be used for contrast stretching the data to 0-255 on the output. To carry out this operation, we prepared a 256 x 256 entry, 8-bit output look-up table, so that we could map each pair of incoming frame and mean image pixel values ("FrameVal" and "MeanVal", below) to the desired 8-bit value ("OutVal"), using the following formula:

$$OutVal = MAX[\ 0, MIN[\ 255, ROUND[\ ((FrameVal / MeanVal) - F_{0.5}) / (F_{95.5} - F_{0.5}) * 255.0\ ]]] \qquad (1)$$

FrameVal and MeanVal vary independently over 0 to 255, and $F_{95.5}$ and $F_{0.5}$ are the values corresponding to the upper and lower percentile bounds, respectively. Using this look-up table, we could quickly remap all of the original image frames to their desired 8-bit, constrast-stretched values, without the loss of precision that would have occurred had we separately performed the division into 8-bit images, and then clipped and contrast-stretched the results, again into 8-bits. With the approach described, we retained all of the precision that we possibly could in producing the 8-bit processed image frames, which then formed the basis for all of the steps that followed.

4. REGISTRATION

After pre-processing each frame for a single sample, the next step was to register all of these frames together. As mentioned earlier, based on the programming of the Compustage coordinates, we knew that the frames were collected on a rectangular raster or grid, and that adjacent pairs of images should overlap nominally by about 180 pixels in each direction. This amount was chosen because the error in the Compustage for any given move was about +-0.5 microns, or around +-30 pixels (standard deviation) at the magnification we were using. Since we were acquiring typically more than 1000 images in a 3 or 4-hour acquisition session, we needed to account for potential errors of more than 3 standard deviations, or roughly +-100 pixels. Even in the worst cases, we still needed to have enough image width left to perform the registration itself. We experimentally determined this width to be minimally 60 pixels, and preferably at least 90.

Essentially then, for any given pair of adjacent images, we needed to search a space of more than 200 by 200, or 40,000 potential offsets. (We only considered translation errors, not rotations or more elaborate warpings, in part for simplicity and feasibility reasons, and in part because the motion of the stage itself was very linear over these short distances of around 20-30 microns). Because it would have been time-prohibitive to actually test every possible offset, we developed a 2-stage approach, first obtaining a reasonable search area or starting point with low resolution data, and then performing a full resolution tree-search to find the precise match position. This strategy kept the number of computations required to register each pair of images to a reasonable level, something closer to a few hundred tests per image pair. The same registration algorithm was used both for matching images in the horizontal direction, as well as the vertical (though we did not register all vertical pairs).

4.1 Low resolution binary correlation

In the first stage of image pair registration, we selected equal-sized subsections from each image and downsampled them by a factor of 8 in each direction. These subimages were both binarized by thresholding at their median levels (50th percentile), under the assumption that this should turn on roughly corresponding sets of pixels in both images, and provide maximum matching accuracy in an information-theoretic sense. Next, a full binary correlation was performed over these small chips; i.e., we computed the value of: (#11+#00) / (#11+#10+#01+#00) over all legitimate translations. (Counts were taken only over the portions of the two chips wluch overlapped, in each case). The peak correlation was used as the starting point for the more elaborate full resolution search, which is described next. This downsampled binary correlation algorithm rarely failed except in very pathological cases, and even when it did not find the "best" location, it was usually close enough for the search algorithm to find the correct registration point.

4.2 Full resolution binary tree search

Given the nominal search starting point found by the low resolution algorithm, we then began a hierarclucal, or "spiraling" tree search, at full resolution (see Figure 4). The starting point was scaled up to full resolution coordinates, we cut out equal-sized subimages from each image that would fully cover the search region, and these subimages were also binarized by thresholding at their medians. They were then filtered morphologically to remove fine structures, so that the registration could be based on more reliable, larger structures. The principle of the search process was as follows: At a given stage, we would compute a score for the current best point, as well as for 8 other points at equal chessboard distances (45 degree increments) away. The point with the best score would then be chosen as the new center, and the same process would be repeated, but with the distances to the 8 other points cut by half. After several rounds, the search process would be complete, and the resulting center point would be returned as the registration match point. To bias the scoring toward matching of the darker structured regions, we computed a score of (#11) / (#11+#10), for each translation, instead of the usual binary correlation score (again, only over the overlapping portions of the subimages). This search strategy quickly zeroed-in on the desired match point (in most cases). While it was not 100% efficient in terms of dividing up the search space (it allowed for multiple paths to the same final result), we preferred to have it work this way as a hedge against cases where anomalies with the match score might cause the search to temporarily go in the wrong direction.

Overall, the registration algorithms described above generally performed well, but they assumed that there would be enough reliable structure present in the limited overlapping regions of adjacent images to guarantee a reasonable

match. While this appeared to be true for the initial set of images that we used to develop the algorithm, during production processing it became clear that this was not always the case. For example, early on we found that the collagen fiber microstructure in the images, which we initially thought would allow for detailed matching, often become slightly distorted due to the electron beam interaction, and acted instead as simply a repetitive texture that actually impeded good registration, forcing us to actively eliminate it from the process. A more general warping process, instead of simply translations, might have overcome this problem, but would have driven up the computational cost substantially. Later, we also found a number of cases where the overlapping region between two frames would fall within a bland area of the sample (having almost no structure), and this led to numerous registration failures. While we have subsequently defined ways to reduce these kinds of problems, (e.g., by registenng all frames in both the horizontal and vertical directions), once the production portion of the study was well underway, we were not allowed to modify any of the processing steps, so as to prevent possible biases from being introduced into the results.

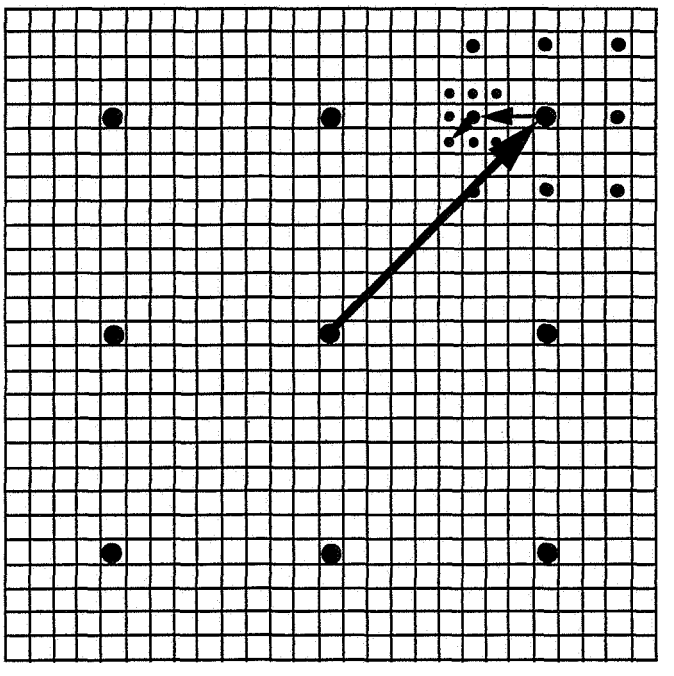


Figure 4—Diagram of tree-search registration algorithm

4.3 Row registration

While the above algorithms were used to register each successive pair of adjacent frames in a row, for registering between rows we adopted a slightly different approach, In principle, if all of the frames in a single row have already been registered, they are, in a sense, "locked" together, so that to register two successive rows, we should simply have to find a single reliable correspondence between them, and then shift all the frames accordingly, In practice, the horizontal registration process was not 100% reliable, and so to guard against these types of errors, we performed pairwise registration in the vertical direction between the 7 central frames in each row, and then used the best match score among these seven to register the successive rows. This algorithm almost never failed to reasonably match the rows, even when serious errors had occurred in the horizontal, or pairwise, registration process.

5. MOSAICKING

When the pairwise and row-wise registration steps were complete, the next stage in the overall process was to mosaic the raw frames together into a composite "virtual image", where all of the image tiles would be adjacent but ***non-overlapping,*** and of a convenient fixed size. (We decided to store the mosaicked result as a set of 1K by 1K tiles, rather than as a single image of 1-2 gigabytes, in order to make it easier to move, handle, and display the processed datasets.) There were two parts to the actual mosaicking process—first to put all of the locations and relationships of the registered raw frames into a single, global address space, and then to use this information to define adjacent but

disjoint tiles that were each 1K by 1K in size (See Figure *5).*

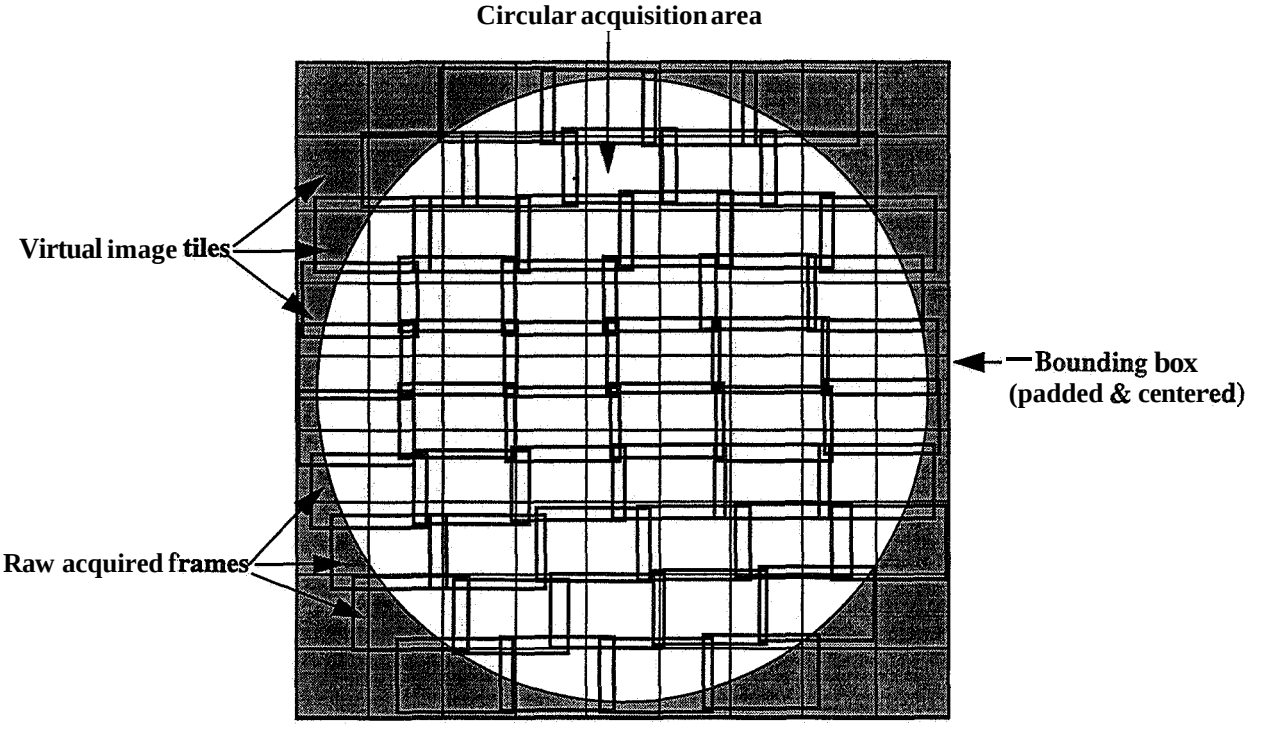


Figure 5—Schematic of mosaicking process

5.1 Global address resolution

To perform the global address resolution process, a tree structure was created with all of the rows of the acquired image frames at the top level, and all of the images within each row as ordered children of that row. This structure was used to retain the coordinates of the upper left comer of each acquired frame, with respect to progressively larger objects, through several "rounds" of address resolution. First, during pairwise registration, the coordinates for each frame represented the position of its upper-left comer with respect to that of the image to the immediate left. In the next pass, upper-left coordinates of each frame in a row were recomputed with respect to that of the leftmost frame in that row.

Next, dunng row registration, as the relationship between each successive row was determined, the upper-left coordinates of the frames below the first row were all updated to reflect the position of the frame with respect to the top-left frame of the first row (i.e., the entire acquisition sequence). At this point, the position of all of the frames was known with respect to the same origin.

The remaining steps before generating the virtual image mosaic were to find the bounding box of all of the frames, and then pad this box to multiples of 1024 pixels, at the same time centering the image data within this area. The bounding box was easily obtained, based on the fixed size of the frames (1024 rows by 1500 columns), and their known relationship to the top-left frame origin. A new origin was then computed as the top-left comer of the bounding box. Next, the size of the bounding box was padded out to the next whole multiple of 1024 pixels in each direction, to make it easy to cut out the 1K by 1K tiles. This larger box was then shifted up and left to center the circular image data within it, and finally the coordinates of every frame were updated with respect to the origin of this padded and centered bounding box The resulting tree data structure with these final coordinates was then saved, as the permanent record of the relationship between the raw acquired EM frames, and the processed virtual image composite tiles.

5.2 Virtual image generation

To produce the individual tiles of the composite virtual image, we next had to find all of the raw frames which intersected a given tile, and determine which subimage portions of these images were involved. For a given tile, (including bordering tiles with no image content), anywhere from 0 to 9 image frames might intersect. Since we only considered translations between frames in the registration process, the intersections were all rectangles of different sizes, In some areas, of course, 2 or 3 images might overlap the same tile area, so we had to decide how to resolve such conflicts. While we had initially contemplated averaging such areas together, initial experiments showed that this approach produced very blurred, complicated images in cases where registration had not precisely matched the subregions To avoid this effect, and produce more comprehensible images, we instead elected to *overwrite* each subarea, based on the raster order of the corresponding frames. This gave priority to the lower-right frames, a somewhat arbitrary but consistent choice. The advantage of this approach is that, while it leads to discontinuities in the images when registration is imperfect, at least the result is more understandable and less confusing to a reader, than the blurred ***melanges*** created by the averaging approach. Often the reader could figure out what had happened visually, and thus could still correctly evaluate the area of the image where the misregistration had occurred — something that was important for obtaining accurate statistical results.

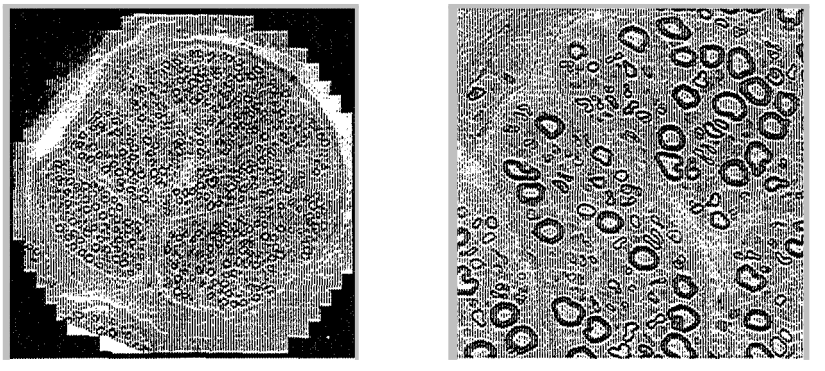


Figure 6—Example lo-res image composite and 4X magnified subregion

The final step in virtual image generation was to generate what we called the ***low resolution,*** or "lo-res" image. This was a single composite image, formed by downsample-averaging each virtual image tile by 8 in both directions (that is, each 8 by 8 block of pixels was replaced with a single average value), and then placing all of these reduced image chips into a single image frame, which typically varied in size from 10 to 40 megabytes (roughly 4-7 feet in diameter at 72 dpi screen resolution). This lo-res image was used both by the cueing algorithms for finding axons and potential clusters, and by the neuroanatomist readers for verifying cues, making additional marks, or noting potential cluster sites or other locations that needed to be reviewed at high resolution. Figure 6 illustrates an example of such a lo-res composite image (greatly reduced), and a 4X magnified portion of it that better illustrates the resolution used for finding and marking axons. The latter image gives an idea of the distribution of sizes and shapes of different axons and their surrounding myelin sheaths, within a single fascicle.

In Figure 7, we have taken a small portion from Figure 6b, and magnified it 8 times. This mage represents the full-resolution data that we have available for any part of the composite virtual image. In fact, Figure 7 is actually a single virtual image tile (1K by 1K), which would normally fill a computer screen. In this image we see both large and small axon fibers, along with a potential cluster (in the middle), and a couple of Schwann cells (with textured interiors) toward the upper left. Note that the Schwann cell along the top left of the image is about the same size as the small axon fiber to its right. In light microscopy (LM) of this kind of nerve tissue, it can be very difficult to distinguish such Schwann cells from axons, based on the stains that are typically used, so they have been a major source of false alarms in counting small axons using LM The grey texture that appears in the background of Figure 7 are collagen fibers, which can actually be seen individually, at full screen magnification. In this image, we can also see the fact that the myelin-surround of the axon fibers vanes in darkness or stain absorption, something that the axon cueing algorithms had to account for. Finally, regarding basement membranes, the pair of small axons shown in the center of this mage are not surrounded by a basement membrane and thus do not represent a true cluster. However, to under-

stand the size of the membranes and the difficulty of verifying them, consider the indented axon near the middle of the image. If we look at the lighter grey structures that sit within the outer concavities of the myelin of this fiber, we can barely see very thin membranes along their outer edges. These membranes are about the same width as a basement membrane, that would have to surround both of these center axons, for them to be considered as a cluster.

In closing this section, we should also point out that the entire set of processing steps mentioned so far, staxting with preprocessing, then registration, mosaicking and virtual image generation, take about 3-8 hours to complete on an HP workstation running at 125MHz, depending on the size of the sample. Most of this time (well over half) is in the pairwise registration and search process, with the rest being divided about equally between the preprocessing calculations and the virtual image generation steps.

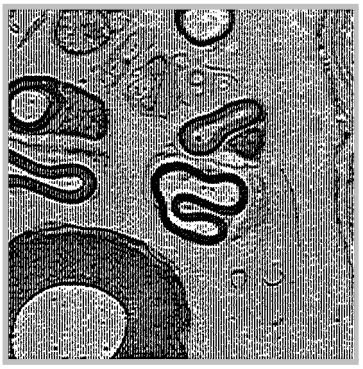


Figure 7—Full-resolution tile of same example showing possible cluster

6. CONCLUSION

In the preceeding pages, we have described a suite of image processing, registration, and mosaicking programs that were developed to support a large-scale clinical study of the effect of a given drug, on diabetic peripheral neuropathy. These programs made it possible to consider use of the newly-developed "cluster counting" approach to verifying nerve regeneration, in conjunction with moderately high magnification electron mcroscopy. What this effort has shown, more than anything else, is that it is now possible to perform large-scale electron microscopy studies of this type, based on constructing very large digital composite images from many hundreds or thousands of individual electron micrograph frames.

7. ACKNOWLEDGEMENTS

This work was supported by Hoffmann-La Roche Limited of Canada, an affiliate of Roche Holding Ltd (Switzerland), through a contract with the University of Michigan Medical School, Department of Anatomy and Cell Biology. The work was carried out as a collaborative effort between the sponsoring organization, the Environmental Research Institute of Michigan (ERIM), and the University of Michigan Department of Anatomy and Cell Biology (School of Medicine), Department of Biostatistics (School of Public Health), and Diabetes and Endocrinology Clinic (Medical Center). We wish to thank all of the participants for their important contributions to this effort.

Two-stage registration for automatic subtraction of intraoral in-vivo radiographs

*Thomas M. Lehmann**, *Walter Oberschelp*†, **Klaus** *Spitzer*

Institute of Medical Informatics
†Institute of Applied Mathematics and Computer Science
Aachen University of Technology (RWTH), Aachen, Germany

Abstract

Digital subtraction radiography has shown to be the best non-invasive tool for diagnosing small changes of internal structures. In dental radiology, subtraction techniques are useful not only for caries diagnostics but also to assess guided tissue regeneration, or to detect bone destruction in implantology. Nevertheless, those techniques require perfect geometrical match of the radiographs to be subtracted. Since mechanical devices such as individual bite blocks are cumbersome to use in clinical routine, automatic image registration can provide a convenient means to detect changes. In clinical routine, more and more X-rays are acquired direct digital by using charge coupled devices or storage phosphor plates, which enabele further processing like automatic subtraction.

Intraoral X-ray examinations are performed by positioning the tube about 20 or 30 cm in front of the patient while the film or digital sensor is touching the maxillo or mandible from inside the patient's mouth. The focus dimensions are usually smaller than 1 mm^2 which allows the approximation that the X-ray beam emenates from a point source. Assuming fixed positions of the X-ray tube and the patient, and a sensor that may be moved and rotated in all directions of the three dimensional space, the geometric transform between two radiographs of the same oral region is given by perspective projection. Because the distance between the source and the object is much larger than the object-target distance, the model of perspective projection could also be used to approximate free-hand radiology Changes in image contents caused by different projections are considered as noise.

Eight parameters must be determined for the perspective back-projection of the subsequent radiograph into the geometry of the reference X-ray. Perspective invariants which allow the determination of the total eight parameters are still unknown. Therefore, the RST-movements: rotation, scaling, and translation are detected in a first stage using the Fourier-Mellii invariant. Calculating the Fourier power-spectra, translations could be decoupled from scalings and rotations. The latter movements are transformed into shifts by mapping cartesian to logarithmic polar coordinates. Cepstrum techniques or phase-only matched filters are reliable methods to register pure translations. They are applied to the detection of rotations and scalings in the Fourier-Mellin domain. After re-rotation and inverse scaling, those methods are also used to determine the remaining translation.

The pre-registration allows the linkage of corresponding landmarks in both radiographs. The edges of teeth, fillings, and implants as well as the spongious structure of bone enables the automatic localization of landmarks. In the second stage, the coordinates of those landmarks are used to determine the perspective parameters for fine-adjustment. At least four corresponding points must be located in both radiographs. Nevertheless, the registration results are improved if more than four points could be labeled. Then, the over-determined system of equations is solved using the least-squares method which is efficiently implemented by a block-wise Cholesky factorization The minimum residual determines the over-all exactness of registration and is optimized by the leaving-one-out method.

The two stages form a robust registration algorithm which is applied to the alignment of both, in-vitro and in-vivo radiographs acquired with and without individual adjustment aids, Excellent results are obtained and allow the clinical establishment of the method.

*Corresponding author Dipl.-Ing. Thomas Lehrnann, Institute of Medical Informatics, Aachen University of Technology (RWTH), 52057 Aachen, Germany, Phone: +49 241 80-88793, Fax: +49 241 8888-426, Email: lehmannOcomputer.org

1 Introduction

In medical diagnostics, images of the same region acquired by equal or different modalities over short or long periods of time have to be compared. Considering only one modality, problems arise if images differ in contrast and/or projection, regardless on whether the comparison is done by experts or with automatic computer algorithms [1].

For instance, spatial registration is a major problem in dental radiology when digital radiographs from the same oral region have to be compared to assess changes occuring over a time interval. Digital subtraction techniques have been evidenced to be the best non-invasive tool for diagnosing small changes of internal structures [2, 3] but they require perfect geometrical match. Since mechanical devices such as bite blocks are cumbersome to use clinically and shown to permit reliable replacements not longer than over a period of one year [4] algorithms for automatic registration of dental radiographs have been presented [5, 6, 7] Such algorithm are usually restricted to the detection of movements such

as rotations, scalings and/or translations (RST-movements). Therefore, those techniques do still not enable automatic alignment prior to subtraction of clinical in-vivo data.

Other registration algorithms are based on landmarks which are extracted from both images. The corresponding points are mapped exactly onto each other, while the residuary pixels are interpolated based on the triangulation of the spatial domain [8] or based on energy minimization models. The latter algorithms are also known as thin-plate spline warping [9]. This group of deformation techniques has two major drawbacks when applied to medical images. On the one hand, the transform of every triangle in the image depends on the neighboring fixpoints and so, they usually differ This complicates quantitative measurements after registration and may cause artefacts in the transition of two triangles [8]. On the other hand, the registration is extremely dependent on the actual positioning of corresponding points and therefore, results based on image warping are scarcely reproducible.

In this paper, a registration technique is presented which uses landmarks to determine the parameters of a certain model of geometric projection. To find the landmarks automatical a pre-registration step is required. After pre-adjustment, dominant points in both images could be assigned within a restricted region of interest. In the next section, the projection model for intraoral free-hand radiography is developed. The two-stage algorithm for automatic registration is presented in detail in Section 3. Applications of the algorithm are discussed in the last section of this paper

2 Image acquisition

In intraoral radiography the focus dimensions are usually smaller than $1\,mm^2$ allowing the approximation that the X-ray beam emanates from a point source. Presupposing fixed positions of the X-ray tube and the patient, while the sensor may be rotated and translated in all directions of the three dimensional space, each pixel (x,y) in the image, acquired before the sensor's displacement, is transformed into the position (x', y') in the image obtained afterwards (Fig. 1) This reversible geometric transform is known as perspective projection and is given by [10]

$$x' = \frac{a_1 x + a_2 y + a3}{a_7 x + a_8 y + 1} \quad \text{and} \quad y' = \frac{a_4 x + a_5 y + as}{a_7 x + a_8 y + 1} \quad (1)$$

Nevertheless, all system components such as tube, object, and sensor may be moved and/or rotated in the meantime of two acquisitions of in-vivo radiographs. Concerning such general displacements, the direction of X-rays penetrating the patient may change and, caused by the summation effect

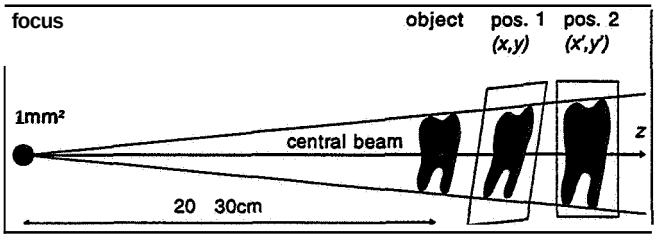


Figure 1 The X-rays emanate from the tube, pass the object, and expose the film or the digital sensor Pos. 1 and pos. 2 describe the location of the sensor acquiring the reference and the subsequent radiograph, respectively. (From [11])

of X-ray imaging, a different image contents will be obtained. In those cases, projection is not reversible.

Table 1 describes 3 rotations (R) and 3 translations (T) of 3 components: X-ray tube (X), patient or object (O), and dental film or digital sensor (S) The possible movements result in 18 displacement vectors. Some of them are neglectable, e.g. rotations of the tube around the projection axis (XR_z) Others are linearly dependent. For example, geometric distortions caused by the movement of the patient perpendicular to the central beam ($+OT_z$) are identical to those obtained by moving tube and sensor in the opposite direction ($-ST_z$ $-XT_z$). Therefore, the set of 18 vectors can be reduced to 8 linear independent vectors, the 6 reversible movements of the sensor and the 2 patient shifts across the projection axis. Note that in intraoral radiology the focus-patient distance usually is larger than 20 cm while the patient-sensor distance is shorter than 1 cm (Fig. 1). Therefore, small patient movements perpendicular to the central beam ($OT_{x,y}$) can be approximated by sensor movements in the same direction ($ST_{x,y}$). In other words, the perspective projection described by equation (1) approximates intraoral freehand radiography.

3 Automatic comparison of in-vivo radiographs

After inverse filtering [12] the digitally acquired X-ray data is adjusted in geometry (Fig. 2). Image regions differing in intensity are segmented automatically after subtraction of the contrast corrected images. Those regions are covered for the next iteration of contrast correction. To register in-vivo radiographs, geometric deformations caused by perspective projection must be eliminated, Since perspective invariants which allow the determination of all eight parameters are still unknown, the geometric adjustment is done in a two-stage algorithm. Data-based pre-registration enables the assignment of dominant points in both images. Those points are linked and used als landmarks to determine the parameters of perspective projection in the fine-adjustment step.

3.1 Pre-adjustment

In this section, the Fourier and the Mellin transform are introduced and utilized to create a rotation-, scaling-, and translation- (RST-)invariant image descriptor Combined with symmetric phase-only matched filter and the cepstrum technique, this descriptor can be used efficiently to decouple and determine the RST-components of the geometric distortion caused by modified imaging geometry.

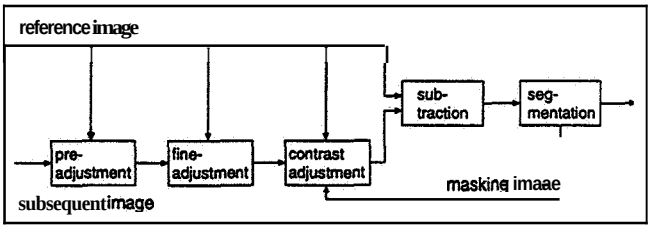


Figure 2: For automatic subtraction of free-hand radiographs the geometric adjustment is done first. Image regions which differ in contents are detected and covered for the subsequent contrast correction.

Displacement	Projection	Linear Combination	Approximation
ST_x	reversible	basic vector	—
ST_y	reversible	basic vector	—
ST_z	reversible	basic vector	—
SR_x	reversible	basic vector	—
SR_y	reversible	basic vector	—
SR_z	reversible	basic vector	—
OT_x	irreversible	basic vector	$-ST_x$
OT_y	irreversible	basic vector	$-ST_y$
OT_z	irreversible	$-ST_z\ -XT_z$	$-ST_z$
O R	irreversible	$-XT_y\ -XT_z\ -XR_x\ +ST_y\ +ST_z\ -SR_x$	$+ST_y\ +ST_z\ -SR_x$
OR_y	irreversible	$+XT_x\ -XT_z\ -XR_y\ -ST_x\ +ST_z\ -SR_y$	$-ST_x\ +ST_z\ -SR_y$
OR_z	reversible	$-SR_z\ -XR_z$	$-SR_z$
XT_x	irreversible	$-ST_x\ -OT_x$	$-ST_x$
XT_y	irreversible	$-ST_y\ -OT_y$	$-ST_y$
XT_z	neglectable	—	0
XR_x	neglectable	—	0
XR_y	neglectable	—	0
XR_z	neglectable	—	0

Table 1 Translations (T) and rotations (R) of the sensor (S), the patient (O), and the X-ray tube (X) result in reversible and irreversible projections. Movements labeled "neclectable" primely affect the image intensity but they only have little influence on the geometric projection of objects. (From: [11])

3.1.1 Fourier transform

The integral transforms [13]

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt \tag{2}$$

and

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}d\omega \tag{3}$$

are well known as Fourier and inverse Fourier transform, respectively One can easily show, that the magnitude of the two-dimensional Fourier transform $|F(u,v)|$ of the image function $f(x,y)$ is invariant to translations in the image plane, while rotations in the (x,y) plane are transformed to equal rotations in the (u,v) plane, and scaling is inverted, e.g. expansions are transformed to shrinkings and vice versa [13].

3.1.2 Mellin transform

The Mellin transform $M(s)$ of a continous one-dimensional function $f(z)$ is given by [14]

$$M(s) = \int_0^{\infty} f(z)z^{s-1}dz \tag{4}$$

and

$$f(z) = \frac{1}{j2\pi}\int_{c-j\infty}^{c+j\infty} M(s)z^{-s}ds \tag{5}$$

denoting the inverse Mellin transform, respectively, where c is the real part of the complex variable $s = c+jw$. The Mellin transform is usually used to compute the moments of $f(z)$ [13]. For example, area, first-, and second-order moments of $f(z)$ are equal to $M(1)$, $M(2)$, and $M(3)$, respectively, and the second-order moment about centroid is given by $M(3) - [M(2)]^2/M(1)$.

Substituting z by e^{-t} we obtain $\ln(z) = -t$ and $dz/dt = -e^{-t}$ Furthermore, the integration limits in (4) transforms like $z \to 0 \Rightarrow t \to \infty$ and $z \to \infty \Rightarrow t \to -\infty$ and the Mellin transform becomes

$$\begin{aligned} M(s) &= \int_{\infty}^{-\infty} f(e^{-t})e^{-t(s-1)}(-e^{-t})dt \\ &= \int_{-\infty}^{\infty} f(e^{-t})e^{-st}dt \\ &= L(s) \end{aligned} \tag{6}$$

with $L(s)$ denoting the two-sided Laplace transform of a function with distorted coordinates $f'(t) = f(e^{-t})$. Further substituting $s = jw$ in (6) yields equation (2). Therefore, the undamped Mellin transform

$$\begin{aligned} M(\omega) &= \int'' f(z)z^{-j\omega-1}dz \\ &= \int_{-\infty}^{\infty} f(e^{-t})e^{-j\omega t}dt \end{aligned} \tag{7}$$

is computable by the Fourier transform, if the coordinate $t \to \ln(1/t)$ is deformed logarithmically before

$$t = e^{-(\ln(\frac{1}{t}))} \equiv e^{-(-\ln(t))} \tag{8}$$

Let f_1 and f_2 be two functions differing only in scale $f_2(z) = f_1(\alpha z)$. Replacing αz by τ, $z = \tau/\alpha$, $dz = d\tau/\alpha$ in the Melliin transform (7) we obtain

$$\begin{aligned} M_2(\omega) &= \int_0^{\infty} f_1(\alpha z)z^{-j\omega-1}dz \\ &= \int_0^{\infty} f_1(\tau)\tau^{-j\omega-1}(\frac{1}{\alpha})^{-j\omega-1}\frac{1}{\alpha}d\tau \\ &= \alpha^{j\omega}M_1(\omega) \end{aligned} \tag{9}$$

One can easily recognize that because of $|\alpha^{j\omega}| = |e^{j\omega\ln\alpha}| = 1$ equation (9) yields $|M_1(\omega)| = |M_2(\omega)|$. Therefore, the magnitude of a Mellin transformed function is invariant to scales.

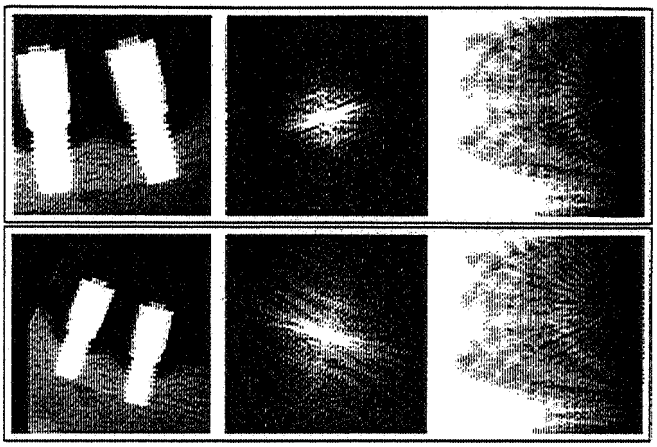


Figure 3: The radiographs $\mathbf{f}(x, y)$ on the left show dental implants. The Fourier power-spectra $|F(u, v)|$ are placed in the middle. Note their invariance to translations, variance to rotation, and inversion of scales. The results of logarithmic polar mapping $|F(\rho, \varphi)|$ are displayed on the right. Rotations and translations are transformed into shifts along the vertical and horizontal axis, respectively. (From [1])

3.1.3 RST-invariant image descriptors

An overview of invariant image descriptors is given elsewhere [10]. In the following, the Fourier and the Mellii transform are combined to a RST-invariant image descriptor similar to those introduced by *Schalkoff* [15] or *Haken* [16]. Therefore, let f_1 and f_2 be two given functions with f_2 being a (x_0, y_0) shifted, a rotated, and β scaled version of f_1

$$f_2(x, y) = \tag{10}$$
$$f_1\big(\beta(x\cos a + y\sin\alpha) - x_0, \beta(-x\sin a + y\cos a) - y_0\big)$$

The Fourier power-spectra are computed to

$$|F_2(u, v)|^2 = \tag{11}$$
$$\left|\frac{1}{\beta^2} F_1\left(\frac{u\cos\alpha + v\sin\alpha}{\beta}, \frac{-u\sin\alpha + v\cos a}{p}\right)\right|^2$$

This representation of the functions f_i only depends on the angle of rotation α and the scaling β. Both may be decoupled applying a transformation to polar coordinates (Fig. 3). Taking into account the addition theorems of trigonometric functions, the substitution $u = r\cos\varphi$ and $v = r\sin\varphi$ in (11) yields

$$|F_2(r\cos\varphi, r\sin\varphi)|^2 = \tag{12}$$
$$\left|\frac{1}{\beta^2} F_1\left(\frac{r\cos(\varphi - a)}{\beta}, \frac{r\sin(\varphi - a)}{p}\right)\right|^2$$

Denoting the power-spectra in (12) as functions of (r, φ) leads to

$$|F_2(r, \varphi)|^2 = \left|\frac{1}{\beta^2} F_1\left(\frac{r}{\beta}, \varphi - \alpha\right)\right|^2 \tag{13}$$

Scalings are transformed into stretches of the r-axis, while rotations are mapped to shifts along the φ-axis. Applying the scale-invariant Mellin transform (7) to the r-coordinate and the shift-invariant Fourier transform (2) to the φ-coordinate

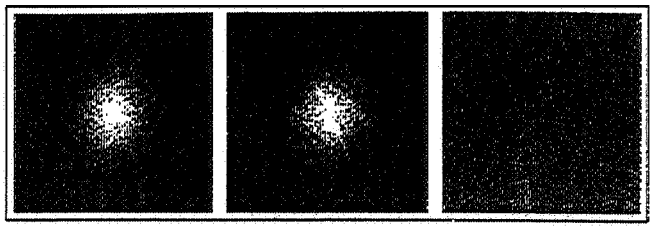


Figure 4 The log-polar representation of the power-spectra from Figure 3 have been Fourier transformed once more. For visualization, the corresponding power-spectra are histogram-optimized. The pixelwise subtraction (right-hand side) demonstrate that there is no structural difference. (From [1])

(Fourier-Mellin transform) equation (13) results in a RST-invariant image representation of both functions f_1 and f_2.

As shown above, the Mellin transform can be calculated using the Fourier transform if the coordinate r is mapped logarithmical $\rho = -\ln(r)$. Combining (8) and (13) yields

$$|F_2(\rho, \varphi)|^2 = \left|\frac{1}{\beta^2} F_1\Big(\rho + \ln(\beta), \varphi - \alpha\Big)\right|^2 \tag{14}$$

In this representation of the entire images f_1 and f_2 the RST-movements are decoupled (Fig. 3). Translations are eliminated while magnifications are represented as shifts in the ρ-axis and rotations are expressed as shifts in the φ-axis. A further calculation of the Fourier power-spectra of the functions given in (14) results in a RST-invariant image descriptor (Fig. 4)

$$|M_2(u, v)|^2 \sim |M_1(u, v)|^2 \tag{15}$$

The mapping of multiplications (coordinate-scaling) to additions (coordinate-shifting) is a well known property of the logarithm. Therefore, related methods are often used without referring to as Fourier-Mellin transform [17].

3.1.4 RST registration

Using the Fourier-Mellin RST-invariant, only translational movements have to be registered for pre-adjustment of radiographs. Usually, such kind of registration is done applying correlation methods which are also referred to as matched filter or template matching. Nevertheless, there are two major drawbacks bind to those methods. Correlation is computational expensive and the maximum of the correlation function is plane which complicates its localization. Applying Fourier techniques such as cepstrum or symmetric phase-only matched filter instead of spatial correlation, registration could be done by detecting sharp peaks.

Matched filter (MF) The aim of image matching is either to determine the presence of a known image in a noisy scene or to determine the parameters of a geometric translation relating the two images $s(x, y) = r(x - x_0, y - y_0)$ and $r(x, y)$ In the presence of white noise with the intensity N, the optimal receiver with respect to the signal-to-noise ratio has the transfer function

$$H_{MF}(u, v) = \frac{1}{|N|^2} R^*(u, v) \tag{16}$$

and the output of the filter $Q_{MF}(u, v)$ equals the convolution

$$q_{MF}(x, y) = \frac{1}{|N|^2} \iint s(a, b)\, r^*(a - x, b - y)\, da\, db$$

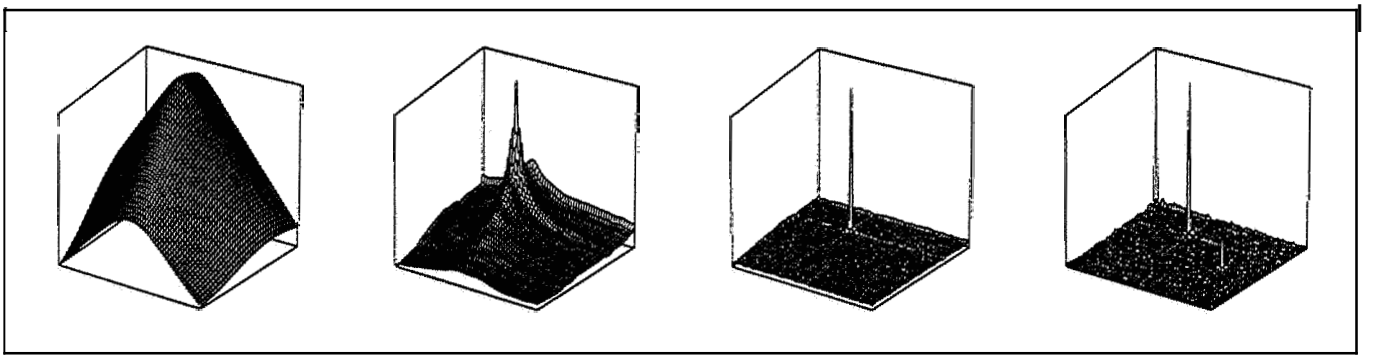


Figure 5: The radiographs from Figure 6 were adjusted. The results of filtering $q(x, y)$ are visualized in the spatial domain. All functions are maximal at the position (24, 16) determing the shift of the radiographs from Figure 6. From left to right: a. matched filter (MF), b phase-only matched filter (POMF), c: symmetric phase-only matched filter (SPOMF), d cepstrum filter (CF) For CF, the images $s(x, y)$ and $r(x, y)$ were pixelwise added to form the cepstrum window $g(x, y)$ which is equivalent to $D = 0$ in equations (21) to (24).

which has a maximum at (x_0, y_0) determing the parameters of the translation [18]. Nevertheless, the MF depends on the image's energy rather than on its spatial structures. This is why the MF provides poor discrimination between objects of different shape but similar size or energy, In addition, the filter output depends on the auto-correlation and its shape is broad around its maximum (Fig. 5)

Phase-only matched filter (POMF) In the Fourier domain, the phase of an image comprise the information about the localization of structures. Therefore, the transfer function of the MF is normalized by its magnitude which defines the POMF [19]

$$H_{\mathrm{POMF}}(u, v) = \mathrm{Phase}\{R^*(u, v)\} = e^{j(-\Phi_r(u,v))} \qquad (17)$$

Using POMF, the detectability of the maximum peak of the filter response $q(x, y)$ is improved (Fig. 5).

Symmetric phase-only matched filter (SPOMF) Following this idea straightforward, not only the MF transfer function $R^*(u, v)$ is normalized, but also the Fourier transform S of the incoming signal. This yields the nonlinear SPOMF [18]

$$H_{\mathrm{SPOMF}}(u, v) = \frac{R^*(u, v)}{|S(u, v)| \; |R^*(u, v)|} \qquad (18)$$

With (18), the filtered signal $Q_{\mathrm{SPOMF}}(u, v)$ contains only phase information but a constant magnitude (Fig. 6)

$$Q_{\mathrm{SPOMF}}(u, v) = e^{j(\Phi_s(u,v) - \Phi_r(u,v))} = e^{-j2\pi(ux_0 + vy_0)} \qquad (19)$$

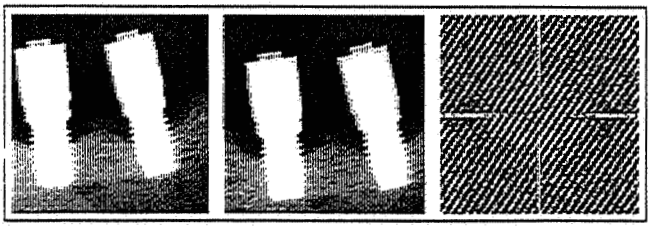


Figure 6: The radiographs were spatially shifted by $x_0 = 24$ and $y_0 = 16$ pixels. The result of phaseonly matched filtering $Q_{\mathrm{SPOMF}}(u, v)$ is shown on the right.

Referring to the shift theorem of the Fourier transform [13], the filter response $q_{\mathrm{SPOMF}}(x, y)$ equals a delta-function which is located at the position (x_0, y_0). Comparing MF, POMF, and SPOMF, this peak determines the shift between the images $r(x, y)$ and $s(x, y)$ best (Fig. 5)

Cepstrum filter (CF) The cepstrum analysis was developed as an one-dimensional technique to examine seismographic data containing echos of some arbitrary wave packets [20, 21]. Based on the separability of the discrete Fourier transform $\mathcal{F}\{\}$ cepstral filtering can easily be transfered into two dimensions. In general, the power-cepstrum $\mathcal{C}\{\}$ of a function f is the power-spectrum of the logarithm of the function's power-spectrum

$$\mathcal{C}\{f(x, y)\} = \left|\mathcal{F}\left\{\log\left|\mathcal{F}\{f(x, y)\}\right|^2\right\}\right|^2 \qquad (20)$$

Commonly, the cepstral filtering is used to estimate a scene's translational shift by placing the two images to be compared successively in the center of a so called cepstrum window Assume the cepstrum window to consist of the reference image $r(x, y)$ on the left-hand side and the subsequent image $s(x, y) = a_0 r(x - x_0, y - y_0)$ on the right-hand side. Then, the resulting cepstrum window $g(x, y)$ may be expressed like

$$\begin{aligned} g(x, y) &= r(x, y) + s(x - D, y) \qquad (21) \\ &= r(x, y) * \left[\delta(x, y) + a_0 \delta(x - (D + x_0), y - y_0)\right] \end{aligned}$$

where a_0 is an amplitude scale factor and D describes the dimension of the square images r and s. Again, the relative shift in x- and y-direction between both images is written in terms of x_0 and y_0, respectively.

Stepwise computing the power-cepstrum (20) we obtain

$$G(u, v) = R(u, v)\left(1 + a_0 e^{-j2\pi(u(D + x_0) + vy_0)}\right) \qquad (22)$$

where $R(u, v)$ and $G(u, v)$ determine the Fourier transform in the frequency domain (u, v) of the reference image and the cepstrum window, respectively. Equation (23) shows the logarithm of the power-spectrum of the composite signal g containiig cosinusoidal ripples with amplitude and frequency

$$\log|G(u,v)|^2 = \log|R(u,v)|^2 + \log\left(1 + 2a_0\cos\left(2\pi(u(D+x_0)+vy_0)\right) + a_0^2\right)$$
$$= \log|R(u,v)|^2 + \log|1+a_0^2| + \frac{2a_0}{1+a_0^2}\cos\left(2\pi(u(D+x_0)+vy_0)\right) \mp \quad (23)$$

related to the intensity factor a_0 and the translational displacements x_0 and y_0. With

$$\log(1+x) = \sum_{n=1}^{\infty}(-1)^{n+1}x^n/n \quad \text{and} \quad -1 < x \leq 1$$

a second power-spectrum operation yields the power-cepstrum of $g(x,y)$

$$\mathcal{C}\{g(x,y)\} = \mathcal{C}\{r(x,y)\} + A\delta(x,y) + B\delta(x \pm (x_0+D), y \pm y_0) + C\delta(x \pm 2(x_0+D), y \pm 2y_0) + \quad (24)$$

which equals the power-cepstrum of the reference image plus a train of delta functions occuring at integer multiplies of the translational shifts $(D+x_0, y_0)$. Thus, the translational difference between two images can simply be detected by inspecting the distance between the origin and the location of the first maximum peak in the cepstral. plane (Fig. 5). Furthermore, the resulting train of delta-functions can be utilized for advanced registration algorithms.

Although this classical cepstrum technique requires no segmentation and is reliable, efficient, and robust to noise, some pre-processing steps are recommended [22, 23]. Since the discrete Fourier transform assumes the images to be periodical, windowing is required to avoid artefacts resulting from the non-periodical property of finite images. In our application, the Kaiser-Bessel window is used [24]. Additional preprocessing by sobel-filtering is also applied to emphasize edge-information. In addition, the Fourier power-spectra are histogram-optimized by stretching their logarithmic scale before they are mapped into polar coordinates [23].

3.1.5 Final algorithm for pre-adjustment

Figure 7 illustrates the resulting algorithm for pre-adjustment (FFT denotes the fast Fourier transform). The Fourier-Mellin invariant is used to decouple the RST-components of movement. The adjustment of rotations and scales is done in the Fourier domain. Hence, important information is coded in the higher frequencies which are located in the outer part of the spectra. Those image areas are affected most by windowing. Therefore, the registration of the log-polar power-spectra is done by SPOMF. CF is applied for the subsequent registration in the spatial domain to determine translations.

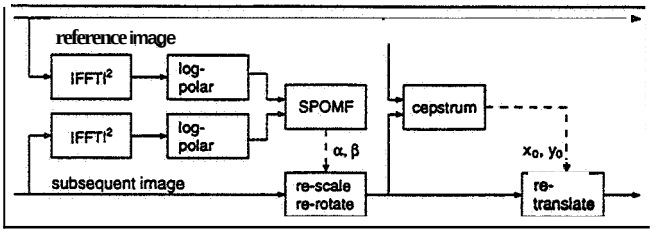


Figure 7 Pre-registration

3.2 Fine-adjustment

Geometric fine-adjustment of radiographs must correct the remaining perspective projection between the reference radiograph and the pre-registered subsequent image. This stage is based on landmarks which are defined next. Local correlation with sub-pixel resolution as well as the leaving-one-out method are applied to optimize the fine-registration step.

3.2.1 Landmarks

A landmark, also referred to as corresponding or reference point, is defined by a dominant pixel in a grey scale image which is highlighted by strong local contrasts. Figure 8 examplifies landmarks in dental radiographs. Small white crosses are overlaid to indicate the landmarks' positions. The edges of teeth, fillings, and implants as well as the spongious structure of bone enable the determination of recognizable landmarks.

The coordinates (x,y) and (x',y') of corresponding landmarks can be used to determine the parameters a_i of the perspective projection mapping the current radiograph onto the reference. Using 4 pairs of landmarks, inversion of equation (1) yields the system of equations (25) which can be solved if the landmarks are linearly independent,

Figure 9 shows the reference radiograph with two sets of four landmarks defined in each, and the resulting subtraction after registration. The quality of adjustment is extremely affected by the exactness of positioning the landmarks. Therefore, the registration depends on the observer or automatic algorithm placing the landmarks in both images.

3.2.2 Least-squares problem

In our registration approach, more than four pairs of points are used. Let $N \in \mathbb{N}$ denote the number of corresponding points. Then, equation (1) yields the overdetermined system of equations (26), which is of the form $(Az = b)$

In general, it is impossible to find a unique set of parameters $z = (a_1, \ldots, a_8)^T$ that will satisfy (1) for all $n = 1, \ldots, N$. Thus, a set of such parameters $\hat{z}$ must be found which, on the average, is optimal. This optimal approximation can be determined using the least-squares method [25].

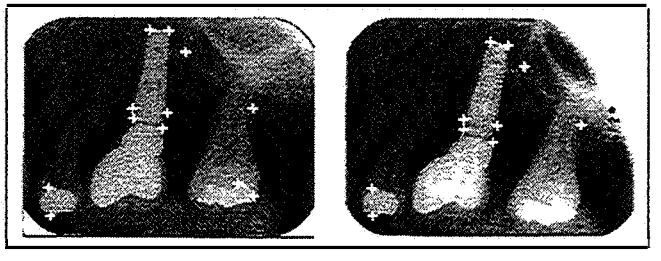


Figure 8 Some landmarks in the reference and subsequent radiograph have been marked by a human observer. They are visualiied by small white crosses.

$$\begin{pmatrix} 21 & y_1 & 1 & 0 & 0 & 0 & -x'_1x_1 & -2.\ y_1 \\ 0 & 0 & 0 & 21 & y_1 & 1 & -y'_1x_1 & -y'_1y_1 \\ x_2 & y2 & 1 & 0 & 0 & 0 & -x'_2x_2 & -x'_2y_2 \\ 0 & 0 & 0 & 22 & y_2 & 1 & -y'_2x_2 & -y'_2y_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x'_3x_3 & -x'_3y_3 \\ 0 & 0 & 0 & 23 & y3 & 1 & -y'_3x_3 & -y'_3y_3 \\ x_4 & y4 & 1 & 0 & 0 & 0 & -x'_4x_4 & -x'_4y_4 \\ 0 & 0 & 0 & 24 & y4 & 1 & -y'_4x_4 & -y'_4y_4 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{pmatrix} = \begin{pmatrix} x'_1 \\ y'_1 \\ x'_2 \\ y'_2 \\ x'_3 \\ y'_3 \\ x'_4 \\ y'_4 \end{pmatrix} \tag{25}$$

$$\begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x'_1x_1 & -x'_1y_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y'_1x_1 & -y'_1y_1 \\ & & & & & & & \\ x_N & y_N & 1 & 0 & 0 & 0 & -x'_Nx_N & -x'_Ny_N \\ 0 & 0 & 0 & x_N & y_N & 1 & -y'_Nx_N & -y'_Ny_N \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{pmatrix} = \begin{pmatrix} x'_1 \\ y'_1 \\ \\ x'_N \\ y'_N \end{pmatrix} \tag{26}$$

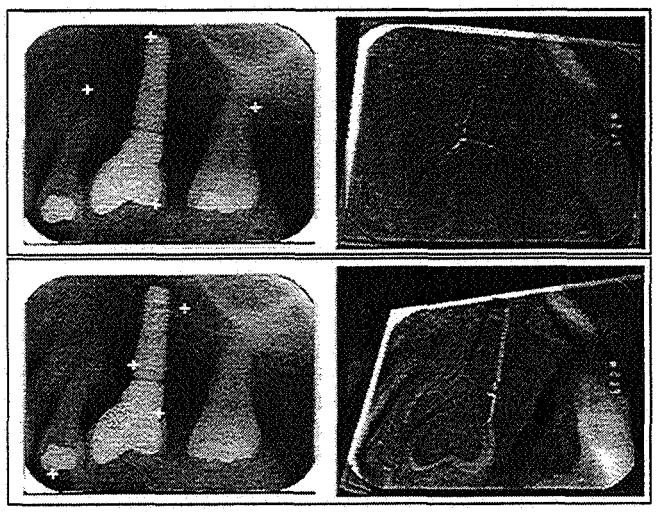


Figure 9: The results of registration based on four landmarks is examplified. The subtraction images on the right demonstrate the strong dependence on the positioning of landmarks.

In equation (26) the data matrix A $\in \mathbb{R}^{2N\times 8}$ and the observation vector b $\in \mathbb{R}^{2N}$ are given. The sought vector $\hat{z}$ is of the dimension $\hat{z} \in \mathbb{R}^8$ This suggests that one strives to minimize $\|A\hat{z} - b\|_p$ for some suitable choice of p. For the 2-norm p = 2 the function

$$\phi(\hat{z}) = \frac{1}{2}\|A\hat{z} - b\|_2^2 \tag{27}$$

is differentiable in $\hat{z}$ and so the minimizers of ϕ satisfy the gradient equation $\nabla\phi(\hat{z}) = A^T(A\hat{z} - b) = 0$. This turns out to be an easily constructed symmetric linear system which is positive definite if A has full column rank. The gradient equation $\nabla\phi(\hat{z}) = 0$ is tantamount to solving the normal equations

$$A^TA\hat{z} = A^Tb \tag{28}$$

Since the matrix $A' = A^TA \in \mathbb{R}^{8\times 8}$ is symmetric positive definite, there exists a unique lower triangular matrix $C \in \mathbb{R}^{3\times 3}$ with positive diagonal entries such that $A' = CC^T$ This factorization is known as the Cholesky factorization and C is referred to as the Cholesky triangle [25].

Using the Cholesky factorization of (28) yields

$$A^TA\hat{z} = A'\hat{z} = (CC^T)\hat{z} = C(C^T\hat{z}) = Cz' = A^Tb \tag{29}$$

and therefore, the least-squares problem is reduced to the solution of the triangular systems: $Cz' = A^Tb$ and $C^T\hat{z} = z'$ Since A' determines perspective projection, the structure of A'

$$A' = A^TA = \begin{pmatrix} A'_{1.1} & 0 & A'^T_{3.1} \\ 0 & A'_{1.1} & A'^T_{3.2} \\ A'_{3.1} & A'_{3.2} & A'_{3.3} \end{pmatrix} \tag{30}$$

can be utilized for an efficient blockwise implementation of the Cholesky factorization. Note, that the blocks $A'_{i.j}$ in the diagonal with $i = j$ are squared and therefore

$$A'_{i.j} = \sum_{k=1}^{j} C_{i.k}C^T_{j.k} \quad \text{with} \quad i,j \in \{1,2,3\} \quad \text{and} \quad i \geq j \tag{31}$$

Let $\sum$ abbreviate $\sum_{n=1}^{N}$, then equations (26), (30) and (31) allow an easy and numerically most efficient calculation of the least-squares approximation $\hat{z} = (a_1, \ldots, a_8)^T$ determining the eight parameters of perspective projection

$$\begin{aligned} A'_{1.1} &= \begin{pmatrix} \sum x_n^2 & \sum x_ny_n & \sum x_n \\ \sum x_ny_i & \sum y_n^2 & \sum y_n \\ \sum x_n & \sum y_n & N \end{pmatrix} \\ &= C_{1.1}C^T_{1.1} \\ A'_{3.1} &= \begin{pmatrix} -\sum x'_nx_n^2 & -\sum x'_nx_ny_n & -\sum x'_nx_n \\ -\sum x'_nx_ny_n & -\sum x'_ny_n^2 & -\sum x'_ny_n \end{pmatrix} \\ &= C_{3.1}C^T_{3.1} \\ A'_{3.2} &= \begin{pmatrix} -\sum y'_nx_n^2 & -\sum y'_nx_ny_n & -\sum y'_nx_n \\ -\sum y'_nx_ny_n & -\sum y'_ny_n^2 & -\sum y'_ny_n \end{pmatrix} \\ &= C_{3.1}C^T_{2.1} + C_{3.2}C^T_{2.2} = C_{3.2}C^T_{1.1} \\ A'_{3.3} &= \begin{pmatrix} \sum(x'^2_n + y'^2_n)x_n^2 & \sum(x'^2_n + y'^2_n)x_ny_n \\ \sum(x'^2_n + y'^2_n)x_ny_n & \sum(x'^2_n + y'^2_n)y_n^2 \end{pmatrix} \\ &= C_{3.1}C^T_{3.1} + C_{3.2}C^T_{3.2} + C_{3.3}C^T_{3.3} \end{aligned}$$

Figure 10 shows some registration examples based on more than four points. The data of coordinates was fitted to the model of perspective projection as described in the previous paragraphs. Although the landmarks were positioned at

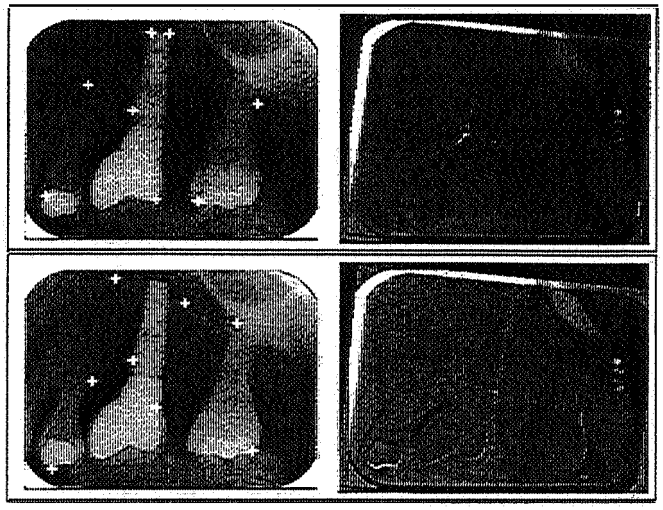


Figure 10: Additional points were added in the radiographs from Figure 9. The subtraction images on the right verify the result of registration being improved and independent on the distribution of landmarks.

completely different positions, the results of registration are nearly the same, and, compared with the registration based on only four landmarks (Fig. 9), the exactness of registration is significantly improved.

3.2.3 Automatic detection of landmarks

The manual placement of corresponding points in a discrete pixel array has two disadvantages. The user is limited to the discrete positions of a digital image although structures in a subsequent radiograph may appear between these discrete coordinates. This is the major drawback. The use of conventional computer mice may cause additional incorrectness. Therefore, the exactness of registration is improved by local correlation. The idea of this refinement is to cut a small template around a given landmark in the subsequent radiograph and use it for template matching in a small window arround the corresponding landmark in the reference X-ray image. In our approach, local correlation is computed with sub-pixel resolution.

This principle is extendet for automatic determination of landmarks. Simple gradient filters [27, 28] are applied to determine candidate points for landmarks in the subsequent image. Since both radiographs are already pre-adjusted, local correlation could be restricted to a certain neighborhood in the reference image. The correlation coefficient is used as threshold to remove point pairs which differ to much Advanced algorithms for the detection of landmarks in medical images are described elsewhere [29].

Figure 11 shows the registration result based on 27 automatical extracted landmarks. The landmarks were determined by the algorithm from *Likar* et *al.* [29] The result of fine-adjustment equals those obtained with manual placed landmarks (Fig. 10).

3.2.4 Leaving-one-out

Once the parameters a_1 to a_8 have been determined the perspective projection described by equation (1) could be solved for each landmark-point (x, y) in the reference image. The distances d_n between the transformed coordinates and the

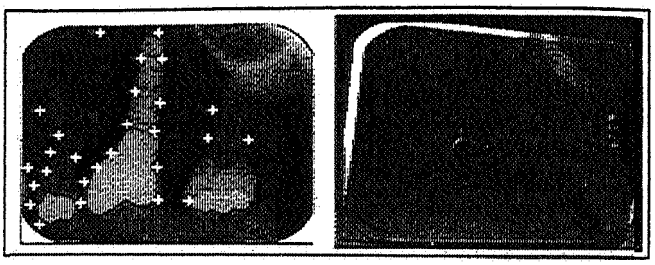


Figure 11. The 27 landmarks used for fine-adjustment were extracted automatical. The subtraction image on the right demonstrates equal quality compared with manual placed landmarks (Fig. 10).

manually marked corresponding points (x', y') indicate the exactness of the estimated geometric projection

$$d = \frac{1}{N} \sum_{n=1}^{N} d_n \qquad (32)$$

with $d_n =$

$$\sqrt{\left(x'_n - \frac{a_1 x_n + a_2 y_n + a_3}{+a_8 y_n + 1}\right)^2 + \left(y'_n - \frac{+a_5 y_n + a_6}{a_7 x_n + a_8 y_n + 1}\right)^2}$$

A numerical more efficient way to determine the quality of the approximation $\hat{z}$ is given by the minimal residual [25]

$$r = b - A\hat{z} \qquad \text{whereas} \qquad \rho = \frac{1}{2N} \|b - A\hat{z}\|_2 \qquad (33)$$

denotes its relative size. In other words, ρ determines the overall quality of the landmarks fitting the model of perspective projection. In our approach, this measure is used to assess the influence of single pairs of corresponding points to the final adjustment. An optimization of ρ leads to an improved registration.

The leaving-one-out method is most widely used to determine the influence of one parameter in a set training a neural network for classification [26]. This procedure is particularly attractive when the number of samples is quite small. Here, this technique is applied to improve the quality of registration. Suppose N points marked in both radiographs. At first, the N subsets containing $N - 1$ landmarks are generated. For the initial set and all subsets the $L = N + 1$ parameter vectors $\hat{z}_l$ and their residuals ρ_l with $l = 1, \ldots, L$ are determined solving equation (28) Finding one ℓ such that $\rho_\ell \leq \rho_l$ for all $l \neq \ell$ defines the best set of parameters. This set is used initially for the next leaving-one-out iteration. When the minimum number $N_{\min}$ of points in a set is reached or the best set equals the initial set, the iteration stops and $\hat{z}_\ell$ is used for final back-projection of the subsequent image into the geometry of the reference image.

3.2.5 Final algorithm for fine-adjustment

The algorithm for he-registration of radiographs is summarized in Figure 12. After the determination of landmarks and the optimization of their coordinates by sub-pixel correlation, the parameters of perspective projection were estimated. The least-squares method is efficiently implemented by a blockwise Cholesky factorization The minimum residual is used to control the leaving-one-out iteration

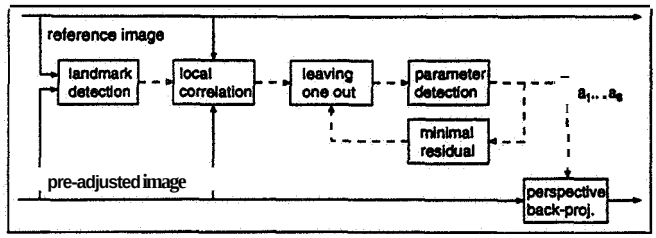


Figure 12: Fine-adjustment

4 Results and Discussion

Automatic subtraction of intraoral radiographs is an important means in oral diagnostics [2, 3]. Although digital imaging devices such as CCD-sensors or storage phosphor plates have become standard in modern dental clinics, subtraction radiography is still restricted to scientific studies and not jet established in clinical routine. This is caused by the enormous problems on standardizing imaging geometry during the large time intervals between the X-ray examinations, which can be up to one year [4].

Computer algorithms for automatic image registration are mostly restricted to affine transforms like rotations, scales, and/or translations [7, 23]. However, geometric transforms occuring in intraoral free-hand radiography can be approximated by perspective projection [11]. Perspective invariants that allow the determination of all model parameters are still unknown [10]. Therefore, correction algorithms for perspective geometry are usually based on landmarks which are placed manually. Four corresponding reference points in both images are required to determine the eight parameters of perspective projection [30] Nevertheless, the registration results obtained with four landmarks extremely depend on the correct localization of those points [11]. In this paper, an automatic and reproducable method for registration of perspective projection was presented.

In the first stage, RST-components of the projection are corrected using the Fourier-Mellin invariant. This data-based method was evaluated and optimized in previous studies [23, 7]. It was shown, that the combination of the Fourier-Mellin invariant with cepstrum techniques allows reliable (correctness > 80%) registration of pictures overlapping more than 70% with rotational components less than 15 degrees, regardless of noise or intensity variations occuring in clinical situations (Tab. 2). As pointed out in paragraph 3.1.4, the CF requires windowing and therefore, it is less useful to register Fourier power-spectra. Applying SPOMF for the rotation and scale detection step, the results are improved. Again, the X-ray sequence was used to quantify the improvement. The algorithm summarized in Figure 7 yields reliable (correctness ~ 90%) registrations up to 35 degree of rotations (Tab 2) In clinical freehand radiography, rotations will be significantly less. Note that the rotations between contiguous frames in Figre 13 are about 35 degrees.

The subsequent registration step is based on corresponding points. In a previous publication was shown, that only 6 fairly distributed landmarks are required to guarantee (correctness > 95%) independence on the positions of the landmarks [11]. Note that automatic algorithms usually extract much more landmarks [29]. Therefore, our two-step algorithm for automatic registration of radiographs acquired intraorally could be established in clinical routine.

Rotation angle	Image pairs	CF No	CF %	SPOMF No	SPOMF %
0.0°	33	33	100	33	100
5.6°	32	32	100	32	100
11.3°	31	31	100	31	100
16.9°	30	29	97	30	100
22.5°	29	23	79	28	97
28.1°	28	19	68	26	93
33.8°	27	12	44	25	93
39.4°	26	11	42	20	77
45.0°	25	7	28	18	72

Table 2: Combining 33 single frames of the rotation sequence (Fig. 13) in pairs of two, altogether 261 different sets of radiographs with a-priori known relative rotation were registered. The results obtained with CF performing the rotation and scale detection step are taken from [7].

Once the images are registered, contrast correction could be easily done by matching the image's histogram functions [1]. Further work will be concentrate on automatic segmentation and quantitative measurement in the subtraction radiographs. Since the algorithm uses no a-priori knowledge on image contents, the method could be transferred into other fields *in* medical imaging, for instance thorax examinations. Adapting the model of projection opens a various field of applications.

Acknowledgement

This work was partly supported by the German Research Community (DFG), grants: Re 427/5-1, Sp 538/1-2, Schm 1268/1-2 The additional financial support by the German Research Community (DFG), grant: Le 1108/1-1, is gratefully acknowledged.

References

[1] Lehmann T, Oberschelp W, Peliian E, Repges R. *Bildverarbeitung für die Medizin: Grundlagen, Modelle, Methoden, Anwendungen.* Springer-Verlag, Berlin, 1997 (in German)

[2] Grondahl HG, Grondahl K. Subtraction radiography for the diagnosis of periodontal bone lesions. *Oral Surg.*, 55:208–213, 1983.

[3] Wenzel A, Warrer K, Karring T Digital subtraction radiography in assessing bone changes in periodontal defects following guided tissue regeneration. *J Clin. Periodontol.*, 19:208–213, 1992.

[4] Janssen PTM, van Palenstein Helderman WH, van Aken J: The effect of in-vivo-occurring errors in the reproducibility of radiographs on the use of the subtraction technique. J *Clin. Periodontol.*, 16:53–58, 1989.

[5] Van der Stelt PF, Ruttimann UE, Webber RL: Determination of projections for subtraction radiography based on image similarity measurements. *Dentomazillofacial Radiology*, 18:113–117, 1989.

[6] Dunn SM, van der Stelt PF, Fenesy K, Shah S: A comparison of two registration techniques for digital subtraction radiography, *Dentomaxillofacial Radiology*, 22:77–80, 1993.

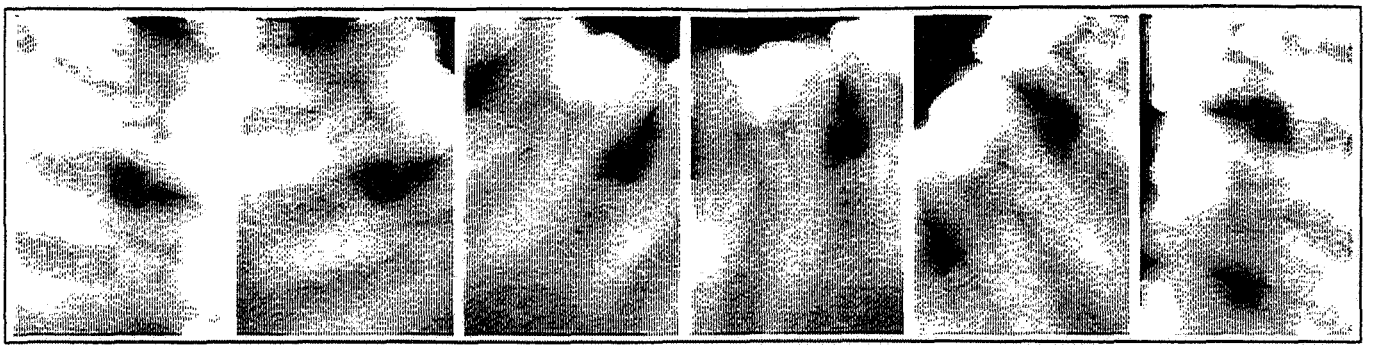


Figure 13: Standardized in-vivo radiographs were acquired with a special device. During a 180 degree rotation, the dry mandible was imaged 33 times. This figure shows the 1st, 7th, 14th, 20th, 27th, and 33th frame of the sequence. (From [7])

[7] Lehmann T, Goerke C, Schmitt W, Kaupp A, Repges R. A rotation-extended cepstrum technique optimized by systematic analysis of various sets of X-ray images. *Proceedings SPIE,* 2710390-401, 1996.

[8] Ruprecht D, Müller H: *Image warping with scattered data znterpolation methods.* Technical Report 443, FB Informatik LS VII, University of Dortmund, Germany, 1992.

[9] Bookstein FL: *Morphometric tools for landmark data.* University Press, Cambridge, 1991

[10] Reiss T H *Recognizing planar* objects *using invariant image features.* Lecture Notes in Computer Science 676, Springer, 1993.

[11] Lehmann T, Grondahl K, Grondahl HG, Schmitt W, Spitzer K. Observer-independent registration of perspective projection prior to subtraction of in-vivo radiographs. *Dentomazillofacial Radiology,* submitted

[12] Peters B, Meyer-Ebrecht D, Lehmann T, Schmitt W System Analysis of X-ray-sensitive CCD's and Adaptive Restoration of Intraoral Radiographs. *Proceedings SPIE* 2710(45):450–466, 1996.

[13] Bracewell RN, *The Fourier Transform and Its Applications.* 2nd ed., McGraw-Hill, New York, 1986.

[14] Oberhettinger F: *Tables of Mellin Transforms.* Springer-Verlag, Berlin, 1974.

[15] Schalkoff RJ: *Digital Image Processing and Computer Vision.* John Wiley & Sons, New York, 1989.

[16] Haken H. *Synergetzc Computers and Cognition. A Top-Down Approach to Neural Nets.* Springer-Verlag, Berlin, 1991.

[17] Cideciyan AV Registration of Ocular Fundus Images. *IEEE Magazine* EMB-14(1):52–58, 1995.

[18] Chen QS, Defrise M, Deconinck F: Symmetric Phase-Only Matched Filtering of Fourier-Mellin Transforms for Image Registration and Recognition. *IEEE Bansactions* PAMI-16(12):1156–1168, 1994.

[19] Horner JL, Gianino PD: Phase-only matched filtering. *Applied Optics* 23(6):812–816, 1984

[20] Bogert BP, Healy MJR, Tukey JW: The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo-Autovariance, Cross-Cepstrum and Saphe Cracking. *Proceedings of the Symposium* of *Time Series Analysis,* pp. 209–242, 1963.

[21] Childers DG, Skinner DP, Kemerait RC: The Cepstrum: A Guide to Processing. *Proceedings of the IEEE* 65:1428–1443, 1977

[22] Lee DJ, Mitra S, Krile TF: Analysis of sequential complex images, using feature extraction and two–dimensional cepstrum techniques. *Journal of the Optical Society of America A* JOSA A-6:863–870, 1989.

[23] Lehmann T, Goerke C, Kaupp A, Schmitt W, Repges R. A Rotation-Extended Cepstrum Technique and its Application to Medical Images. *Pattern Recognition and Image Analysis* 6(3):592–604, 1996.

[24] Harris FJ: On the Use of Windows for Harmonic Analysis with the Discrete Fourier–Transform. *Proceedings of the IEEE* 66:51–83, 1978.

[25] Golub GH, van Loan CF: *Matrix Computations.* Johns Hopkins, Baltimore, 1989.

[26] Duda RO, Hart P E *Pattern Classification and Scene Analysas.* John Wiley & Sons, New York, 1973.

[27] Förstner W A feature based correspondence algorithm for imaging matching. *Int. Arch. Photogramm. Remote Sensing* 26:150–166, 1986.

[28] Rohr K. Modelling and identification of characteristic intensity variations. *Image and Vision Computing* 10:66–76, 1992.

[29] Likar B, Bernard R, Pemus F: *Automatzc extraction of point pairs for matching 2D medical images.* World Congress on Medical Physics and Biomedical Engineering, Nice, France, September 1997

[30] Ostuni J, Fisher E, van der Stelt P, Dunn S: Registration of dental radiographs using projective geometry *Dentomazillofacial Radiology* 22:199-203, 1993.

Regional Registration of Texture Images with Application to Mammogram Followup

D Brzakovic and N Vujovic
Department of Electrical Engineering and Computer Science
Lehigh University
Bethlehem, PA 18015

Abstract

The work is motivated by the problem of mammogram screening based on comparison between mammograms of the same patient acquired in different screenings. Misregistration between temporally spaced screenings arises from minor differences in 3-D positioning and compression, as well as, normal changes in tissue that are functions of time Since the relationship between images cannot be modeled with available information, precise registration and pixel-to-pixel comparison as an intractable problem Instead, this work proposes defining corresponding regions in two images and carrying out comparison between them, similarly to what is done by medical experts. The locations of the regions are determined based on the locations of identifiable landmark points, and regions' extents are determined by characteristics of the older mammogram.

1 Introduction

There is strong evidence that early breast cancer detection is the key to successful treatment Mammography is the primary means of early detection and regular screenings are recommended for a significant percentage of the female population Maintaining high reliability in massive screenings calls for introducing computers into the present protocols. The primary area where computer technology is useful is in ensuring adequate scanning thus allowing the experts to concentrate on interpretation Moreover, computer-based screening can provide increased sensitivity to detecting abnormalities that are masked by underlying normal anatomical structures. The additional benefits of using computers include reducing the cost, providing efficient storage media, and telemedicine capabilities. In spite of these important advantages and considerable research effort in the past decade, techniques for aiding mammogram screening have not yet yielded clinically acceptable rates of false negatives and/or false positives. The most fundamental problem with computer algorithms is that they require description of what is considered normal and/or abnormal in a mammogram. The tremendous variability of both has made formulations of such descriptions very difficult and so far developed algorithms lack the adaptability necessary to handle variations normally encountered in mammograms

The work described in this paper aims at overcoming this lack of adaptability by searching for abnormalities by comparing temporally spaced screenings of a patient An advantage of considering mammogram sequences relative to analyzing a single mammogram is that it uses the older mammogram as a reference, and thus, is likely to reduce the high rate of false positives associated with many computer-based methods. More importantly, following subtle changes may provide for very early cancer detection The constituent parts necessary in realizing mammogram comparison are: (i) algorithms for putting two temporally spaced screenings of the same breast into correspondence, (ii) algorithms for characterizing mammogram regions, and (iii) decision making paradigm. The concentration of this paper is on the first part, i.e., putting pairs of images into correspondence. This part is technically the most challenging because precise mammogram registration is intractable. This is due to the fact that these images correspond to compressed, elastic 3-D objects and the images differ primarily due to the fact that there are variations in positioning and compression. These variations are in essence changes in viewpoint and cannot be modeled based on available information, and thus it is impossible to countereffect 3-D changes in viewpoint based on the available 2-D images. The concentration of this work is on developing an alternative to precise registration by defining corresponding regions in two images, as is done by medical experts. These regions can then be analyzed for changes indicative of cancer

Technically, regional registration of mammograms is the problem of identifying similar regions in weakly structured texture patterns The proposed approach logically divides into two steps The first step involves establishing landmark points in a mammogram pair and establishing their correspondence The second step involves formation of regions using the landmark points. The paper describes both steps and is organized as follows. The algorithm for extraction of potential control points is described in Section 2, the matching procedure together with evaluation is outlined in Section 3 Section 4 describes using the points to form the corresponding regions. Section 5 summarizes the results and discusses continuation of the work

2 Identification of landmark points

Considering that mammograms are nonstructured texture patterns and that the relationship between two images cannot be modeled, establishing regional correspondence requires at least determining a set of landmarks common to both images. These landmarks can be used as reference points in establishing locations and geometry of corresponding regions Based on the fact that the only distinct elements appearing in practically any

mammogram are elongated structures, the most prominent of which correspond to milk ducts and blood vessels, we select the intersections between these structures to be the landmarks After extracting the intersections in both images we establish correspondence between the common subset of points. In order to make establishing correspondence practical, we consider only the intersections that are the most likely to appear in both mammograms; consequently, we limit the extraction to the intersections between the most prominent elongated structures. The obtained points are referred to as the potential control points Locations of the potential control points may vary in two images due to minor positioning differences and changes in compression; the algorithms described in this paper handle minor variations and reject images where correspondence between image pairs cannot be established with confidence.

The identification of the points is based on the model for elongated structures proposed in [3] which effectively shows that the cross-sections of elongated structures are characterized by roof edge profiles. Consequently, extraction of the elongated structures is, in essence, a task of identifying peaks of the roof edge profiles. Similar tasks are encountered in other applications, e.g. in fingerprint analysis, line drawing understanding, and digital angiography Many researchers have investigated this problem and have proposed appropriate methods, however the emphasis of proposed algorithms is on detailed image analysis, extraction of every roof edge with the objective of retaining continuity of the structures. Mammograms are complex images with many intensity variations, thus, extraction of every roof edge yields overwhelming detail Instead, we propose an alternative algorithm which identifies only the most prominent roof edges Specifically, we limit the identification to the widest roof edge profiles. In the following we summarize the processing steps, detailed exposition of the algorithm can be found in [10]

Prior to applying the proposed operator, images are smoothed using the multi-pass Gaussian filter The filter is approximated by a 5 x 5 kernel, as described in [2] Smoothing removes minor intensity variations and does not in effect change the intensity profiles. The roposed operator considers two concentric neighborgoods of different size a large neighborhood m x n and a small neighborhood k x l The operator assigns a value to an image pixel by counting the pixels in the large neighborhood that have intensity values smaller than the smallest intensity value in the smaller neighborhood Next, the images are turned into binary form by utilizing a threshold, T The parameters used with the set of mammograms described in this study are as follows: (i) detection of elongated vertical structures $m = 11$, n $= 1$, $k = 3$, $1 = 1$, and $T = 5$, and (ii) detection of elongated horizontal structures: m $= 1$, $n = 11$, $k = 1$, $l = 3$, and $T = 5$. The rationale for selecting these values is explained in [10]

The operator is applied in two steps, one that extracts predominantly horizontal structures and the other that extracts predominantly vertical structures The elongated structures in arbitrary directions are obtained by combining the two results by the logical OR operation, and their intersections are obtained by combining the two results by the logical AND operation The centroids of obtained binary objects (typically smaller than 10 pixels) constitute the set of potential control points An example of the extracted points in a mammogram pair is shown in Figure 1

The performance of the algorithm was evaluated relative to effects of width of linear structures, their contrast relative to background, background type, and noise on algorithm's performance Evaluation was carried out one synthetic images obtained by superimposing elongated structures generated by the Poissonian line model, [6], onto background of varying contrast, texture patterns and real mammograms. The results were summarized in form of ROC curves which show the true positive fraction and false positive fraction of pixels detected by the algorithm, [4], [5] The evaluation has confirmed that the parameter selection was appropriate for the resolution of images considered in the study Furthermore, the algorithm is insensitive to change in the structure's width for structures wider than one pixel and its performance is not affected by the type or contrast of the background, [9].

3 Establishing control point correspondence

Given two sets of potential control points, the objective of the correspondence algorithm is to identify the common subset and within it individual point correspondences Since the relation between the two images in the case of mammograms is unknown, it is necessary to use a model-free matching algorithm

Matching is attempted between a point $p(x_p, y_p)$ in the older mammogramand a point $q(x_q, y_q)$ in the newer mammogram only if point q lies in a region likely to provide a match[1] This region is determined using a pair of reference points $o(x_o, y_o)$ and $n(x_n, y_n)$ in the old and new mammogram, respectively The correspondence between o and n is established accurately, either manually or by using an appropriate algorithm. The algorithm used in this work to establish the locations of o and n is described in [10] Point q is a match of point p if it lies in the neighborhood k x 1 centered at x_c, y_c determined by the intersection[2] of the line

$$y = \frac{y_p - y_o}{x_p - x_o}(x - x_n) + y_n \qquad (1)$$

and the circle

$$(x - x_n)^2 + (y - y_n)^2 = (x_o - x_p)^2 + (y_p - y_o)^2 \qquad (2)$$

Values k and l are determined by expected differences between the two mammograms and in this work they are set to 60 pixels.

The correspondence is established by measuring similarity between regions around potential matches using accumulator matrix as follows.

- *Measuring similarity between potential matches* A match between points is established using signatures and a similarity criterion

  - *Signature formation* The idea behind the signatures can be best explained when considering the retino-cortical projection, i.e., a geometrical transformation described as a conformal mapping of the polar $(p, \mathbf{0})$ plane onto the Cartesian $(log(\rho), \mathbf{0})$ plane. A point $z(\rho, \mathbf{0}) =$

[1] In principle, the proposed algorithm does not require limiting point locations; however. constraining the search to image subregions significantly speeds up the processing time.

[2] There exist two points of intersection and the one of interest lies in the area of breast tissue since the reference point is the tip of the nipple.

$\rho e^{j\theta}$ in the retinal plane maps onto the point $w(u,v) = u + jv$ in the cortical plane as

$$w = log(z) = log(\rho) + j(\theta + 2\pi k), \quad k = 0, \pm 1, \tag{3}$$

When taking into account only the principal branch of the logarithmic function, the representation of a cortical projection point is $u = log(\rho), \quad v = \theta$ The transformation may be implemented efficiently in a form of smart sensor, [7],[8] Signature vectors are obtained from the cortical images by integrating the images along the p axis in intervals $\Delta\theta$ It is necessary that the sampling rates along the θ axis provide for local deviations from linearity; through experimentation the sampling rate is selected to be $\Delta\theta = 20^0$ The high values in a signature correspond to the elongated structures forming the intersection Typically, a signature has 2-4 maxima. The locations of these maxima are encoded by thresholding the signature entries Assuming that the length of an elongated structure encoded by the signature is l and its average width is w, the threshold value is selected to be $\frac{1}{2}(l \times w)$ Coefficient $\frac{1}{2}$ is due to the fact that the signature encompasses a wedge which covers approximately one half of the rectangular area that may be encompassed by the elongated structure. Thus, for mammograms considered in this work, where based on resolution on average $w = 4$, and $l = 10$, the threshold is set to 20

- *Similarity criterion* Two signatures $S_1 = [s_k^1]$ and $S_2 = [s_k^2]$, $k = 1, 2, \ldots, \frac{360^0}{\Delta\theta}$ match if $r \geq t$. where

$$r = S_1 S_2 = \sum_{k=1}^{360^0/\Delta\theta} s_k^1 s_k^2, \tag{4}$$

and t is the threshold Since the signatures are binary vectors, the similarity criterion given by Equation (4) counts the number of 1's that appear at the same positions in two signature vectors. Assuming that at least two parts of elongated structures are captured by a signature, we select $t = 2$

- *Accumulator matrzx* This matrix tallies votes for possible matches Given a set of N_o potential control points in the older mammogram, M_o, and a set of N_n potential control points in the newer mammogram, M_n, the accumulator matrix is an $N_o \times N_n$, where entry $e(p,q)$ corresponds to points labeled p and q in M_o and M_n, respectively Entry $e(p,q)$ is updated each time a possible match between points p and q is established

  - *Step 1 Locatzon test* For each point in the old mammogram any potential control point in the new mammogram that satisfies the location criterion, i.e., it lies in the neighborhood $k \times l$ centered at (x_c, y_c) determined by Equations (1) and (2), is considered.
  - *Step 2: Matrix updating* If a matching criterion between a point p in the old mammogram and a point q in the new mammogram is satisfied, the accumulator entry $e(p,q)$ is updated In the case that no match is established, return to Step 1, otherwise, proceed to Step 3
  - *Step 3: Processzng the remaznzng poznts* If points $p(x_p, y_p)$ and $q(x_q, y_q)$ are matched in Step 2, then for each remaining point in the old mammogram, $r(x_r, y_r)$, a neighborhood of size $\Delta_2 \times \Delta_2$ of point (x_c', y_c') in the new mammogram is examined for a possible match The coordinates (x_c', y_c') are determined by the intersection of the line

$$y = \frac{y_r - y_p}{x_r - x_p}(x - x_q) + y_q \tag{5}$$

and the circle

$$(x - x_q)^2 + (y - y_q)^2 = (x_r - x_p)^2 + (y_r - y_p)^2 \tag{6}$$

Each time a matching criterion is satisfied, the appropriate accumulator entry is updated After considering all possible matches return to Step 1

The signatures of 29 mammogram pairs were analyzed to establish successfully that (i) control points are characterized by unique signatures, and (ii) signatures are preserved in subsequent screenings, [10]

At the end of the matching process, the entries which correspond to true matches tend to have high values while the ones that correspond to wrong matches are small. The wrong matches are eliminated by thresholding. A good threshold value provides high values both for sensitivity (or the fraction of the established matches computed relative to the all possible matches) and specificity (or accuracy of the established matches) The emphasis of this work is on high specificity and, therefore, in some cases the number of established matches is relatively small

Two approaches to thresholding the accumulator matrix were considered The first one involved an optimum thresholding algorithm and the second used for obtaining results in this paper used a fixed threshold value of 80% of the maximum entry This approach was verified through algorithm evaluation, Section 3.2 After thresholding, the accumulator matrix is scanned and the most likely matches are determined Scanning is done by first examining rows and then columns of the accumulator matrix. If there is more than one non-zero entry in a particular row, then the point in the new mammogram which corresponds to the largest accumulator entry is selected as the best match The rest of the entries in the same row are set to zero. Columns are scanned in a similar fashion. After scanning, the accumulator matrix contains non-zero entries only at positions which correspond to the final matches. As an example, matches established between potential control points in Figure 1 are shown in Figure 2

31 Algorithm evaluation

The purpose of the tests was to determine the robustness of the correspondence algorithm with respect to parameters employed by the algorithm The evaluation of the correspondence algorithm was carried out in two stages, (i) using synthetic images and (ii) using archival film Each image pair comprised an original and a corresponding deformed pattern The underlying model in generating the original images was the Poissonian model and deformations were modeled using the

deflected elastic beam model [9]. For each test, a set of **100** image pairs were used and the results were summarized in form of ROC curves. The experiments are described in detail in [9] The conclusions of the experiments were that both **good** specificity and sensitivity of the algorithm can be achieved, since for true positive fraction of about **97%** false positive fraction is smaller than 5%. Furthermore, it was established that the accumulator matrix threshold can always be determined to accommodate for high specificity, and thresholds of 80% of the maximum accumulator entry are always a good selection In general, the algorithm's performance decreases with the increase in distortion, but even for high levels of distortion a working point with low false positive fraction can be achieved for T=5, [9]

The second part of the evaluation was carried out on **31** mammogram pairs representing screenings in standard cranio-caudal and medio-lateral views. Due to lack of ground truth regarding control points, the evaluation was performed by visual inspection. Validation was made independently by three unbiased observers, i.e., observers who had no knowledge about the algorithm. an experienced radiologist and two untrained observers. The validation experiment was conducted as follows Each observer was shown a mammogram pair, where the older mammogyam contained highlighted points selected by the algorithm The observer was then asked to identify the corresponding points in the newer mammogram Observers were also asked to assign a "level of confidence" to the established matches. Namely, if a match was established with a high degree of certainty, the confidence level was "sure", if a match was determined with a lower degree of certainty, the confidence level was "not sure", in the case when no match could be assigned, a point was labeled "don't know" No time limit was imposed upon observers during the validation experiment, The observers were trained and encouraged to use enhancement techniques in the experiment (e.g., histogram equalization) Based on the spatial resolution [for majority of mammograms 1 mm/pixel), the algorithm and an observer were considered to be in agreement if the distance between the two points was less than 10 pixels. The algorithm established correspondence between **11** pairs of points on average, this is in contrast to the earlier work, [10], that was establishing correspondence between 5 pairs on average. Table 1 summarizes the validation results when only points labeled "sure" were considered Each entry in the table represents the percentage of points "in agreement" For example, the entry *(algorithm,radiologist)* = 91 indicates that **91%** of "sure" control points determined by the radiologist are "in agreement" with corresponding points determined by the computer Table 2 includes both "sure" and "not sure" points.

The validity of the evaluation, i.e. consistency of performance and lack of bias was established by presenting three pairs of mammograms eight times to two of the observers. In the first four presentations, the observers were shown the identical mammogram pair In the fifth, sixth, and seventh presentations, the mammograms were flipped around x axis, around y axis, and around both x and y axes, respectively In the eighth presentation only a portion of the mammograms that contained points of interest were shown, i.e., the images contained no visible breast outline. This was done to check if the human perception of control points is affected by landmarks such as image borders and/or breast outline. The obtained results were then compared using five statistical tests. This evaluation concluded that regardless of the mammogram type the human perception is consistent [9].

	radiologist	observer 1	observer 2
algorithm	**91**	71	75
radiologist		**83**	82
observer 1		-	78

Table **1** Result of validation when using signatures and considering only "sure" points. Entries represent the percentage of points "in agreement"

	radiologist	observer 1	observer 2
algorithm	86	72	69
radiologist	-	78	77
observer 1	-	-	67

Table 2: Result of validation when when using signatures and consideringboth "sure" and "not sure" points. Entries represent the percentage of points "in agreement"

4 Establishing regional correspondence

The corresponding regions are selected based on the control points and similarity of texture characteristics In the present research the regions aye circular. The circular geometry is selected because it is invariant to rotational misalignments arising from image digitization and requires determinin only the corresponding centers and radii. The essentiaf:elements of the procedure are illustrated in Figure **3,** and they involve the following

- *Determining the extent of corresponding regzons* The radii of the corresponding regions are the same and are determined so that the corresponding regions have similar texture characteristics (this is important for the comparison paradigm) Partitioning of the older mammogram is used to ensure that each region is characterized by predominantly homogeneous intensity pattern Partitioning employs a hierarchical region growing paradigm that uses pyramidal multiresolution image representation. Relationships between pixels at different resolutions are established by a fuzzy membership function as described in [1] The selection of the parameters of the fuzzy membership function allows for fine-tuning the method to specific segmentation objectives. Additional criteria may be used to limit a region extent, e.g., regions may be limited to a specific size dictated by the spatial resolution and expected types of cancerous changes

- *Covering of a mammogram pair* The covering algorithm generates corresponding circular regions and covers an image pair with overlapping circles with no gaps. The distances between landmark points are used to determine the centers and the radius for each pair is such that it ensures that the region in the other mammogram belongs predominantly to a single partition Since the distances between control points are not guaranteed to be preserved in two screenings, the region centers are determined as the cluster centers of elements obtained by considering all pairwise combinations of matched control points.

Figure 4 shows a subset of corresponding regions established for mammograms shown in Figure 2.

5 Summary and conclusions

The focus of this work is on identifying landmarks and establishing their correspondence in temporal pairs of mammograms. The motivation of the work is automation of analysis of mammogram sequences. Two algorithms were developed and tested The first algorithm identifies potential control points in mammograms, and it was evaluated on synthetic images and was shown to be robust to changes in width of elongated structures, background complexity, and presence of noise. The second algorithm establishes correspondence between potential control points, and it was evaluated both on synthetic images and archived film. One of the major challenges of the work was developing evaluation protocols because no "groundtruth images" exist. The bulk of the evaluation involved human subjects, and it was necessary to evaluate reliability of performance of an observer and to quantify consistence of the observer's visual perception of control points selected by the algorithm The most important conclusions are that there is the agreement in **90%** of cases between the algorithm and the medical expert.

Our present research focuses on evaluating the algorithms for establishing similar regions The appropriate region size and geometry are under study The ultimate test of performance will be in the next stage of development when the complete system performance, including characterization of corresponding regions, is compared to results of biopsies

References

[1] D Brzakovic and M Neskovic "Mammogram Screening Using Multiresolution-based Image Segmentation," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. **6,** pp **1437-1459, 1993** Also in *State of the Art on Digital Mammographic Imaging*, K Bowyer and S Astley (Eds.), World Scientific, **1994.**

[2] P J Burt, "Fast Filter Transforms for Image Processing," *Computer Graphics and Image Processing*, vol 16, **1981**

[3] N Cerneaz and M Brady, "Finding Curvi-Linear Structures in Mammograms," Proc of the First International Conference, CVRMed **'95,** Nice, France, pp. **372-382, 1995**

[4] C E. Metz, "ROC Methodology in Radiographic Imaging," *Investigative Radiology*, vol. **21,** pp 720-733, **1986**

[5] C.E. Metz, "Statistical Analysis of ROC Data in Evaluating Diagnostic Performance," m *Multiple Regression Analysis: Applications in Health Sciences*, Medical Physics Monograph, Eds. D E Helbert and R.H Myers, vol. **13,** pp **365-384, 1986**

[6] R.E. Miles, "A Various Aggregates of Random Polygons Determined by Random Lines in a Plane," *Adv. in Math*, vol. **10,** pp **256-290, 1973**

[7] G Sandini and P Dario, "Active Vision Based on Space-Variant Sensing," *Proc. of the Fifth International Symposium on Robotics Research*, Tokyo, Japan, **1989**

[8] G Sandini and V Tagliasco, "An Antropomorphic Retina-Like Structure for Scene Analysis," *Computer Graphics and Image Processing*, vol **14,** pp 362-372, **1980**

[9] N Vujovic, Registration of Time-Sequences of Random Textures with Applications to Mammogram Followup, Ph D Dissertation, Lehigh University, 1997

[10] N Vujovic and D Brzakovic, "Establishing Correspondence Between Pairs of Mammograms," *IEEE Trans. on Image Processing*, vol 6, no. 10, pp. **1388-1399, 1997**

ACKNOWLEDGMENT

This work was supported b the NSF under Grant **IRI-9504363** and by the **U.8.** Army under Grant **DAMD17-96-1-6128** The mammograms used in the study were supplied by the St Luke's Hospital Bethlehem, PA The authors thank Dr P Brzakovic, O Alsayegh and P Bakic for their help in algorithm evaluation

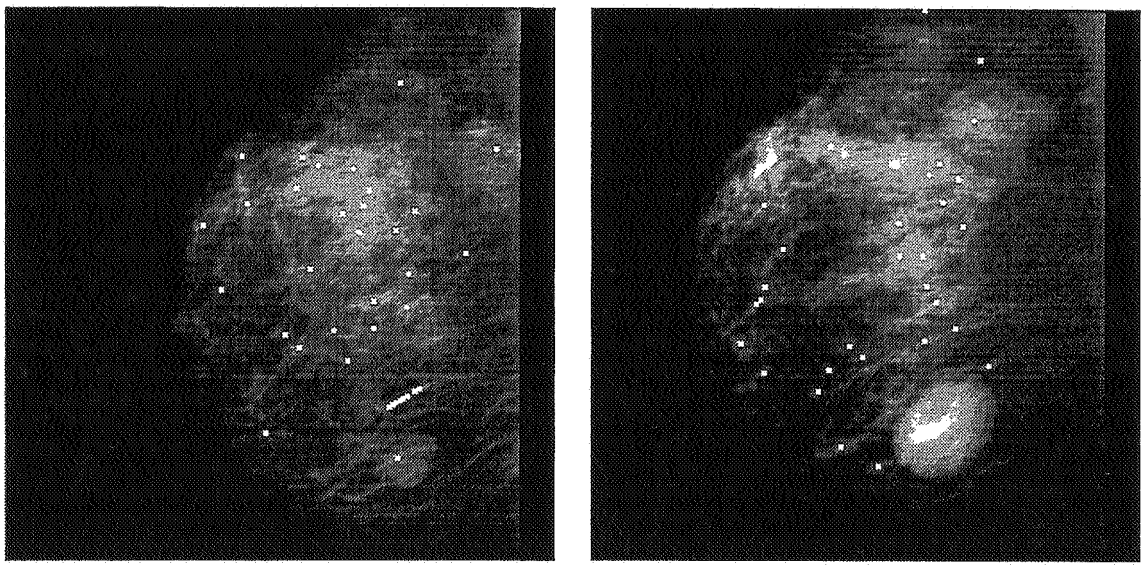


Figure 1 An example of detected potential control points in a mammogram pair The points are superimposed on the images

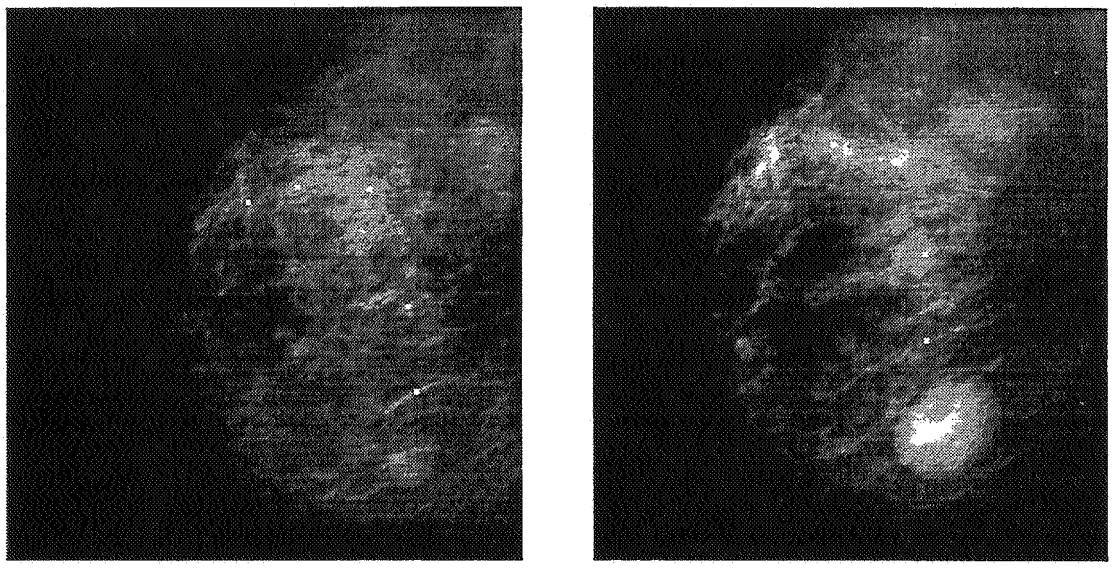


Figure 2 Matched points between potential control points shown in Figure 1

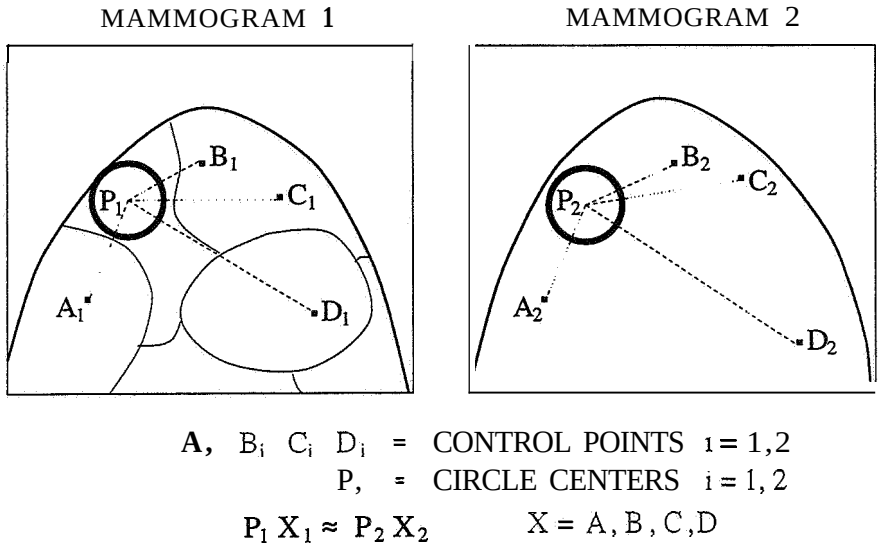


Figure 3 Conceptual description of formation of corresponding regions. The region centers are determined based on the distances from the matched control points and the extent is determined by the segmentation of the older mammogram.

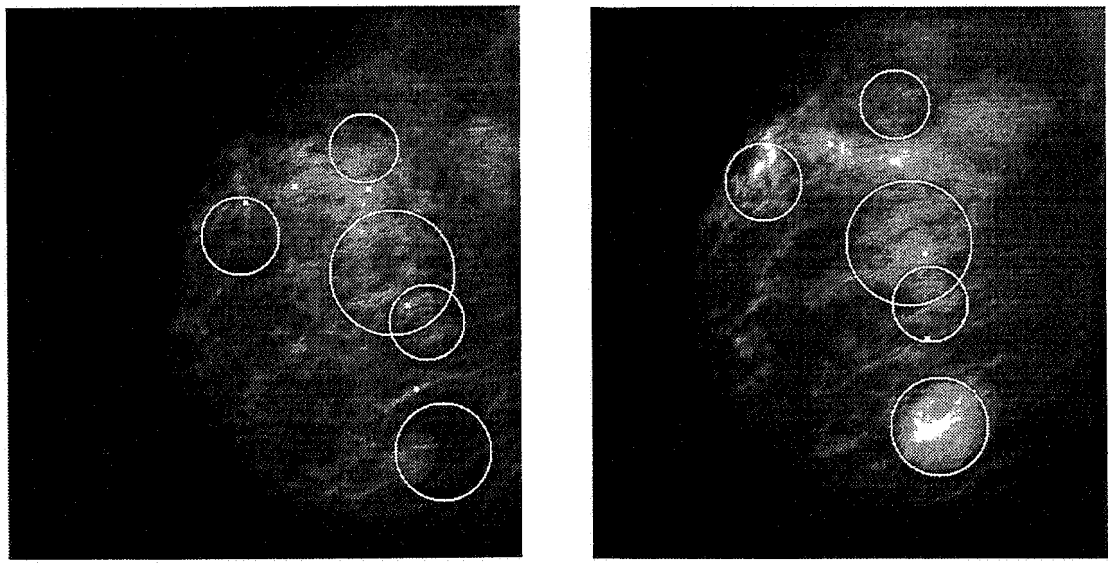


Figure 4 An example of a subset of corresponding regions established based on the control points shown in Figure 2

Session VI

General Methods with Application to Medical Imagery

Consistent Feature Extraction Based on Steerable Filters and Application in Recovering Rigid Transforms

Maha Sdam, Kyong Chang and Kevin Bowyer
University of South Florida
Tampa, FL 33620
e-mail: sdam,chang,kwb@csee.usf.edu

Abstract

an image operator is proposed for extracting consistent features of interest in corresponding images. The feature selector is based on using steerable filters for extracting translation and rotation invariant features As an application of the feature selector, we describe a registration technique for recovering rigid transforms in images which contain both textured and non-textured regions Consistency of feature extraction allows establishing correct correspondence between a number of features which can be used to calculate an accurate measure of the transform between a pair of images. The method is fully automated and is highly robust. The technique was applied to 14 pairs of mammogram images where each pair represents two instances of a mammogram film originally digitized at different resolutions. Performance of the Automated registration method is compared to performance of a similar method based on manually selected and matched features. The technique was also applied to 71 pairs of mammograms as part of a process for transcribing radiologist markings of abnormal regions in mammograms.

1 Introduction

Traditional correlation based techniques are only feasible for recovering small transforms between corresponding images. Large movement creates a search space too large to be traversed in a reasonable amount of time. This is applicable to most types of image transforms including movement defined by simple translation and rotation. The time complexity of these techniques remains high even when the search is not performed exhaustively, or if correlation is calculated for subregions of the images rather than the full images (see [6] for a survey of correlation–based, and other, registration techniques) An alternative method is selecting and matching corresponding interest features across the two images Feature selection is performed independently in each image then features are matched based on their respective image properties.

Earlier techniques such as the Moravec operator [3, 10] extract features based on local image variance. Although they attempted to measure directional variance to adapt to rotation changes, the number OF directions considered was too small to allow consistent feature extraction. Other researcher [9, 11] used image features such as lines and regions to locate feature points. These techniques, especially the region–based ones, are dependent on image content and may fail in highly textured regions or regions not fully included in the images. Another method which depends on image content [12] searches for specific structures using template matching then iteratively refines the feature locations.

Depending on the degrees of freedom of the transform, a minimum number of matched features is required to calculate the transform between the images. The main challenge in a fully automated process is selecting a subset of features consistently across corresponding images. In practice, it is difficult to design generic image operators capable of robustly extracting a small number of consistent features across images related by some transform, even if the images are very similar As a result, some registration methods require manual selection and matching of features in order to ensure proper performance. Thls paper introduces a translation and rotation invariant image operator for extracting features more consistently across corresponding images Rotation invariance is achieved through using steerable filters The interest feature selector is a generic image operator which can be applied to both textured and non–textured image regions. This is a desirable property whch allows registration irrespective of image content

To illustrate the use of the proposed operator it is applied in an algorithm for extracting features across images related by a rigid transform which may include a large degree of rotation and translation. A rigid transform can be recovered using only two matched pairs of features, provided the features are matched perfectly. Extracting a larger set of features using less consistent operators improves the chances that the same features will be extracted in both images, however, it also increases the processing required for matching and eliminating incorrect matches Using a rotation and translation invariant operator allows extracting a smaller set of features without compromising the degree of consistency required

While the interest selector and the technique described are in no way application specific, they were motivated by the need for registration of digitized instances of mammogram images. As image analysis techniques become more mature, there is an ever increasing need for images with established ground truth against which performance of various techniques can be measured. Images with ground truth are also essential for learning optimal parameters for many al-

gorithms. This is especially the case in mammogram image analysis which has become an active research area in recent years [5, 1, 2].

For mammogram images, as well as most medical images, ground truth is established by experienced radiologists. Until electronic display technologies meet the high requirements of displaying mammogram images, and until radiologists become experienced in reading electronically displayed images, the most reliable "ground truthing" method is for radiologists to mark regions of interest on the actual film. One way of transcribing these markings in electronic form requires digitizing the film without markings, and with markings. The two digitized versions can then be registered to establish the exact location of the markings in the digitized unmarked version. This is needed since the unmarked version is clearly the one to be used in training or measuring performance of image analysis techniques.

The following section provides an overview of steerable filters based on Gaussian derivatives and introduces a translation and rotation invariant interest feature selector Section Three describes a general method for recovering the rigid transform. Section Four provides experimental results on 14 pairs of corresponding images used to evaluate the registration method and results on additional 71 pairs of images The last section provides discussion and conclusions.

2 Selecting Interest Features

In selecting candidate features for matching across images, it is not feasible to consider all pairs of pixels in the two images as potential matches even if matches are known to lie within a local region. An exhaustive search on all pixels is extremely inefficient and is very susceptible to matching non-corresponding pixels based on accidental similarities between their local properties. A better solution is extracting a relatively small set of interesting points or features to be used as candidates for matching. An interesting point is one which is distinct from its surrounding pixels in some sense.

One of the main challenges in identifying interest features is extracting features consistently across the images being matched. As a result of the transform between images, features will shift and rotate in one image relative to their corresponding features in the other image. An image operator sensitive to shift and rotation variations will fail to recover features consistently. Designing basic image operators invariant to such changes has gained attention in recent years [8]. This section describes an invariant feature selector which attempts to extract features without bias towards their location or orientation in an image

A basic class of angularly adaptive image operators called steerable filters is of particular interest in this work. The most useful property of steerable filters in the context of selecting features for matching is that their response is not biased towards a particular orientation or set of omentations. These operators can be applied to every image point and hence can become shift invariant in responding to features adequately sampled in both images. These properties make steerable filters suitable candidates for constructing a consistent interest point selector

2.1 Steerable Filters

The information presented in this subsection was adapted from the paper by Freeman and Adelson [8] and is included here for completeness. As described by Freeman and Adelson, steerable filters are oriented linear operators that can be specified for an arbitrary orientation from a linear combination of basis filters. The idea is that the space of all rotated versions of a function with limited angular frequency can be adequately sampled by a finite, and possibly small, set of basis functions. Basis filters can be applied to an image in order to generate a fixed set of filtered images which can be linearly combined to generate an image filtered at an arbitrary orientation Derivatives of Gaussian functions are often used to detect locally maximal changes m image gray scale. These functions have a finite angular frequency and hence they are steerable. As an example, consider a circularly symmetric Gaussian function of the form.

$$G = e^{-(x^2+y^2)} \tag{1}$$

Local changes in gray scale along the horizontal direction in an image can be detected using a linear filter constructed from sampling the partial derivative of G with respect to x (or the first derivative of G along onentation 0):

$$G_1^0 = \frac{\delta G}{\delta x} = -2xe^{-(x^2+y^2)} \tag{2}$$

To represent gray scale variation along the vertical direction, the function $\frac{\delta G}{\delta y}$ (or G_1^0 rotated by $90°$) can be used:

$$G_1^{90} = \frac{\delta G}{\delta y} = -2ye^{-(x^2+y^2)} \tag{3}$$

G_1^0 and G_1^{90} are basis functions for the space of first derivatives of G along all possible directions. For an arbitrary angle 8, G_1^θ can be constructed from a linear combination of the two basis functions defined by

$$G_1^\theta = cos\theta G_1^0 + \sin\theta G_1^{90} \tag{4}$$

Since the basis filters are linear, generating an image by linearly combining two image versions filtered with each of the basis filters is equivalent to generating the image by applying a linear combination of the basis filters. The advantage of performing the earlier operation is that it completely specifies the filter response as a function of its rotation angle while the latter specifies a single value of that function. Specifying a filter response as a function of its rotation angle allows for analytically calculating the filter orientation with maximal response.

Freeman and Adelson explicitly derive x–y separable basis filters for second order Gaussian derivative at an arbitrary orientation G_2^θ. They also derive an expression for the Hilbert transform H_2^θ of G_2^θ, and the x–y separable basis functions needed to steer H_2^θ The Hilbert transform introduces phase independence so that the maximal filter response estimate is invariant to phase changes. The derived expressions for G_2^θ and its basis functions G_{2a}, G_{2b} and G_{2c} are:

$$\begin{aligned} G_2^\theta &= cos^2(\theta)G_{2a} - 2cos(\theta)sin(\theta)G_{2b} \\ &\quad + sin^2(\theta)G_{2c}, where \\ G_{2a} &= 0.9213(2x^2 - 1)e^{-(x^2+y^2)}, \\ G_{2b} &= 1.843xye^{-(x^2+y^2)}, and \\ G_{2c} &= 0.9213(2y^2 - 1)e^{-(x^2+y^2)} \end{aligned}$$

H_2^θ and its basis functions H_{2a}, H_{2b}, H_{2c} and H_{2d} are defined by:

$$
\begin{aligned}
H_2^\theta &= cos^3(\theta)h_{2a} - 3cos^2(\theta)sin(\theta)H_{2b} \\
&\quad +3cos(\theta)sin^2(\theta)H_{2c} - sin^3(\theta)H_{2d}, where \\
H_{2a} &= 0.9780(-2.254x + x^3)e^{-(x^2+y^2)}, \\
H_{2b} &= 0.9780(-7515 + x^2)(y)e^{-(x^2+y^2)}, \\
H_{2c} &= 0.9780(-7515 + y^2)(x)e^{-(x^2+y^2)}, and \\
H_{2d} &= 0.9780(-2.254y + y^3)e^{-(x^2+y^2)}
\end{aligned}
$$

The basis functions can be sampled to generate filters which can be applied to an image. The resulting 7 filtered images can be used to calculate the angle of dominant orientation θ_d at each mage pixel and the strength of that orientation s_d using the following expressions:

$$
\begin{aligned}
\theta_d &= \frac{arg[C_2, C_3]}{2}, and \\
s_d &= \sqrt{C_2^2 + C_3^2}, where \\
C2 &= 0.5([G_{2a}]^2 - [G_{2c}]^2) \\
&\quad +0.46875([H_{2a}]^2 - [H_{2d}]^2) \\
&\quad +0.28125([H_{2b}]^2 - [H_{2c}]^2) \\
&\quad +0.1875([H_{2a}][H_{2c}] - [H_{2b}][H_{2d}]), and \\
C3 &= -[G_{2a}][G_{2b}] - [G_{2b}][G_{2c}] \\
&\quad -0.9375([H_{2c}][H_{2d}] + [H_{2a}][H_{2b}]) \\
&\quad -1.6875[H_{2b}][H_{2c}] - 0.1875[H_{2a}][H_{2d}]
\end{aligned}
$$

It is this orientation and strength of maximal response that are used in extracting features of interest as will be explained in the following subsection.

2.2 The Feature Selector

Filters created by sampling the basis functions of G_2 and H_2 can be used to create a map of dominant orientation and strength of that orientation at each image pixel. For an image $I(x, y)$, these maps are denoted by $O(x, y)$ and $S(x, y)$ respectively The orientation estimate is most accurate for pixels which have a single dominant orientation of grey-scale change. Pixels at which there are overlapping structures will have less reliable orientation estimates reflected by their lower orientation strength.

Some textured image regions may contain overlapping structures which reduce the reliability of orientation estimate at many image pixels. This is also the case at points of sharp orientation changes such as corners and locally maximal curvature points. To reduce this effect, each pixel is assigned an orientation measure given as a weighted average of orientations in a region centered around the pixel. Each orientation is weighted by the relative orientation strength at its respective pixel. The smoothed orientation map generated is denoted by $O_s(x, y)$,

$$
O_s(x_i, y_j) = \frac{\sum O(x_m, y_n)S(x_m, y_n)}{\sum S(x_m, y_n)} \qquad (5)
$$

m and *n* index through a window centered at (x_i, y_j) In the current implementation, the window size is 5x5 pixels. This averaging has the effect of assigning an intermediate orientation to pixels at which overlapping structures, or sharp orientation changes are present. This intermediate orientation is biased towards the strongest orientation in the surrounding region

A measure can be assigned to each pixel based on the difference between its onentation and the orientations of immediately surrounding pixels. This measure is defined as:

$$
M(x_i, y_j) = min\{|O_s(x_i, y_j) - O_s(x_m, y_n)|\} \qquad (6)
$$

m and *n* index through the eight neighbor pixels of (x_i, y_j).

The original image orientation map O(x,y) assigns orientations in a possible range of $[-\frac{\pi}{2}, \frac{\pi}{2}]$. $O_s(x, y)$ is a normalized weighted average of $O(x, y)$, so it also has the same possible range. Each pixel in $O_s(x, y)$ can be considered to have an intermediate orientation between its neighboring pixels and hence the range of $M(x, y)$ is approximately $[0, \frac{\pi}{2}]$ A minimum difference of 0 suggests strong neighbors with similar orientations. A minimum difference of $\frac{\pi}{2}$ suggests strong neighbors with orientations of $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ A difference of $\frac{\pi}{4}$ suggests neighbors that differ in orientation by $\frac{\pi}{2}$ A difference of $\frac{\pi}{2}$ represents maximum distinctness between orientations.

A distinctness measure $D(x_i, y_j)$ is assigned to each pixel (x_i, y_j) based on the value of $M(x_i, y_j)$ as follows:

$$
D(x_i, y_j) = \begin{cases} M(x_i, y_j) & \text{if} M(x_i, y_j) < \frac{\pi}{4} \\ \frac{\pi}{2} - M(x_i, y_j) & \text{if} M(x_i, y_j) > \frac{\pi}{4} \end{cases}
$$

This measure reflects how distinguishable a point is from its most similar neighbor The measure has a maximum value when a pixel has an intermediate orientation between strong neighbors that differ by $\frac{\pi}{2}$ Points with a locally maximal distinctness measure are the points of interest.

3 Recovering Rigid Transforms

Recovering transforms between images based on a small set of matched features is an efficient registration technique. The success of the technique is completely dependent on accurate selection and matching of features. Using the interest feature selector described in the previous section, it is possible to design a highly robust method for recovering rotation and translation between a pair of images. With appropriate modifications, the overall approach is applicable to more complicated types of image transformations. The interest selector is also useful in images which contain considerably more complex deformations since it is adaptive to local changes in orientation that do not necessarily have to comply with a simple global rotation. Figure *1* outlines the registration method described in the following subsections.

3 1 Matching Features

After extracting features in each image, a match is estab lished between pairs of features across the two images. To determine this match, a pool of potential matching features is created in the second image for each feature in the first image. The pool is simply formed using features in the second image which lie within a window centered around the location of each feature in the first image. The size of the window can be as large as needed to capture the degree of displacement and rotation possible between the two images. Once this requirement is satisfied, further increase in the window size will increase the processing needed for

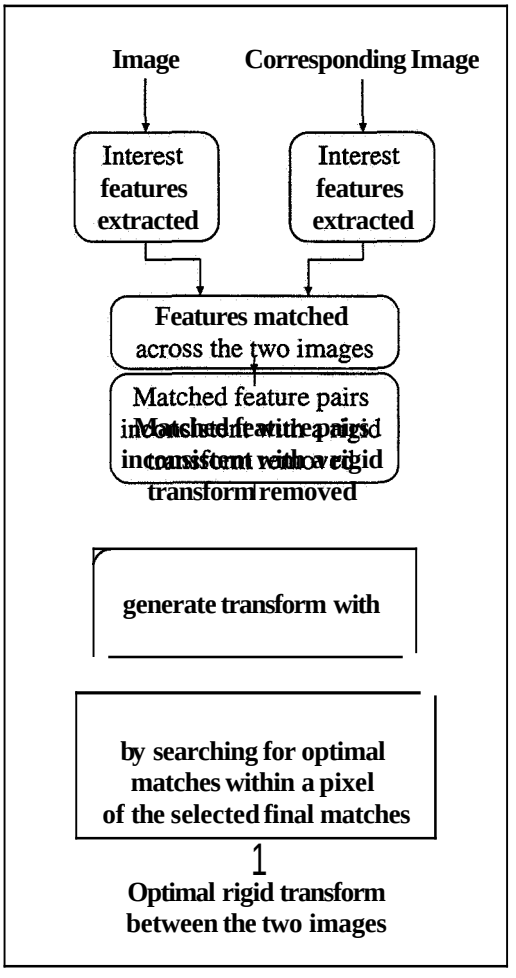


Figure 1 General approach used for recovering rigid transforms.

matching. It can also make the matching more difficult by introducing too many potential matches.

Matching each feature to one of the members in its pool of potential matches can be assigned an error measure inversely proportional to the degree of similarity between the features. A possible measure is the square root of the average square difference between corresponding regions centered around each of the two features in its respective mage [7]. If the first image is denoted by $FI(x,y)$ and the second image by $SI(x,y)$, then a possible error measure for matching feature A at point (x_i, y_j) in FI to feature B at point (x_k, y_l) in SI is given by

$$E = (\sum (SI(x_{k+m}, y_{l+n}) - FI(x_{i+m}, y_{j+n}))^2)^{\frac{1}{2}} \qquad (7)$$

m and n index through a square region centered around each feature in its respective image.

Because of the transform which takes place between the two images, this measure will potentially fail to capture the similarity between many true matches by incorrectly assigning their match a large error value. To correct for the transform which takes place around the features being matched, it is possible to use the orientation of maximal grey–scale change at each feature. This is the estimate calculated for each image pixel in the process of extracting the features of interest in the previous section. Before calculating the difference between the local regions around the features, the local region in the second image is rotated to compensate for the difference in dominant grey–scale change orientation between the first image feature and each potential match. The error measure corrected for possible change in feature onentation is given by:

$$E_\theta = (\sum (SI(\tilde{x}_{k+m}, \tilde{y}_{l+n}) - FI(x_{i+m}, y_{j+n}))^2)^{\frac{1}{2}},$$
$$\tilde{x}_{k+m} = x_{k+m}\cos\theta - y_{l+n}\sin\theta, \text{and}$$
$$\tilde{y}_{l+n} = x_{k+m}\sin\theta + y_{l+n}\cos\theta$$

Where θ is the orientation at the feature in the second image relative to the onentation at the feature in the first image. The transformation applied to the original (x_{k+m}, y_{l+n}) points in equation 7 to calculate E_θ is a simple rotation transform defined by the rotation matrix:

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

Once an error measure is assigned to matching each first image feature to each of its potential matches, a greedy method is used to establish the final match. The method proceeds by matching a pair of features which produces the minimum error among all potential matches, eliminating that pair from future consideration, and then repeating the process using the remaining potential pairs of matches.

3.2 Removing Inconsistent Pairs

Since the type of transform between the images is known, it is possible to obtain a highly accurate set of matched features by eliminating matches inconsistent with the transform. In the case of rigid transforms, each two pairs of matches must maintain the same distance between features in their respective images. Further, traversing a triplet of matched features must maintain the same clock–wise, or

counter–clock–wise, direction in both images. To eliminate inconsistent matches, list of all possible triplets of matches is formed. This is feasible since the number of matches is relatively small. A maximum of **100** interest features are extracted in the current implementation and typically a maximum of 80 feature pairs are matched

Each triplet of matched features is examined to determine if it maintains constant distance and direction between features in the two images. Matches which participate in the largest number of inconsistent tnplets are eliminated first Matched feature elimination stops when all remaining triplets are consistent. Since features are extracted at exact pixel locations, distances between features are considered equal if they agree to within a single pixel. Again this is a greedy method which potentially eliminates more matches than needed.

3.3 Final Matches

After eliminating inconsistent features, it is highly unlikely that any of the remaining matches are incorrect, however, it is possible that some combinations of the remaining features **will** help recover a more accurate transform because of their arrangement in the images. To determine which combination gives the best transform) each combination is used to calculate a transform (see [13] for detailed explanation of recovering a transform using matched points) The transform is then applied to the second image. An error measure is calculated **as** the average absolute difference between the first image pixels and corresponding transformed second image pixels. The combination of matched points which produces the minimum error is the final pixel–level match used

3.3.1 Subpixel Accuracy

To achieve subpixel accuracy, the same method used for selecting the final pixel–level match can be used on matched features generated by displacing the final pixel–level match by fractions of a pixel. In this implementation, features are matched to within **0.25** of a pixel

4 Experimental Results

To measure the performance of recovering rigid transforms based on the proposed method of feature selection and matching, **14** pairs of mammogram images are used. Each pair consists of two digitized versions of the same film at different resolutions. The first digitized version is of a clean film at **42** microns per pixel (images are up to approximately **6,000** by **7,000** pixels in size). The second digitized version is at 210 microns per pixel (images are up to approximately **1,200** by **1,400** pixels in size) The smaller digitized version is of the film after being marked with a greased pencil by **an** expert radiologist The purpose of marking the film is establishing ground truth for regions of interest, such **as** suspicious or abnormal regions. The **42** micron images are reduced to **210** micron by averaging blocks of **5** by **5** pixels. The reduced version of the clean film will be referred to as the first image while the **210** micron marked version will be called the second image. Each image pixel is represented by **16** bits per pixel for a possible grey–scale range of 0 to **65,535.**

The rigid transform between the first and second image in each of the **14** pairs was recovered using the automated method described in this paper A similar semi–automated method recovers the rigid transform by **allowing** a human operator to select matching points across an image pair The operator has the ability to magnify a region before selecting a point within that region The method has a tolerance to error in manually selecting points within three pixels of their actual location. To compensate for this, searching for optimal matches is performed within **3** pixels around each selected point at increments of **0.25** pixels. **As** described earlier, our fully automated method **also** searches at increments of **0.25** pixels, however, it assumes accuracy in feature selection to within a pixel. Table **1** shows the transform recovered for each of the **14** pairs using the fully automated method and the semi–automated method. The error associated with the transform corresponds to the average absolute difference between each pixel in the first image and its corresponding pixel in the transformed second image. The transformed second image pixels are calculated using bilinear interpolation of original image pixels. It is clear from the error measures for all transforms that the automated method is capable of recovering a more accurate transform

Although the image pairs contain significant translation differences, they do not seem to contain a high degree of rotation. All image pairs seem to have rotation differences of less than one degree about their origin. To further test our method, we rotated each original **42** micron image by **5** degrees about its origin The rotated images were then reduced to **210** micron again by averaging. The transform between the rotated images and the marked **210** micron images was calculated using the automated method. Table **2** shows the recovered transform and error associated with each image pair The technique was able to recover the rotation introduced, combined with the original transform between the images, in all **14** pairs tested

Figure **2** shows an example of a pair of mammograms used The top image is the clean film digitized at **42** micron, then rotated by **5** degrees and reduced to a **420** micron resolution. The bottom image is the marked film digitized at **420** micron Since image pixels are **16** bits, they were histogram equalized into a range of 0 to **255** for display purposes. All processing was performed on the full **16** bits The marked white points are the final pixel level matches found by the algorithm described. The recovered rigid transform is applied to the bottom image then the difference between the top image and transformed image was calculated. The difference image is shown in Figure **3.** To display the difference image, the center 98% difference values around the average difference were scaled to values between 0 and **255.** More positive differences appear brighter, while more negative differences appear darker The radiologist markings become very clear with the subtraction of images. The difference images for the remaining **13** pairs show similar properties.

The registration method is currently being used in construction of the Digital Database for Screening Mammography [4] to transcribe radiologist markings. The registration technique has been applied to **71** mammogram pairs. Based on subjective evaluation of the difference images generated, the fully automated registration method gave accurate results for **68** (95for the remaining three pairs did not show the radiologist markings clearly.

5 Conclusion

We have introduced an image operator for extracting translation and rotation invariant features for matching across corresponding images. The feature selector was applied in a highly robust algorithm for recovering rigid transforms in

Table 1 Difference in performance for recovering rigid transforms **using** the fully automated method, and the method based on manual feature selection. The displacement in the x and y directions is given in pixels and the angle is given in degrees.

Image Pair	A	:omate	Metho	
		y disp.		error
1	0.50	28.75	0.00	46.2
2	7.92	20.68	0.79	55.3
3	-8.28	32.62	-0.67	109.8
4	-3.27	25.31	-0.13	67.8
5	12.57	16.78	1.07	43.1
6	26.24	18.91	0.78	43.0
7	-2.05	32.50	-0.21	30.0
8	-2.40	31.28	-0.06	26.85
9	0.50	26.50	0.00	77 90
10	-7 12	30.21	-0.38	139.0
11	0.50	26.50	0.00	56.67
12	3.39	15.87	0.61	176.9
13	0.50	27.25	0.00	89.39
14	-4.37	23.76	-0.03	40.81

Image Pair	Manual		ed Fea	
	x disp.	y disp.	angle	error
1	116	29.51	-0.02	163.0
2	7.61	20.79	0.76	146.0
3	-5.99	35.99	-0 78	462.7
4	-4.20	26.42	-0.16	222.3
5	12.72	16.35	1.00	129.0
6	26.24	19.06	0.73	106.3
7	0.36	35.17	-0.24	312.6
8	-0.41	31 71	-0.01	110.8
9	0.96	26.31	0.00	150.4
10	-8.20	27 90	-0.29	487 4
11	1.66	28.84	-0.10	271 7
12	1.89	15.01	0.58	434.4
13	-0.63	30.18	-0.23	181.8
14	-2.83	24.46	-0.01	184.6

Table 2: Performance of the automated method on image pairs with added rotation. The displacement in the x and y directions is given in pixels and the angle is given in degrees.

Image Pair				error
1				36.0
2	7.89	20.44	5.77	42.3
3	-9.34	34.55	4.22	48.9
4	-2.95	25.32	4.88	58.2
5	12.38	16.72	6.05	29.3
6	26.24	18.86	5.78	21.0
7	-2.14	32.43	4 75	20.5
8	-2.64	31.33	4.96	23.9
9	-0.21	26.27	4.97	71.1
10	-7 34	30.38	4.58	145.2
11	-1 28	26.81	4.82	68.3
12	2.95	16.25	5.52	186.5
13	0.46	26.69	5.01	57.0
14	-3.98	23.98	4.96	32.67

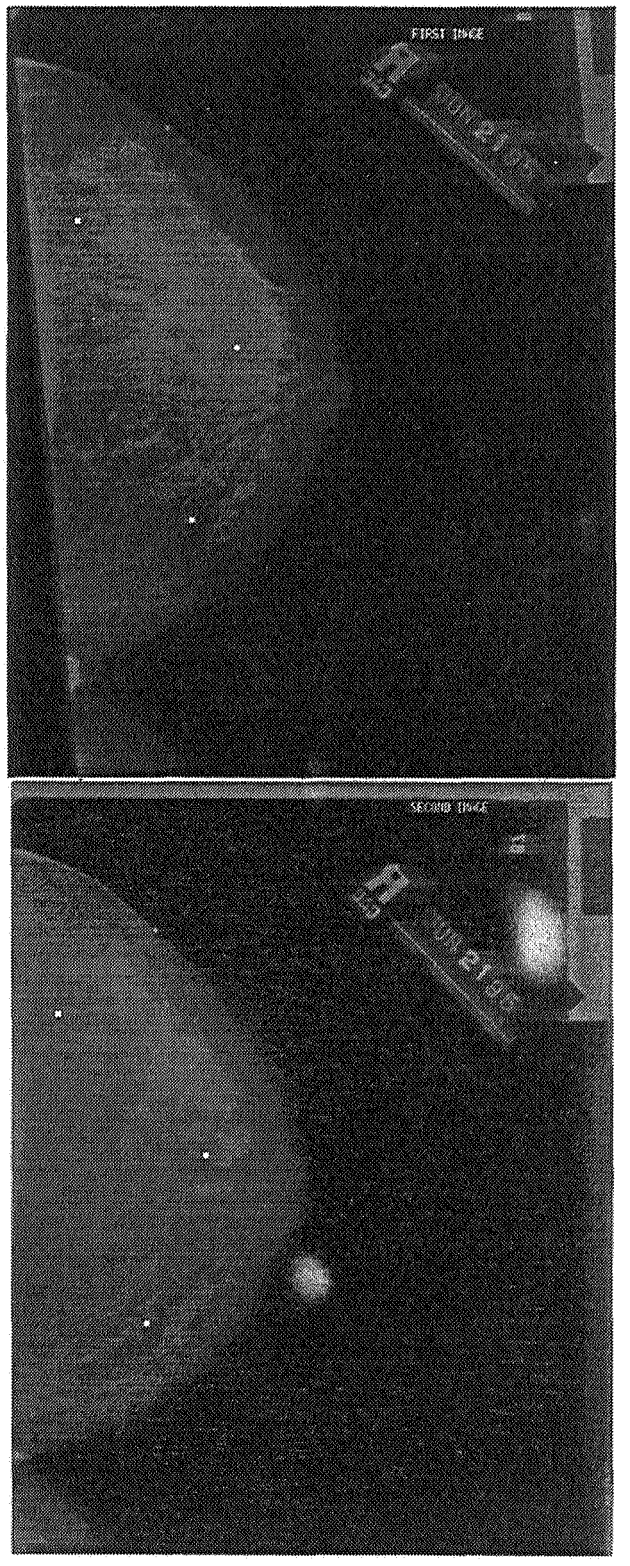


Figure 2: Features axe consistently extracted and correctly matched in textured regions in spite **of** the high degree of rotation between the *images.*

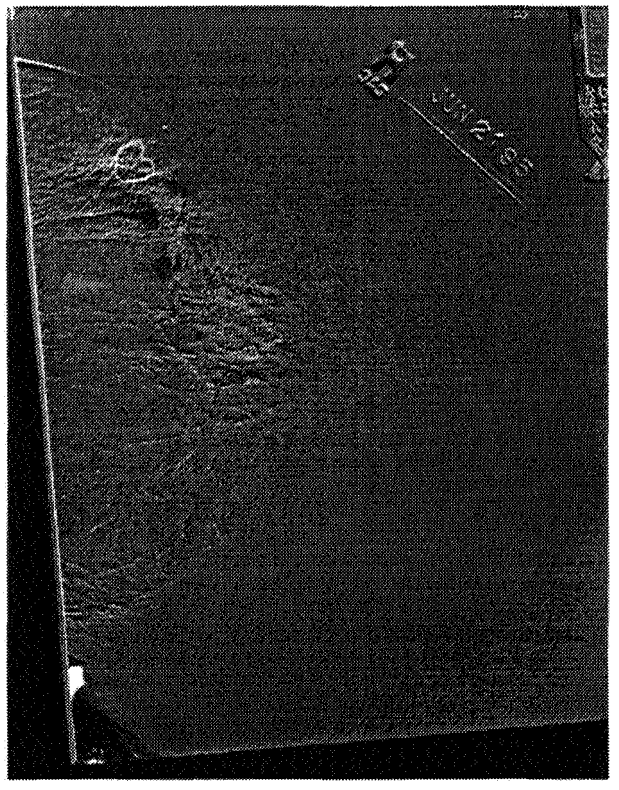


Figure 3: The difference between the transformed second image and the first image. The radiologist markings become very clear in the difference image which is in the same coordinate system as the unmarked image.

images which contain textured regions. The operator does not depend on image content which makes it suitable for different types of images.

The feature selector is invariant to local changes in rotation and not only to a single global rotation transform. This means that it can be applied to images which have undergone more complicated transforms reflected by non–linear changes in local orientation between an image pair

Acknowledgement

Support for this research was provided by U.S. Army Medical Research and Development Command, DAMD17-94-J-4015.

References

[1] *Proceedings* of *The Second International Workshop on Digital Mammography,* July 1994.

[2] *Proceedings* of *The Third Internatronal Workshop on Dagital Mammography,* June 1996.

[3] D Barnea and H. Silverman "A class of algorithms for fast digital image registration" *IEEE Transactions on Computers,* C–21:179–186, 1972.

[4] K. Bowyer, D. Kopans, W P Kegelmeyer, R. Moore, M Sallam, K Chang, and K Woods. "The Digital Database for Screening Mammography" In *Third International Workshop on Digital Mammography,* pages 431–434, June 1996

[5] K W Bowyer and S. Astley. *State of the Art in Digital Mammographic Image Analyszs.* World Scientific, 1994

[6] L Brown. "A survey of image registration techniques" *A CM Computing Surveys,* 24(4):225–376, Dec 1992.

[7] R. Duda and P Hart. *Pattern Classification and Scene Analysis.* John Wiley & Sons, 1973.

[8] W Freeman and E Adelson The design and use of steerable filters. *"IEEE Transactions on Pattern Analysis and Machrne Intelligence",* 13(9):891–906, Sep. 1991

[9] A. Goshtasby and G. Stockman. "A region based approach to digital image registration with subpixel accuracy " *IEEE Trasactions* on *Geosetence and Remote Sensing,* GE–24:390–399, 1986.

[10] H.P Moravec. "Rover visual obstacle avoidance" In *Proceedings of International Joint Conference on Artificral Intelligence,* pages 785–790, 1981

[11] G Stockman, S Kopstein, and S. Benett "Matching images to models for registration and object detection via clustering" *IEEE Transactions on Pattern Analysas and Machane Intellrgence,* 4229–241, 1982.

[12] M. Unser, B Trus, and M Eden. "Iterative restoration of noisy elastically distorted periodic images" *Signal Processing,* 17(3):191–200, July 1989.

[13] G. Wolberg. *Digital Image Warping.* IEEE Computer Society Press, Los Alamitos, CA, 1990.

Surface Based Matching using Elastic Transformations

Maria Gabrani and Oleh J. Tretiak
Electrical & Computer Engineering Department,
Drexel University, Philadelphia, PA 19104,
E-mail: maria@cbis.ece.drexel.edu, tretiak@coe.drexel.edu

Abstract

In this paper we present a methodology of nonlinear alignment of multidimensional data sets that is based on Elastic Transformations. The proposed approach is designed for surface based matching. Results on the existence and uniqueness of the solution are presented, and experimental results on two- and three-dimensional alignment of brain data used in neurochemistry research are shown.

1 Introduction

Studies that require the collation of volumetric medical images from different individuals call for the use of nonlinear transformations. We propose a methodology for nonlinear alignment of d-dimensional objects on the basis of geometric features such as surfaces. Gray-scale correspondence is not a reliable criterion for alignment, particularly in multimodality imaging. Object boundaries, on the other hand, are independent on imaging modality.

The proposed formulation follows the elastic deformation methodology introduced by Bajcsy et al. [4], [5], and further developed by Miller and co-workers [6]. The energy functional of the proposed transformations has a general form that includes previously used energy functionals; thin-plate bending energy of Bookstein [7], combination of thin-plate and membrane energy of Kass et. al [8], strain energy of Miller et. al [6]. Unlike Bajcsy and Miller, the solution to the problem has the form of a multivariate interpolation formula. The use of interpolation theory in this field was pioneered by Bookstein [7]. One of the major differences of the proposed work with previous methods ([1], [2], [7]) is the form of the constraints.

Elastic transformations are valid for any dimensionality d of the problem and any order of energy m, as long as the energy operator satisfies some specific constraints ([11]). One of the major advantages of this work is that overcomes the "curse of dimensionality" The order of the transformations doesn't necessarily increase with the dimensionality of the application Problems such as singularities at zero and slow decay and ripples at infinity are absent in surface based matching. Furthermore, elastic transformations provide stability. Thus, we are able to pick different order basis functions according to desired properties; fact that provides a flexibility in the design of the application.

This paper is organized the following way. In Section 2 we present the mathematical definition of the problem. The solution and the conditions of validity and existence are provided in Section 3. In Section 4 we elaborate the procedure of designing interpolating functions. Sections 5 and 6 are devoted to the stability of the algorithm and the point correspondence evaluation procedure, respectively. In Section 7 we demonstrate the performance of the proposed transformations in 2D and 3D mapping problems. Finally, we conclude our presentation with discussion in Section 8.

2 Problem's Mathematical Definition

In an alignment task we are given two "images"

$$f_1(\mathbf{x}), f_2(\mathbf{x}), \quad \mathbf{x} \in \mathcal{R}^d \tag{1}$$

We are required to find a space *transformation* $T(\mathbf{x})$ to compute $f_1'(\mathbf{x}) = f_1(T(\mathbf{x}))$ in which homologous points in f_1 and f_2 are aligned This is done identifying homologous point sets (surfaces) S_1, $S_2 \subset \mathcal{R}^d$ in the corresponding images, and finding a transformation such that $T(S_1) \subset S_2$ Our formulation circumvents the need to establish point correspondence between the spaces. This general approach was used to find rigid body (Kozinska et. al [9]) and affine (Ozturk [10]) transformations for multidimensional alignment.

The formulation is based on multidimensional spline theory as developed by Duchon [1] and further developed by Meinguet [2]. We seek to find a function $\mathbf{T} : \mathcal{R}^d \to \mathcal{R}^d$ such that $T(a_i) = a_i + u_i = a_i, a_i \in S_1, a_i \in S_2, i = 1, \ldots, N$, with

$$T(x) = L_y[\Phi](x) + x, \quad \forall x \in f_1 \tag{2}$$

and with $\Phi(x)$ minimizing

$$\int w(\rho)|\hat{\Phi}(\xi)|^2 d\xi \tag{3}$$

In the above formulas $\rho = |\xi|$ denotes frequency, $\hat{\Phi}(\xi)$ is the Fourier transform of $\Phi(x)$, w $\mathcal{R}^d \to \mathcal{R}, w(\rho) > 0$ for $\rho > 0$, the limit of $w(\rho)/\rho^{2m}$ for some positive integer m exists and is positive as $\rho \to 0^+$, and $w(\rho) = O(\rho^{2k})$ for large ρ. In (2) $L_y[\Phi](x) = c\frac{\partial \Phi}{\partial y}|_{y=x}$

3 Interpolation Solution

Many authors have worked towards the solution of constrained minimization problems with the energy functional similar to our form ([1] - [3]). The difference of these works with our problem is mainly on the form of the closeness constraints. We were inspired by these works and we followed a similar analysis to obtain the closed form interpolation solution to our registration problem. We found that if $2k > m + d/2 + 1$ function T exists and has the form $T(x) = x + v(x)$ where $v(x)$ has the form

$$v(x) = \sum_{i=1}^{N} H(x - a_i) \; \gamma_i + \sum_{k=1}^{M} d_k L_x[p_k(x)](x) \qquad \forall x \in E_1 \tag{4}$$

Here $H(x)$ are the tensor-valued functions defined as

$$H_{qn}(x - a_i) = L_x[L_y^*[\mathcal{E}(y - x)](a_i)_q](x)_n \tag{5}$$

with $\gamma_i \in \mathcal{R}^d$, $H \; \gamma$ denotes the (tensor) inner product, and $d_1, \ldots, d_M$ are scalars. The coefficients γ_i and d_k are the solutions of the simultaneous equations,

$$v(a_i) = u_i, \quad a = 1, \ldots, N \tag{6}$$

$$\sum_{i=1}^{N} L_x[p_k(x)](a_i) \; \gamma_i = 0, \quad k = 1, \ldots, M \tag{7}$$

In the above formulas, E is given by

$$\hat{\mathcal{E}}(\rho) = \frac{1}{w(\rho)} \tag{8}$$

with $w(\rho)$ defined as in (3), L^* is the adjoint of L, and $p_k \; \mathcal{R}^d \to \mathcal{R}$, $k = 1, \ldots, M$ is a set of linearly independent polynomials in the variables $(x_1, \ldots, x_d)$, of total degree less or equal to $m - 1$, and there are M of them, with $M = \binom{d+m-1}{d}$. Function T is unique whenever the point set a_i, $a = 1, \ldots, N$ is unisolvent for the polynomials.
A necessary and sufficient condition for the energy operator $w(\rho)$ to give a valid interpolation solution in X is that for $\rho \to \infty$

$$\int \rho^{2m+2} w^{-2}(\rho) d\xi \tag{9}$$

be integrable ([11]). If the order of $w(\rho)$ at infinity is $2k$ then the condition (9) becomes $2k > m + d/2 + 1$. If $k = m$ then the above condition is modified to $m > d/2 + 1$ that is the case of study for Duchon, Meinguet, Wahba, and Bookstein. However, it is not necessary that $k = m$. Any k that satisfies the above condition at infinity provides valid solution in X. However, the order of $w(\rho)$ at zero is fixed to m, depending strictly on the order of polynomials $\mathcal{P}$. Hence, the order of the operator $w(\rho)$ may not be necessarily the same at infinity and at zero. Fact that allows some flexibility in the design of the interpolating functions.

4 Designing Interpolating Functions

From (2) and (9) we see that the Elastic Transformation is specified by a choice of either $\mathcal{E}(r)$ or $w(\rho)$. One solution can be constructed by choosing

$$\mathcal{E}(r) = \begin{cases} f_1(r), & |r| \geq 1 \\ f_2(r), & |r| < 1 \end{cases} \tag{10}$$

with $\Delta^{m_1} f_1 = 6$. Thus, $f_1 = r^{n_1} \ln r$ for d even and $f_1 = r^{n_1}$ for d odd, where $n_1 = 2m_1 - d$. Function f_2 for d even is defined by $(ar^{n_2} + br^{n_3}) \ln r$. For d odd it is given by $f_2 = r^{n_2}(a + br + cr^2)$, $n_2 \geq 2$. The coefficients a, b, and c are chosen so that $\mathcal{E}(r)$ and its first two derivatives are continuous at $r = 1$. This construction leads to an $\hat{\mathcal{E}}(\rho) \geq 0$, $O(\hat{\mathcal{E}}(\rho)) = -2m_2$ as $\rho \to \infty$, and $O(\hat{\mathcal{E}}(\rho)) = -2m_1$ as $\rho \to 0$, where $2m_2 = n_2 + d$. From (9), since $w(\rho) = 1/\hat{\mathcal{E}}(\rho)$, we obtain

$$O(w(\rho)) = \begin{cases} 2m_1, & \rho \to 0 \\ 2m_2, & \rho \to \infty \end{cases} \tag{11}$$

The above corollary is a mathematical form of the discussion that we give in the previous paragraph about the behavior of the energy operator at zero for $m_1 = m$ and at infinity for $m_2 = k$. Given a function of the form (10) we can prove that we can use this function in our registration problem and obtain a valid interpolation solution ([11]).

We can also tackle the issue the opposite way. We can consider an operator that satisfies the necessary and sufficient condition (9) and is not an iterated Laplacian. Such an operator in Fourier domain is the $w(\rho) = \rho^4 + \varepsilon\rho^6$, with $\rho = |\xi|$, and ε a small constant. At infinity $w(\rho) \simeq \rho^6$ and at the neighborhood of the origin $w(\rho) \simeq \rho^4$. That is, for small r, where the second order iterated Laplacian ρ^4 leads to a solution with infinite energy, the order of the operator increases, compensating for the singularity. For large r the second order operator leads to smooth solution. This solution is the one corresponding to elasticity theory. The interpolating function is then given by $\hat{\mathcal{E}} = 1/(\rho^4 + \varepsilon\rho^6)$. By partial fraction expansion the Fourier transform of the interpolating function takes the form

$$\frac{1}{\rho^4 + \varepsilon\rho^6} = -\frac{\varepsilon}{\rho^2} + \frac{1}{\rho^4} + \frac{\varepsilon}{\rho^2 + 1/\varepsilon} \tag{12}$$

or, in symbolic form $\hat{\mathcal{E}} = c_1\hat{g}_1 + c_2\hat{g}_2 + c_3\hat{g}_3$, respectively. Taking the inverse Fourier transform we find that, $g_1 = \ln r$, $g_2 = r^2 \ln r$ for d even, and r^2 for d odd, and finally, $g_3 = \varepsilon^{1-\nu/2} K_\nu(\varepsilon^{1/2} r) r^{-\nu}$, for $d < 5$, and $r > 0$. Here K_ν is Bessel function of the third type of order ν, with $\nu = d/2 - 1$. Thus, for example, for d = 2 we obtain $E = c_1 \ln r + c_2 r^2 \ln r + c_3 K_0(\varepsilon^{1/2} r)$. At infinity K_0 is zero thus the interpolating function is the one of order $m = 2$; $E = r^2 \ln r$. In the neighborhood of zero, the singularities of the g_1 and g_2 cancel out with the Bessel function and the total function acts as a third order interpolating function.

A major outcome from the above discussion is that we can find operators that do not follow the order of the dimensionality d. In other words the order m does not necessarily have to follow the dimensionality. Thus, we can keep m low as long as we satisfy (9) for all ρ. However, as we have already mentioned, our aim is to solve an interpolation problem. From the interpolation theory point of view, taking higher order energy functionals we introduce smoother basis functions. The smoother the basis functions the better the interpolation approximation. An advantage of the proposed transformations with respect to the thin-plate splines is that for the same order of polynomials the proposed transformations have smoother basis functions.

5 Stability

Formulas (4) - (7) look quite complicated. However, as we show in [11] the registration problem may be reduced

to a problem in linear algebra of the form $Az = y$. A major problem of the application of Thin-Plate splines, as well as Elastic transformations is that their implementation may lead to ill-conditioned systems. The main reason for that is the form of the interpolating functions $\mathcal{E}$. The interpolating functions are radial basis functions that cause the system matrix diagonal elements to be zero and the off-diagonal elements to increase with r Below we present two new techniques to deal with the problem.

In the first technique, the system of equations (7) used to construct the transformation is modified by expanding the polynomial basis. Operationally, when the coefficients used to construct the transformation $v(r)$ satisfy (9) then the rate of growth of $v(r)$ at large r is reduced, The set of polynomials of total degree less or equal to $m - d/2 - 1$ assures a growth for large r sufficiently small so that the integrand in (9) is integrable for small ρ ([12]). If we augment system (7) by adding polynomials of higher degree then the transformation rate of growth is decreased, and with polynomials of high enough degree we can assure that $v(r)$ goes to zero for large r This causes the system of equations (6), (7) to have better condition number By doing this we are, in effect, using a different set of basis functions, ones that have a more rapid decay, leading to a smaller interaction among the interpolating points. The introduction of higher order polynomials leads to a non-square system of equations, but this problem can be solved effectively by using efficient method for approximate solution, such as QR decomposition

A second technique is concentrated in the form of the basis functions. The diagonal elements in matrix A are equal to zero, which is a major cause of ill-conditioning. An efficient way to overcome the problem is to scale the elements of the matrix in such a way that the difference between them is decreased. Another way to deal with the problem is to introduce a new basis. We propose a new basis that is a scaled sum of translated versions of the original basis ([12]) Thus if we impose the correct restrictions the new basis will be smoother than the original one. Furthermore each point will be encountered more times and thus scaled in such a way that the matrix A will have smaller differences between its elements. A way to develop this new basis is described in details in [12]. Using this technique the condition number of matrix A is reduced by 3 orders of magnitude.

6 Point Correspondence

The formulas developed in Section 3 assume that we know the correspondence between points a_i in their locations α_i in the transformed image. Our algorithm takes input boundaries of objects, chooses the set of points a_i from the boundary in one image, and searches for the points α_i in the other image that correspond to them by using the "principle of minimum energy", of the theory of continuum mechanics. According to the principle the energy of the deformation of the object in equilibrium state is minimum. Thus the configuration of points that corresponds to the homolog pairs of points gives the minimum energy. None other configuration will give less energy Lets consider that we want to find N control points. We initially pick N points in both surfaces. We distribute the points in such a way that we split the outer surface in equally spaced neighborhoods. The initial configuration is random and thus does not give us the minimum energy. Considering that the rotation of the deformed object is not very big we assume that the correct position of each point is in its neighborhood. We keep the position of the N points in the original image fixed and perturb each point of the target image on the surface in all possible directions. For each configuration of points we estimate its energy gradient. The set of points that will give us zero gradient energy is the one that we consider as control points. This procedure is independent on the order m of the energy and gives us satisfactory results. The process is rapid since all calculations are in matrix format. For further improvement of the running time one can check the sign of the gradient in all directions of each point and permit further perturbation in the direction of reaching zero. Moreover one can stop the algorithm when the gradient reaches zero.

In the above procedure the initial configuration is very important. The size of the neighborhood is also critical. One can vary the size according to the problem and to the initial configuration The neighborhoods may also overlap. We never faced the problem of the swapping of two points. Another more efficient approach is to initially find the areas of the surface with high curvature. Then sample the points in such a way that in areas of higher curvature the point distribution is more dense and in areas of smaller curvature the points are fewer, This way the correspondence is more certified and the neighborhoods remain small. Moreover from the interpolation point of view there are more basis functions in areas of high variation without overloading the system.

In the case of 2D mapping the neighborhoods are intervals of the outline. In 3D mapping the neighborhoods are parts of the outer surface consisting of intervals of outlines of adjacent sections. More detailed presentation is given and illustrated in [13]

7 Experimental Results

In this paper we are presenting experimental results obtained mapping rat brains by matching their outer surfaces. According to Section 3, the solution to this problem is an interpolating formula which is a sum of two parts. The first part consists of a weighted sum of the second derivatives of the Green's function $\mathcal{E}$ centered at the control points a_i, a_j with $i = 1, \ldots, N$ and j fixed. The second part consists of the derivates of the polynomials of total degree less or equal to $m - 1$. We present results on both two- and three-dimensional alignment. In the two-dimensional case we are considering the following cases; $m = 2$, $m = 3$ and $m = 4$. For the case $m = 2$ we use

$$\mathcal{E}(r) = \begin{cases} \frac{1}{8\pi} r^2 \ln r, & |r| \geq 1 \\ \frac{1}{64\pi} r^4 \ln r - \frac{1}{4608\pi} r^6 \ln r, & |r| < 1 \end{cases} \quad (13)$$

where r is the distance between two points (x_1, x_2) and (y_1, y_2). For $m = 3$ and $m = 4$ the function $\mathcal{E}$ is given by,

$$\mathcal{E}(r) = \begin{cases} -\frac{1}{128\pi} r^4 \ln r, & \text{for } m = 3 \\ \frac{1}{4608\pi} r^6 \ln r, & \text{for } m = 4 \end{cases} \quad (14)$$

In the three dimensional case we consider $m = 3$ and the interpolating function

$$\mathcal{E}(r) = \begin{cases} \frac{-\Gamma(-1.5)}{2^7 \pi^{1.5}} r^3, & |r| \geq 1 \\ \frac{\Gamma(-2.5)}{2^9 \pi^{1.5}} r^5 + \frac{\Gamma(-3.5)}{2^{13} \pi^{1.5}} r^7 + \frac{\Gamma(-4.5)}{2^{15} 15 \pi^{1.5}} r^9, & |r| < 1 \end{cases} \quad (15)$$

The rat brain images are images of rat data used in neurochemistry research, obtained in our lab.

7.1 2D Rat Brain Mapping

In the first experiment we deal with 2D data. The objects are horizontal histological sections of two different rat

brains. The size of the images is 240 x 320 square pixels.

Fig. 1(a), shows the image that we consider as the original one and the image, whose deformation we want to find, superimposed. We see that there is an affine transformation (rotation), as well as a nonaffine one (local variabilities) which corresponds to normal variation between individuals and to differences due to the acquisition methods. In this case, we have no information about corresponding points, we only know that the two outlines correspond Therefore, we make use of the algorithm that is described in section 4. We apply the proposed algorithm for $m = 2$, $m = 3$ and $m = 4$, and for $N = 8$ control points.

The deformations that we find using this algorithm are applied to the original image, and the outcomes superimposed on the data are shown in Fig. 1(b), (c) and (d), for $m = 2$, $m = 3$ and $m = 4$, respectively. We see that in all cases the results are very satisfactory and approximately the same. The data outlines have been deformed globally as well as locally to match the atlas. Significant difference between the various energy orders is not visible.

The difference is expected in the inside of the images where we presume that the higher orders will better predict the inner structures. We thus apply the transformations found previously to the whole image. In Fig. 2 we demonstrate the deformation of the atlas by presenting the grayscale difference of the deformed images and the data. For comparison reasons, in Fig. 2 (a) we show the grayscaled difference of the original image and the data. In Fig. 2 (b), (c), and (d) we show the difference of data and the transformed original image using energy order $m = 2$, $m = 3$, and $m = 4$, respectively. We see that using only 8 points of the outline we are able to satisfactory predict the location of inner structures of the brain Observe also that with the increase of m the prediction becomes better This is due to the smoothness of the Green's functions.

7.2 3D Rat Brain Mapping

In this paper the experimental results in the case of 3D mapping are restricted in the problem of known point correspondence. The original object is a set of images of 50 coronal histological sections of a rat brain. The distance between sections is $10\mu m$, and the thickness of each section is approximately $42\mu m$. The data set that we exploit in the present experiment is the points of the outer surface. The number of points on the surface is 87892. In Fig.3 we show the whole data set (outer surface of the brain) highlighting the portion used in the experiment. As target object we consider the second order polynomial transformation of the original brain. The original and the target brains are illustrated in Fig. 4, left and right, respectively In these figures we have, on purpose, emphasized the z-coordinates for better visualization The two data sets superimposed are shown in the left side of Fig. 5 where the gray corresponds to the target image and the darker surface corresponds to the original image. For better visualization, on the right of Fig. 5 we depict the projection in a plane of the target (gray) superimposed on the original (black) brain,

For the mapping we consider 10 control points randomly selected from the outer surface of the original image. The point set is illustrated in Fig. 6, left with white dots. The points are distributed in three-dimensions but for better visualization we depict their projection to a plane. The corresponding points in the target image are illustrated with black dots in Fig. 6 right. The outcome of the 3D elastic transformation algorithm is presented in Fig. 7 In Fig. 7, left we show the projection into a plane of the outcome. In the same image we show the new position of the transformed initial set of points. Finally, in Fig. 7, right we show the outcome and the target image superimposed. The alignment is obvious, and we see that the outcome is very satisfactory. The differences between the two objects are minimal. With 10 points out of 87892 we are able to obtain very good results considering second order transformations and energy order $m = 3$.

8 Discussion and Conclusions

In this paper we are presenting a new methodology for nonlinear alignment of d-dimensional objects on the basis of geometric features such as surfaces, The formulation follows the elastic deformation methodology introduced earlier generating an energy functional of a more general form. Conditions on the existence and uniqueness of the solution are described One of the main advantages of the proposed transformations is the introduction and validation of the idea of using more generalized differential operators than the iterated Laplacian. This permits the use of lower order energy functionals in space of high dimensions. Techniques that compensate singularities at zero and slow decay and ripples at infinity are described.

Our formulation automatically establishes point correspondence between the spaces being aligned. An important issue in implementing the proposed methodology is how to choose the initial alignment point set. This matter certainly deserves more study

Elastic transformations are used in the area of brain mapping. Several original results in two- and three- dimensional alignment are presented to demonstrate their performance and validate their application in the area of image matching. Further improvement of our algorithm is the aim of our future work. Evaluation of this work is in progress.

9 Acknowledgments

This work is supported by NIH Grant P41-RR01638.

References

[1] Duchon, J., "Jnterpolation des fonctions de deux variables suivant le principle de la flexion des plaques minces," *RAZRO analyse Numérique*, vol. 10, pp. 512,1976.

[2] Meinguet, J., "Multivariate Interpolation at Arbitrary Points Made simple," *Journal of Applied Mathematics (ZAMP)*, vol. 30, 1979.

[3] Wahba, G.; and Wendelberger, J., "Some new mathematical methods for variational objective analysis using splines and cross validation," *Monthly Weather Review*, vol. 108, pp. 1122-1143, April 1980.

[4] Broit, C., "Optimal registration of deformed images," doctoral dissertation, Univ. Pennsylvania, 1981.

[5] Bajcsy, R.; and Kovačič, S. "Multiresolution Elastic Matching." Computer Vision, Graphics, and Image Processing, vol. 46, pp. 1-21, 1989.

[6] Miller, M.; Christensen, G.; Amit, Y.; and Grenander, U., "Mathematical textbook of deformable neuroanatomies," *Proc. Natl. Acad. Scr. USA*, vol. 90, pp. 1194411948, December 1993.

[7] Bookstein, F.L., "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations," *ZEEE trans. on PAMZ*, vol. 11, no. 6, June 1989.

[8] Kass, M.; Witkin, A.; and Terzopoulos, D., "Snakes: Active contourmodels," *Int. J. Comput. Vision,* vol. 1, pp. 321-331, 1987

[9] Kozinska, D; Tretiak, O.J.; Nissanov, J; and Ozturk, C., "Multidimensional Alignment Using the Euclidean Distance Transform," submitted to Graphical Models and Image Processing

[10] Ozturk, C, "Alignment Package for Matlab", at http://cbis.ece.drexel.edu/ICVC/Align/align.html.

[11] Gabrani, M.; and Tretiak, O.J., "Elastic Transformations," in preparation.

[12] Gabrani, M.; and Tretiak, O.J., "Algorithmic Aspects of Elastic Transformations," in preparation.

[13] Gabrani M., and Tretiak, O.J., " Surface based mapping using Elastic Transformations", in preparation.

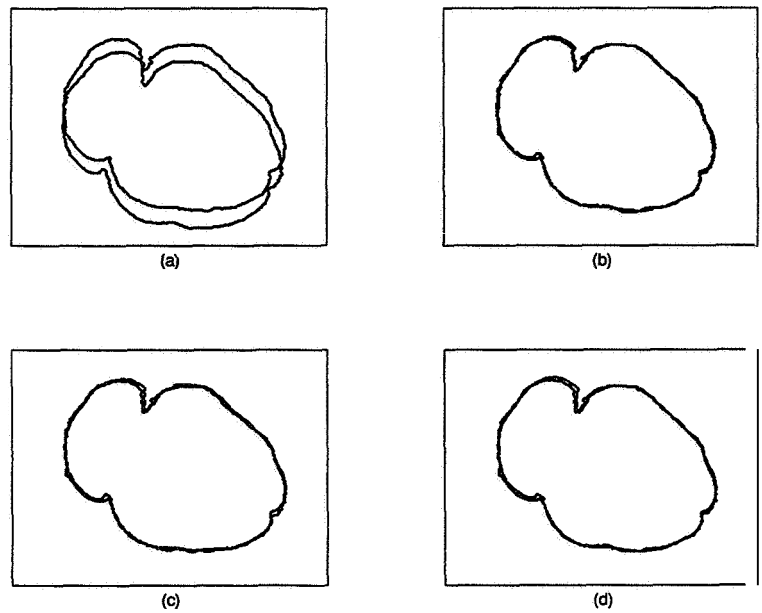


Figure 1 Demonstration of the elastic transformations performance in 2D mapping and for **various** values of energy orders. For comparison reasons the data is superimposed to the (a) atlas, (b) outcome for $m = 2$, (c) outcome for $m = 3$, and (b) outcome for $m = 4$ (gray).

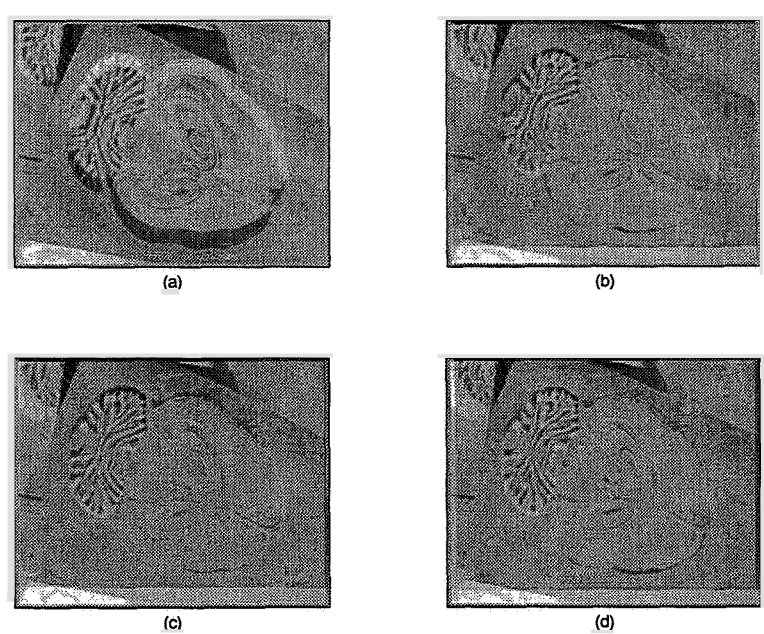


Figure 2: Demonstration of the transformation of the whole 2D image for various values of energy orders. The grayscaled difference of the data and the (a) original, (b) outcome for $m = 2$, (c) outcome for $m = 3$, and (b) outcome for $m = 4$.

Figure 3: Three different views of the rat brain and its part under study (darker). The images are in the form of a **3D** plot of the points of the outer surface of the brain.

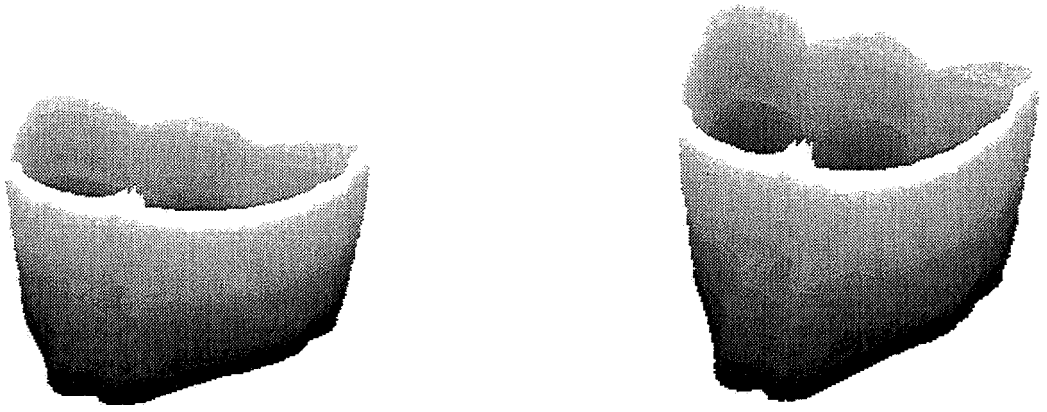


Figure **4:** Left. the object considered to be the original one; ***50*** adjacent coronal sections of a rat brain Right; the object considered to be the target; second order polynomial transformation of the original. In these figures the z-axis is enhanced on purpose for better visual quality,

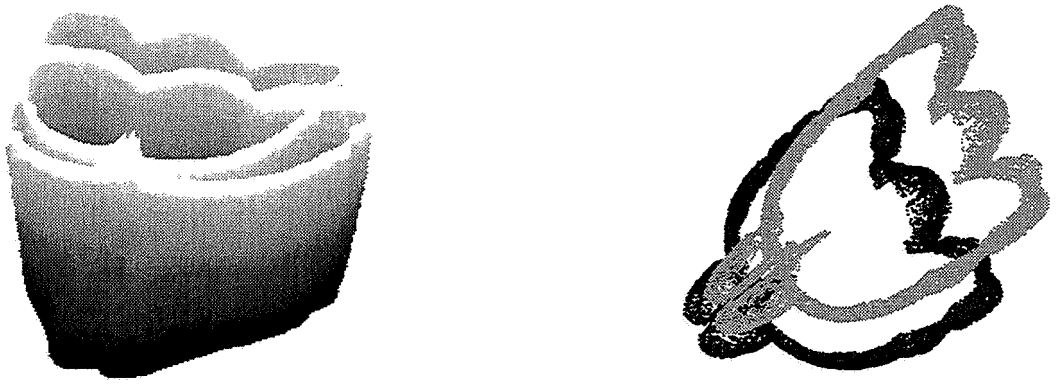


Figure 5: Left. The onginal object (darker) and the target one (gray) superimposed Right. for better demonstration of the differece of the objects we depict the projection in a **2D** plane of the target brain (gray) superimposed to the original (black);

Figure 6: For the transformation we use 10 control points. Left: the white dots are the control points of the original object. Right: the black dots are the control points of the target object.

Figure 7- Left: Projection of the deformed using elastic transformations original brain in a 2D plane; for the transformation 10 points of the outer surface were used. Right: for better demonstration of the performance of the proposed algorithm in the case of 3D mapping we depict the deformed original brain superimposed to the original. No difference between the two of them is visible.

Fast Multimodality Image Registration Using Multiresolution Gradient-based Maximization of Mutual Information

Frederik Maes*, Dirk Vandermeulen, Guy Marchal, Paul Suetens.
Laboratory for Medical Imaging Research!
Department of Electrical Engineering, ESAT† and
Department of Radiology, University Hospital Gasthuisberg§
Katholieke Universiteit Leuven, Belgium.

Address for correspondence:
Laboratory for Medical Imaging Research (ESAT/Radiologie)
Departement Radiologie, UZ Gasthuisberg
Herestraat 49, B-3000 Leuven, Belgium
Email: Frederik.Maes@uz.kuleuven.ac.be

Abstract

We investigate the robustness and performance of various multiresolution gradient-based optimization strategies for maximization of mutual information of image intensities for registration of CT and MR images of the brain. Our results demonstrate that high resolution images can be matched with subvoxel accuracy compared to the external marker-based reference solution in less than 5 minutes CPU time an a state-of-the-art workstation.

1 Introduction

Maximization of mutual information (MMI) has been proposed as a new approach to the problem of multimodality medical image registration [2] The method applies the concept of mutual information (MI) or relative entropy to measure the statistical dependence or information redundancy between the image intensities of corresponding voxels in both images, which is assumed to be maximal if the images are geometrically aligned Because no assumptions are made regarding the nature of this dependence and no limiting constraints are imposed on the image content of the modalities involved, MMI is a very general and powerful criterion, allowing robust, fully automated affine registration of multimodal images with different contrast and resolution without prior segmentation, feature extraction or other preprocessing steps.

In this paper, we investigate the performance of various gradient-based optimization strategies for maximization of mutual information and compare their efficiency with Powell's method used in [2]. The MI criterion varies smoothly as a function of the 12 affine registration parameters and the gradient of MI can be computed exactly, using analytic expressions for the partial derivatives of MI with respect to the registration parameters. Because evaluation of MI and of its gradient is proportional to the number of samples in the image, large speedups are possible using a multiresolution optimization approach, starting the optimization at a low resolution using a coarsely sampled image and continuing at a higher resolution by increasing the number of samples as the optimization proceeds.

Various multiresolution MMI strategies were evaluated for CT/MR matching of the brain on a set of 9 patient data sets, each consisting of CT, low resolution MR-PD, MR-T1, MR-T2 and high resolution MR-MPRAGE images. The accuracy of the MI criterion was evaluated by comparison with the external marker-based reference registration solution All optimization methods achieve subvoxel accuracy compared to the reference solution. Our results demonstrate that, compared to Powell's method at full resolution, speedups up to a factor 4 can be achieved without a loss of robustness by using the Levenberg-Marquardt method and a 2-level multiresolution scheme. Large datasets such as 256x256x128 MR-MPRAGE and 512x512x48 CT images can be registered in less than 5 minutes CPU time on a state-of-the-art workstation (SGI Octane, IRIX 6.4, R10000 195 MHz, 17 SPECfp95) Registration of the image volumes is therefore no longer a bottleneck in the analysis of multimodal images in clinical practice.

Other groups have experimented with the mutual information matching criterion and have presented different strategies to increase registration performance. Viola and Wells *et al.* [7] use stochastic sampling to estimate the joint image intensity histogram from a small number of samples. But this may introduce local optima and decrease registration robustness. Studholme *et al.* [6] use a multi-resolution technique to speed up an heuristic search procedure. They have reported results for CT, MR and PET matching that are very similar to ours [8]

Analytic expressions for the gradient of MI w.r.t. the affine registration parameters are derived in section 2 Section 3 gives an overview of various gradient-based optimization strategies for maximization of mutual information. Section 4 evaluates the robustness and performance of these methods for matching of CT and MR images of the brain

*Frederik Maes is Aspirant of the Belgian National Fund for Scientific Research (NFWO)
†Directors: André Oosterlinck & Albert L Baert
‡Kardinaal Mercierlaan 94, B-3001 Heverlee, Belgium
§Herestraat 49. B-3000 Leuven, Belgium.

Conclusions are presented in section 5.

2 The gradient of MI w.r.t. the affine transformation parameters

The mutual information I [1] of the images 1 and 2 to be registered, with image intensities $\{a\}$ and $\{b\}$ respectively, is computed from the joint image intensity histogram $\mathbf{H} = \{h_{ab}\}$ of their overlapping volume:

$$h_a = \sum_b h_{ab}\,, \quad h_b = \sum_a h_{ab}\,, \quad N = \sum_{ab} h_{ab}$$

$$p_{ab} = \frac{h_{ab}}{N}\,, \quad p_a = \sum_b p_{ab}\,, \quad p_b = \sum_a p_{ab}$$

$$I = \sum_{a,b} p_{ab} \log_2 \frac{p_{ab}}{p_a.p_b} = \sum_{a,b} \frac{h_{ab}}{N} \log_2 \frac{N.h_{ab}}{h_a.h_b}$$

The images are first linearly rescaled such that $1 \leq a \leq n_1 - 1$ and $1 \leq b \leq n_2 - 1$, with n_1 and n_2 the number of histogram entries for image 1 and 2 respectively. The joint histogram H is computed by binning the intensities of geometrically corresponding voxels in both images. $\mathbf{H}$ and I are thus function of the geometric registration parameters $\boldsymbol{\alpha} = \{\alpha_k\}$

If the derivatives $\frac{d\mathbf{H}}{d\alpha_k}$ of $\mathbf{H}$ w.r.t. α_k exist, the gradient of I with respect to $\boldsymbol{\alpha}$ can be expressed as

$$\nabla I(\boldsymbol{\alpha}) = \{\frac{dI}{d\alpha_k}\}$$

$$\frac{dI}{d\alpha_k} = \sum_{a,b} \frac{dI}{dh_{ab}} \frac{dh_{ab}}{d\alpha_k} = \sum_{a,b} \frac{1}{N}.(\log_2 \frac{p_{ab}}{p_a.p_b} - I).\frac{dh_{ab}}{d\alpha_k}$$

Each gradient component k is thus expressed as the sum over all histogram entries of the change in each entry when changing registration parameter α_k, weighted by the influence of this change on I

The derivative $\frac{d\mathbf{H}}{d\alpha_k} = \{\frac{dh_{ab}}{d\alpha_k}\}$ of the joint histogram $\mathbf{H}$ w.r.t. the affine registration parameter α_k is itself a histogram with the same size as $\mathbf{H}$. The dependence of $\mathbf{H}$ on the transformation parameters $\boldsymbol{\alpha}$ depends on the interpolation scheme and the binning strategy that is used to construct H. While both nearest neighbour (NN) and trilinear (TRI) interpolation might result in discontinuous changes in the histogram for small changes in the registration parameters, trilinear partial volume distribution (PV) interpolation assures that $\mathbf{H}$ varies continuously with $\boldsymbol{a}$, provided that the border of the region of overlap is treated specially. The histogram derivatives $\frac{d\mathbf{H}}{d\alpha_k}$ exist almost everywhere when PV interpolation is used (they can be infinite) and can be computed exactly using analytic expressions These expressions are derived below

2.1 The affine registration transformation

The affine geometric transformation of image coordinates $\boldsymbol{p}$ in image 1 into image coordinates $\boldsymbol{q}$ in image 2 is given by

$$\mathbf{q} = (\mathbf{T}_{i,w,2}^{-1} * \mathbf{A}(\boldsymbol{\alpha}) * \mathbf{T}_{i,w,1}).\mathbf{p}$$

$\mathbf{T}_{i,w,1}$ and $\mathbf{T}_{i,w,2}$ are 4 x 4 image-to-world coordinate transformation matrices that take the voxelsizes and orientation of each of the images into account. $\mathbf{A}(\boldsymbol{\alpha})$ is the f f i e transformation matrix determined by the affine transformation parameters cy

$$\mathbf{A}(\boldsymbol{\alpha}) = \boldsymbol{T} * \mathbf{R} * \boldsymbol{G} * \mathbf{S} \qquad (1)$$

with the 4 x 4 matrices $\boldsymbol{T}$, $\mathbf{R}$, $\boldsymbol{G}$ and S representiag 3-D translation, rotation, skew and scaling respectively.

The affine image-to-image transformation $\mathbf{T}_{12}$ that transforms image coordinates in image 1 into image coordinates in image 2 is thus given by

$$\mathbf{T}_{12}(\boldsymbol{\alpha}) = \mathbf{T}_{i,w,2}^{-1} * \mathbf{A}(\boldsymbol{\alpha}) * \mathbf{T}_{i,w,1}$$

The derivatives of $\mathbf{T}_{12}$ with respect to its parameters α_k are itself 4 x 4 matrices given by

$$\frac{\delta \mathbf{T}_{12}}{\delta \alpha_k} = \mathbf{T}_{i,w,2}^{-1} * \frac{\delta \mathbf{A}}{\delta \alpha_k} * \mathbf{T}_{i,w,1} \qquad (2)$$

Expressions for $\frac{\delta \mathbf{A}}{\delta \alpha_k}$ can be straightforwardly computed for each of the translation, rotation, skew and scale parameters from the expression 1 given above.

2.2 Computing H using PV interpolation

The joint histogram $\mathbf{H}$ of the images 1 and 2 with registration parameters $\boldsymbol{\alpha}$ is computed by transforming all voxels in image 1 into image 2 by the image-to-image coordinate transformation $\mathbf{T}_{12}(\boldsymbol{\alpha})$ and binning all pairs of corresponding voxel intensities after interpolation in image 2. For each sample, the PV interpolation method updates the joint histogram H at 8 entries according to the following scheme.

Let k be a voxel of image 1 at position $\mathbf{p}_k$ Let $\mathbf{q}_k = \mathbf{T}_{12}.\mathbf{p}_k$ be the transformed position of voxel k in image 2. Let $l_{k,m} = l_{k,\mathbf{n}_m} = l_{k,(\alpha\beta\gamma)_m}$, with $m = 4\alpha + 2\beta + \gamma + 1, \alpha, \beta, \gamma \; E \; \{0,1\}, m = 1,2,3,.,8$, denote the 8 nearest neighbour voxels of $\mathbf{q}_k$ in image 2, at positions $\mathbf{q}_{k,m} = \mathbf{q}_{k,\mathbf{n}_m}$ such that $\mathbf{q}_{k,\mathbf{n}_{m_1}} - \mathbf{q}_{k,\mathbf{n}_{m_2}} = \mathbf{n}_{m_1} - \mathbf{n}_{m_2}$ Let $\mathbf{r}_k$ be the distance of $\mathbf{q}_k$ to $\mathbf{q}_{k,1}$ $\mathbf{r}_k = \mathbf{q}_k - \mathbf{q}_{k,1} = (r_{kx}, r_{ky}, r_{kz})$ with $0 \leq r_{ks} \leq 1$ The trilinear partial volume distribution weights $w_{k,m}$ of $\mathbf{q}_k$ w.r.t. its neighbours $\mathbf{q}_{k,m}$ are then given by

$$\begin{aligned}
w_{k,1} &= w_{k,(000)} = (1 - r_{kx}).(1 - r_{ky}).(1 - r_{kz}) \\
w_{k,2} &= w_{k,(001)} = (1 - r_{kx}).(1 - r_{ky}).r_{kz} \\
w_{k,3} &= w_{k,(010)} = (1 - r_{kx}).r_{ky}.(1 - r_{kz}) \\
w_{k,4} &= w_{k,(011)} = (1 - r_{kx}).r_{ky}.r_{kz} \\
w_{k,5} &= w_{k,(100)} = r_{kx}.(1 - r_{ky}).(1 - r_{kz}) \\
w_{k,6} &= w_{k,(101)} = r_{kx}.(1 - r_{ky}).r_{kz} \\
w_{k,7} &= w_{k,(110)} = r_{kx}.r_{ky}.(1 - r_{kz}) \\
w_{k,8} &= w_{k,(111)} = r_{kx}.r_{ky}.r_{kz}
\end{aligned}$$

Let a_k be the intensity of image 1 at voxel k Let $b_{k,m}$ be the intensity of image 2 at voxel $l_{k,m}$ if $\mathbf{q}_{k,m}$ falls inside the volume of image 2 and 0 otherwise. For each sample k in image 1, $\mathbf{H}$ is updated at all 8 entries $h_{a_k b_{k,m}}$ with the weight $w_{k,m}$

$$h_{a_k b_{k,m}} \mathrel{+}= w_{k,m}\,, \quad m = 1,2,3,.,8$$

Each entry h_{ab} in the histogram $\boldsymbol{H}$ is thus build up as the sum over all samples k of all fractions $w_{k,m}$:

$$\mathbf{H}(a,b) = \sum_k \sum_{m=1}^{8} w_{k,m}.\delta(a - a_k, b - b_{k,m}) \qquad (3)$$

Entries of $\mathbf{H}(a,b)$ corresponding to $a = 0$ or $b = 0$ are not taken into account when computing I to exclude samples that fall outside the volume of overlap. Because all fractions $w_{k,m}$ vary smoothly with the transformed position $\mathbf{q}_k$, which varies itself smoothly with the transformation Parameters $\boldsymbol{a}$, the histogram $\mathbf{H}$ and hence the mutual information I are continuous functions of $\boldsymbol{a}$

2.3 Computing derivatives of H using PV interpolation

The derivatives of H w.r.t. the affine transformation parameters α_k can be computed as a linear combination of the derivatives of H w.r.t the elements $T_{12_{ij}}$ of the image-to-image transformation matrix itself:

$$\frac{d\mathbf{H}}{d\alpha_k} = \sum_{ij} \left(\frac{\delta \mathbf{T}_{12}}{\delta \alpha_k}\right)_{ij} \frac{d\mathbf{H}}{dT_{12_{ij}}}$$

Using expression 3 for $\boldsymbol{H}$, the derivatives of $\mathbf{H}$ w.r.t. $T_{12_{ij}}$ can be computed from the derivatives of the weights $w_{k,m}$ w.r.t. $T_{12_{ij}}$

$$\frac{d\mathbf{H}}{dT_{12_{ij}}}(a,b) = \sum_k \sum_{m=1}^{8} \frac{dw_{k,m}}{dT_{12_{ij}}} . \delta(a - a_k, b - b_{k,m}) \qquad (4)$$

Thus, the derivative $\frac{d\mathbf{H}}{dT_{12_{ij}}}$ is itself a histogram which can be computed using the PV interpolation scheme similarly to the way H itself is computed, but updating each entry with $\frac{dw_{k,m}}{dT_{12_{ij}}}$ instead of with $w_{k,m}$

Expressions for $\frac{dw_{k,m}}{dT_{12_{ij}}}$ can be obtained as follows:

$$\frac{dw_{k,m}}{dT_{12_{ij}}} = \nabla w_{k,m}(\mathbf{r}_k) \frac{\delta \mathbf{r}_k}{\delta T_{12_{ij}}} = \sum_s \frac{\delta w_{k,m}}{\delta r_{ks}} \frac{\delta r_{ks}}{\delta T_{12_{ij}}} \qquad (5)$$

The derivatives $\frac{\delta w_{k,m}}{\delta r_{ks}}$ of $w_{k,m}$ w.r.t. r_{kx}, r_{ky} and r_{kz} are simply computed from the expressions for $w_{k,m}$ given above, while the derivatives $\frac{\delta \mathbf{r}_k}{\delta T_{12_{ij}}}$ of $\mathbf{r}_k$ w.r.t. the elements $T_{12_{ij}}$ of the image-to-image transformation are

$$\mathbf{r}_k = \mathbf{q}_k - \mathbf{q}_{k,1}$$

$$\frac{\delta \mathbf{r}_k}{\delta T_{12_{ij}}} = \frac{\delta \mathbf{q}_k}{\delta T_{12_{ij}}} = \frac{\delta \mathbf{T}_{12}}{\delta T_{12_{ij}}} . \mathbf{p}_k$$

Due to the sparseness of $\frac{\delta \mathbf{T}_{12}}{\delta T_{12_{ij}}}$, expression 5 can be simplified:

$$\frac{\delta r_{ks}}{\delta T_{12_{ij}}} = \begin{cases} p_{kj} & \text{if } s = i \\ 0 & \text{otherwise} \end{cases}$$

such that

$$\frac{dw_{k,m}}{dT_{12_{ij}}} = \frac{\delta w_{k,m}}{\delta r_{ki}} . p_{kj} \qquad (6)$$

2.4 Computing derivatives of I using PV interpolation

Combining the expressions derived above, the derivatives of $I(\alpha)$ are given by

$$\frac{dI}{d\alpha_k} = \sum_{a,b} \frac{1}{N} . (\log_2 \frac{N.h_{ab}}{h_a.h_b} - I) . \left(\sum_{rs} \left(\frac{\delta T_{12}}{\delta \alpha_k}\right)_{rs} \frac{d\mathbf{H}}{dT_{12_{rs}}} \right)_{ab}$$

The computation of the derivatives $\frac{\delta I}{\delta \alpha_k}$ thus involves

- computing the 12 histograms $\frac{d\mathbf{H}}{dT_{12_{rs}}}$ using expressions 4 and 6; these histogram derivatives can be computed together with $\boldsymbol{H}$ by scanning the image **1** once.
- computing the derivatives $\frac{\delta T_{12}}{\delta \alpha_k}$ of the image-to-image transformation w.r.t. the affine transformation parameters using expression 2; this involves 4 x 4 matrix manipulations only.
- computing the histogram $\frac{d\mathbf{H}}{d\alpha_k}$ as a linear combination of the histograms $\frac{d\mathbf{H}}{dT_{12_{rs}}}$ and taking the sum of all entries of $\frac{d\mathbf{H}}{d\alpha_k}$ weighted by $\frac{dI}{dh_{ab}}$; this involves scanning the histogram $\mathbf{H}$ and all histograms $\frac{d\mathbf{H}}{d\alpha_k}$ once.

The computation of the gradient $\boldsymbol{VI}$ was implemented in C++ and integrated in the ***MIRIT*** program [3]. As the gradient $\boldsymbol{VI}$ consists of 12 components, computation of VI is expected to be 12 times as expensive as computation of I itself Because of some redundancy in the computations, it was found experimentally that evaluation of VI is about 10 times more expensive than evaluation of I for typical CT and MR image volumes using 256 histogram bins for both images.

Figure 1 shows traces of I and $\frac{\delta I}{\delta t_x}$ for right-to-left translation of a CT and an MR image of the brain around the registered position. Note that although I itself is continuous, its gradient $\boldsymbol{VI}$ is not A discontinuity in $\boldsymbol{VI}$ might appear each time the set of 8 nearest neighbours $l_{k,m}$ in image 2 of any sample k in image **1** changes.

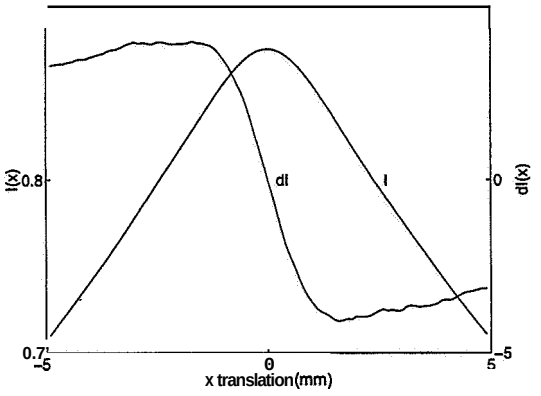


Figure 1 Traces of I and $\frac{\delta I}{\delta t_x}$ using PV interpolation for right-to-left translation of a CT and an MR image of the brain around the registered position.

3 Optimization strategies for maximization of mutual information

With the expressions for $I(\alpha)$ and its gradient $\nabla I(\alpha)$ derived in section 2, various optimization strategies can be investigated for maximization of $I(\alpha)$ These are shortly discussed below for the general case of minimization of the N-dimensional function f **(p)** with gradient $\boldsymbol{V}$f ***(p)***

3.1 Powell's direction set method

Powell's direction set method only requires evaluations of f itself and not of ∇f The method finds the N-dimensional minimum of f by repeatedly minimizing f in one-dimension along a set of N different directions, each time starting from the minimum found in the previous direction using a one-dimensional minimization method such as Brent's [4]. If all directions in the set were conjugate one to another with respect to the function to be minimized, optimization in one direction would not spoil the optimization in previous ones and only one iteration over all directions would be required to exactly find the (possibly local) minimum Powell's method incorporates a scheme to update the direction set as the optimization proceeds and to iteratively construct

such a set of conjugate directions. The set is initialized with the basis vectors in each dimension, but after each iteration in which all directions in the set are optimized over in turn, a new direction vector is constructed as the overall distance moved in parameter space in that iteration. Another line minimization is performed along this direction, which is then appended to the direction set while the first direction in the set is dropped. It can be shown that N iterations of this basic procedure yields a set of N mutually conjugate directions if $\mathbf{f}$ is a quadratic function Powell's algorithm thus exactly minimizes a quadratic $\mathbf{f}$ in $N(N+1)$ line minimizations in all.

However, for non-quadratic $\mathbf{f}$, this procedure of generating new directions might result in the directions in the set becoming linear dependent, such that after a few iterations some of the variables to be optimized over might not be represented anymore in the set The direction set has therefore to be reinitialized to the basis vectors (or to the colums of any orthogonal matrix) after every N iterations of the basic procedure. We prefer to *"give up the property of quadratzc convergence in favor of a more heuristic scheme which tries to find a few good directions along narrow valleys instead of* N *necessarily conjugate directions."*([4], pp. **416).** This involves heuristics to decide after each iteration whether to consider the overall distance moved in that iteration as a new direction and if so, which direction in the current set should be replaced by this new one. A good strategy consists in discarding the best direction, i.e. the direction along which $\mathbf{f}$ made its largest decrease, as it is likely to be a major component of the new direction that is being added [4].

Due to differences in image resolution in different directions and due to the specific shape of the objects in the scene, the order in which the different parameters are considered and are optimized over might strongly influence optimization performance and registration robustness. For instance for axial images of the brain, optimization of the horizontal translations and of the rotation around the vertical axis is much better conditioned than optimization of the vertical translation or of the pitch rotation around the left-to-right horizontal axis.

3.2 Steepest gradient descent

The steepest gradient descent method is the most straightforward method for incorporating gradient information in the minimization process. The minimum of the function is found by a number of consecutive one-dimensional line minimization steps using for instance Brent's algorithm, each step starting at the minimum found in the previous step and proceeding in the direction of the gradient at that point, i.e. the direction of steepest descent. Because each starting point is the minimum along the previous direction, the new gradient direction along which will be optimized in the next iteration is necessarily orthogonal to the previous one.

Steepest gradient descent is in general not a very good algorithm. Because the gradient generally does not point to the optimum directly and because consecutive steps towards the optimum are always at orthogonal angles, many tiny steps are usually needed before reaching the optimum, especially when going down a long and narrow valley.

3.3 Conjugate gradient methods

Conjugate gradient methods try to overcome the problems associated with the steepest gradient descent approach by trying to construct the new direction as being conjugate to the previous one with respect to the function to minimize, rather than down the gradient direction.

A scheme to construct a set of conjugate directions iteratively as the optimization proceeds has been proposed by Fletcher and Reeves [4]. In each iteration i a line minimization is performed in the direction $\mathbf{d}_i$, starting at point $\mathbf{x}_i$ and leading to point $\mathbf{x}_{i+1}$ Initially, $\mathbf{d}_1$ is set to the gradient direction $\mathbf{g}_1$ at point $\mathbf{x}_1$ After each iteration a new direction $\mathbf{d}_{i+1}$ is constructed by

$$\gamma_{i,FR} = \frac{\mathbf{g}_{i+1}.\mathbf{g}_{i+1}}{\mathbf{g}_i.\mathbf{g}_i}$$

$$\mathbf{d}_{i+1} = \mathbf{g}_{i+1} + \gamma_i.\mathbf{d}_i$$

It can be shown that, iff is quadratic, the vectors $\mathbf{d}_i$ are mutually conjugate with respect to f [4], such that this scheme arrives at the exact minimum of $\mathbf{f}$ in N iterations, requiring N gradient evaluations in all.

For non-quadratic $\mathbf{f}$, more iterations are usually required. Polak and Ribiere introduced a small change to the Fletcher-Reeves algorithm using the form

$$\gamma_{i,PR} = \frac{(\mathbf{g}_{i+1} - \mathbf{g}_i).\mathbf{g}_{i+1}}{\mathbf{g}_i.\mathbf{g}_i}$$

For quadratic f both expressions are identical because then $\mathbf{g}_i.\mathbf{g}_j = 0, i \neq j$, but for non-quadratic f there is some evidence that the Polak-Ribiere scheme converges faster than Fletcher-Reeves.

3.4 Quasi-Newton methods

The basic idea of variable metric or quasi-Newton methods is to iteratively build up a good approximation to the inverse Hessian matrix $\mathbf{J}^{-1}$ of the N-dimensional function $\mathbf{f}$ to be optimized. If the function *can* be assumed to be quadratic and $\mathbf{J}^{-1}$ is known, the step to take from the current point $\mathbf{x}_i$ in iteration i to the exact optimum $\mathbf{x}^*$ of the function can be determined by setting $\nabla f = 0$ as in Newton's one-dimensional optimization method This gives $\mathbf{x}^* = \mathbf{x}_i - \mathbf{J}^{-1}\nabla f(\mathbf{x}_i)$. Hence, $\mathbf{d}_{i+1} = -\mathbf{J}^{-1}\nabla f(\mathbf{x}_i)$ is taken as the new direction in which a line minimization is performed in iteration i Initially, $\mathbf{d}_1$ is set to the gradient direction and the first approximation $\mathbf{J}_1^{-1}$ of $\mathbf{J}^{-1}$ is set to the identity matrix. Two approaches to iteratively update $\mathbf{J}_i^{-1}$ after each iteration using function and gradient evaluations have been proposed by Davidon-Fletcher-Powell (DFP) and by Broyden-Fletcher-Goldfarb-Shanno (BFGS) [4]. For quadratic functions $\mathbf{f}$, both schemes converge to $\mathbf{J}^{-1}$ in N iterations, hence requiring N gradient evaluations in all to arrive at the exact minimum of f For non-quadratic $\mathbf{f}$, BFGS has been recognized to be superior in details of roundoff error and convergence tolerances [4].

3.5 Least squares methods

Every minimization problem of a function $\mathbf{f}$ in N dimensions can be considered as a least squares problem of recovering the N parameters $\mathbf{x}$ for which the least squares figure of merit f^2 is minimal. An elegant method for general nonlinear least squares has been put forth by Levenberg and Marquardt [4]. The method varies smoothly between the steepest descent and inverse-Hessian approaches by solving at each iteration i the incremental update $\delta\mathbf{x}_i$ from the current estimate $\mathbf{x}_i$ of the optimum to the next one from the equation

$$(\mathbf{J}, + \lambda\mathbf{I}).\delta\mathbf{x}_i = -\nabla f^2 = -2\mathbf{f}\ \nabla f \qquad (7)$$

with $\mathbf{J}_i$ the Hessian of f^2 at $\mathbf{x}_i$, $\mathbf{I}$ the identity matrix and λ a regularization parameter For λ approaching zero, equation

7 reduces to $\mathbf{J}_i.\delta\mathbf{x}_i = -\nabla f^2$, which is the inverse-Hessian method, bringing us from $\mathbf{x}_i$ directly to the optimum if f^2 is a quadratic function. For very large values of λ, $\mathbf{J}_i + \lambda\mathbf{I}$ is diagonally dominant and equation 7 reduces to $\delta\mathbf{x}_i \sim -\nabla f^2$, which is the steepest descent method.

The Hessian matrix $\mathbf{J}_i$ of f^2 is approximated by $J_{kl} \simeq 2.\nabla f_k \nabla f_l$ by ignoring the second derivative terms. At each iteration, 7 is solved for δx_i using a moderate value for λ and f^2 is evaluated at $\mathbf{x}_i + \delta\mathbf{x}_i$. If $f^2(\mathbf{x}_i + \delta\mathbf{x}_i) \geq f^2(\mathbf{x}_i)$, λ is increased to favor the steepest gradient approach and 7 is solved for a new trial step $\delta\mathbf{x}_i$. If $f^2(\mathbf{x}_i + \delta\mathbf{x}_i) < f^2(\mathbf{x}_i)$, $\mathbf{x}_{i+1} = \mathbf{x}_i + \delta\mathbf{x}_i$ is taken as the new estimate for the optimum and a new iteration is started after having decreased λ to favor the inverse-Hessian approach. Convergence is declared when f^2 decreases by a negligible amount.

The Levenberg-Marquardt method has the advantage over the other gradient-based methods discussed here that no line minimization is being performed in each iteration, which saves a lot of function evaluations.

3.6 Multiresolution techniques

All the above methods can be incorporated in a multiresolution scheme to increase speed performance. The optimization is performed first at lower resolution by using only a fraction of the voxels in image 1 to construct the joint histograms from which I and VI are computed. After convergence, the optimization proceeds at higher resolution, and eventually at full resolution, by taking more voxels in image 1 into account.

Subsampling image 1 with sampling factors f_x, f_y and f_z along the x, y and z coordinate dimension respectively, results in a speedup with a factor of $F = f_x.f_y.f_z$ in the evaluation of I and VI. If subsampling image 1 only slightly affects registration robustness such that the optimum converged to at lower resolution is close to the optimum at full resolution, it is to be expected that most of the evaluations of I and VI will be performed at lower resolution and that the number of evaluations at full resolution will be much smaller than when the optimization is done entirely at full resolution. The 2-level multi-resolution optimization approach with the lower resolution level using subsampling factors f_x, f_y, f_z may thus be up to F times faster than the single-resolution optimization approach.

However, the sampling factors f_x, f_y and f_z can not be made arbitrarily large without deteriorating registration robustness. If F is too large such that the registration at the lower resolution level does not converge to the global optimum but to some local optimum that might be still far off from the global one, the optimization at the lower resolution level does not provide a good starting position for the optimization at the higher resolution level and the number of evaluations at full resolution will be about the same as for single-resolution optimization. As a result, the multiresolution strategy might be a lot slower than single-resolution optimization, due to the additional, but useless, evaluations at lower resolution. Appropriate values for f_x, f_y and f_z can only be determined experimentally, although it is clear that this choice is linked to the image resolution along each direction.

The use of more than 2 resolution levels is usually not very efficient. If the optimization at the selected lowest resolution level already converges to an optimum which is close to the global optimum at full resolution, it does not make sense to first do another optimization at a somewhat higher resolution instead of considering full resolution right away. Indeed, the additional number of evaluations that is needed to reach convergence at any additional resolution level will increase the overall number of evaluations required, such that the speedup gained at lower resolution might be lost completely. However, considering more than 1 lower resolution level might be appropriate if the optimization at the selected lowest resolution level fails. It might then be advantageous to first try another lower resolution level at a somewhat higher resolution than the first one tried, instead of switching to full resolution immediately, because of the speedup that might be gained if the optimization at this second resolution level is indeed succesful.

4 Validation

The performance of the gradient-based optimization methods using PV interpolation was compared to that of Powell's method for 6-parameter rigid-body registration of CT and MR images of the brain for various 2-level multiresolution strategies.

4.1 Experiments

The experiments were performed on a set of 9 patient data sets, consisting of CT, MR-MPRAGE, MR-PD, MR-T1 and MR-T2 images[1]. MR-PD images were available for patients 1 to 4 only, while no MR-T2 image was available for patient 3. The characteristics of these images are summarized in table 1. The MR-MPRAGE images for patients 1 to 4 are sagittal images, those of patients 5 to 9 are coronal images. All images have a 12-bit intensity range.

For each set, all MR images were registered to the CT image using the CT image as the reference image, resulting in 30 registration experiments in total, 9 involving the high-resolution MR-MPRAGE images and 21 involving the lower resolution MR-PD, MR-T1 and MR-T2 images. The number of histogram bins used to compute the mutual information criterion was 256 for both the floating and the reference image in all cases. For each case the registration solution obtained using external markers[2] was used as reference to evaluate registration accuracy.

All experiments were performed using Powell's direction set method (POW), steepest gradient descent (STD), conjugate gradient (CJG), quasi-Newton (QSN) and Levenberg-Marquardt (LVM) optimization methods using the full floating image resolution and each of 5 2-level multiresolution strategies, resulting in a total of 900 registration results. The subsampling factors used at the lower resolution level in the $\{xyz\}$ dimensions were: {441}, {332}, {331}, {222}, {221} and {111}, with overall subsampling factors of 16, 18, 9, 8, 4 and 1 respectively. No smoothing of the images was applied prior to subsampling.

All optimizations for a given pair of images started from the same initial position with all parameters set to 0, except for a rotational offset of 5 degrees around one axis to avoid PV interpolation artefacts. The convergence parameters of the Brent line minimization algorithm were identical for all methods: the absolute precision was set to 10^{-3}, the fractional precision to 10^{-2} and the maximum number of iterations to 10. Overall convergence was declared at each resolution level if the fractional decrease of the criterion in one iteration was smaller than 10^{-5} for POW and LVM and 10^{-6} for STD, CJG and QSN, as each iteration comprises 6 line minimizations for POW, but only 1 for the other methods, and because STD, CJG and QSN are more likely to get stuck in a weak local optimum than LVM because of the line minimization.

[1] Provided by J.M. Fitzpatrick as part of the RREP project [5].
[2] Provided by J.M. Fitzpatrick as part of the RREP project [5].

Image	#	Size	Voxels (mm)	Orientation
CT	9	512^2 x (40 − 49)	(0.4 − 0.45)' x 3.0	axial
MR-MPRAGE	9	256^2 x 128	0.98^2 x (1.25 − 1.66)	sagittal/coronal
MR-PD	4	256^2 x 52	$(0.78 - 0.86)^2 \times 3.0$	axial
MR-T1	9	256^2 x (51 − 52)	$(0.78 - 0.86)^2$ x 3.0	axial
MR-T2	8	256^2 x 52	$(0.78 - 0.86)^2$ x 3.0	axial

Table 1 Characteristics of the images used in the experiments discussed in section 4.

All experiments were conducted on an SGI Octane workstation (IRIX 6.4, R10000 195 MHz, 17 SPECfp95). Because the load of this machine varied while the experiments were done, it is unsave to compare different methods by the CPU-time required to reach convergence. Instead, different experiments are compared by their number of equivalent full resolution criterion evaluations N_e. N_e is computed as the weigthed sum of the number of function and gradient evaluations at each resolution level, taking into account the subsampling factors at each level and the relative difference in complexity between function and gradient evaluations:

$$N_e = \sum_{r}^{R} \frac{N_{f,r} + \eta . N_{g,r}}{F_r}$$

with R the number of resolution levels (1 or 2 in our experiments), $N_{f,r}$ and $N_{g,r}$ the number of function and gradient evaluations at resolution level r, F_r the overall subsampling factor at that level and η the relative complexity of the gradient evaluation compared to that of the function evaluation. We used $\eta = 12$ as all 12 affine gradient components are computed each time, assuming that evaluation of each of these is equivalent to one function evaluation This is rather conservative, as a factor of 10 was found experimentally.

The values of N_e were aggregated over all patients for each method and each multiresolution strategy for the higher resolution MR-MPRAGE images in one group and for the lower resolution MR-PD, MR-T1 and MR-T2 in another The performance of the various methods was evaluated by comparing their mean N_e values using the one-sided t-test with significance level $\alpha = 0.005$.

4.2 Results

The recovered registration parameters ranged from -29 to +13 mm for translation and from -28 to +14 degrees for rotation. The accuracy of all 900 registration results was evaluated by comparison with the external marker-based reference registration solution. The error was evaluated as in [2] by the average norm of the transformation difference vector evaluated at 8 points near the brain surface. Errors ranged from 0.5 to 3.1mm and were all smaller than 1CT voxel. It was found that the error is independent of the optimization method that was used for each of the 30 different registration experiments, error variations among the 24 different optimization results was smaller than 5%, indicating that each of these converged to the same optimum.

The performance of each of the optimization strategies is summarized in figures 2 and 3. Each plot shows the values of N_e obtained for each of the 5 optimization methods using each of the 6 multiresolution strategies aggregated over all 30 CT/MR experiments. The results were aggregated for the CT/MR-MPRAGE and the CT/MR-(PD,T1,T2) experiments as no significant differences were observed between both groups. Each of the boxes in each plot has lines at the lower quartile, median and upper quartile values of N_e The lines extending from each end of the box show the extent of the rest of the data. Outliers are indicated by '+' The mean, minimum and maximal values of N_e for each case and of the recorded CPU-time are tabulated in tables 2 and 3 respectively.

The results for POW show a somewhat larger spread than those for the other methods. This is due to the fact that if after a number of POW iterations the convergence criterion has not been satisfied, another set of 6 line minimizations is being performed, while for STD, CJG and QSN the convergence criterion is evaluated after every single line minimization and for LVM after every update of the trial solution. Although the number of iterations might vary a lot for the multiresolution strategies using STD, CJG, QSN or LVM at the lower resolution levels, computational complexity is mainly determined by the number of iterations and gradient evaluations at full resolution, which is usually only 2 or 3 for most cases.

The plots show that the use of any of the 2-level multiresolution approaches results in a decrease of computational complexity by a factor of 2 for POW and 3 to 5 for the gradient-based methods For all methods, multiresolution strategies [441], [332], [331] and [222] perform significantly better than [111] and also significantly better than [221] for STD, CJG and QSN However, subsampling factors larger than 8 at the lower resolution level do not offer any significant speed advantage: for all methods there is no significant difference in performance between the multiresolution approaches [441], [332], [331] and [222] (except for QSN for which [441] performs significantly better than [331]) This can be explained by the observation that the distance between the optima converged to at lower and at full resolution increases for larger subsampling factors, such that more full resolution evaluations are required to reach convergence. Hence, the advantage of faster evaluations at lower resolution is cancelled by the additional number of evaluations required at full resolution.

LVM performs significantly better than any other method for all multiresolution strategies. At full resolution, POW and CJG perform equally well, but for any of the 2-level multiresolution strategies, CJG performs significantly better than POW STD performs significantly better than POW for most 2-level multiresolution strategies, but never significantly better than CJG QSN does never perform significantly better than any other method for all multiresolution strategies.

LVM using multiresolution strategies [441], [332], [331] or [222] performs systematically 4 to 5 times better than POW at full resolution. The CPU-times in table 3 show that the registration of the CT and MR-MPRAGE images is performed in 3 minutes CPU-time on average using LVM and the [441] multiresolution strategy, while about 13 minutes CPU-time is required on average using POW at full resolution.

4.3 Discussion

The robustness of various gradient-based optimization methods for maximization of mutual information has been inves-

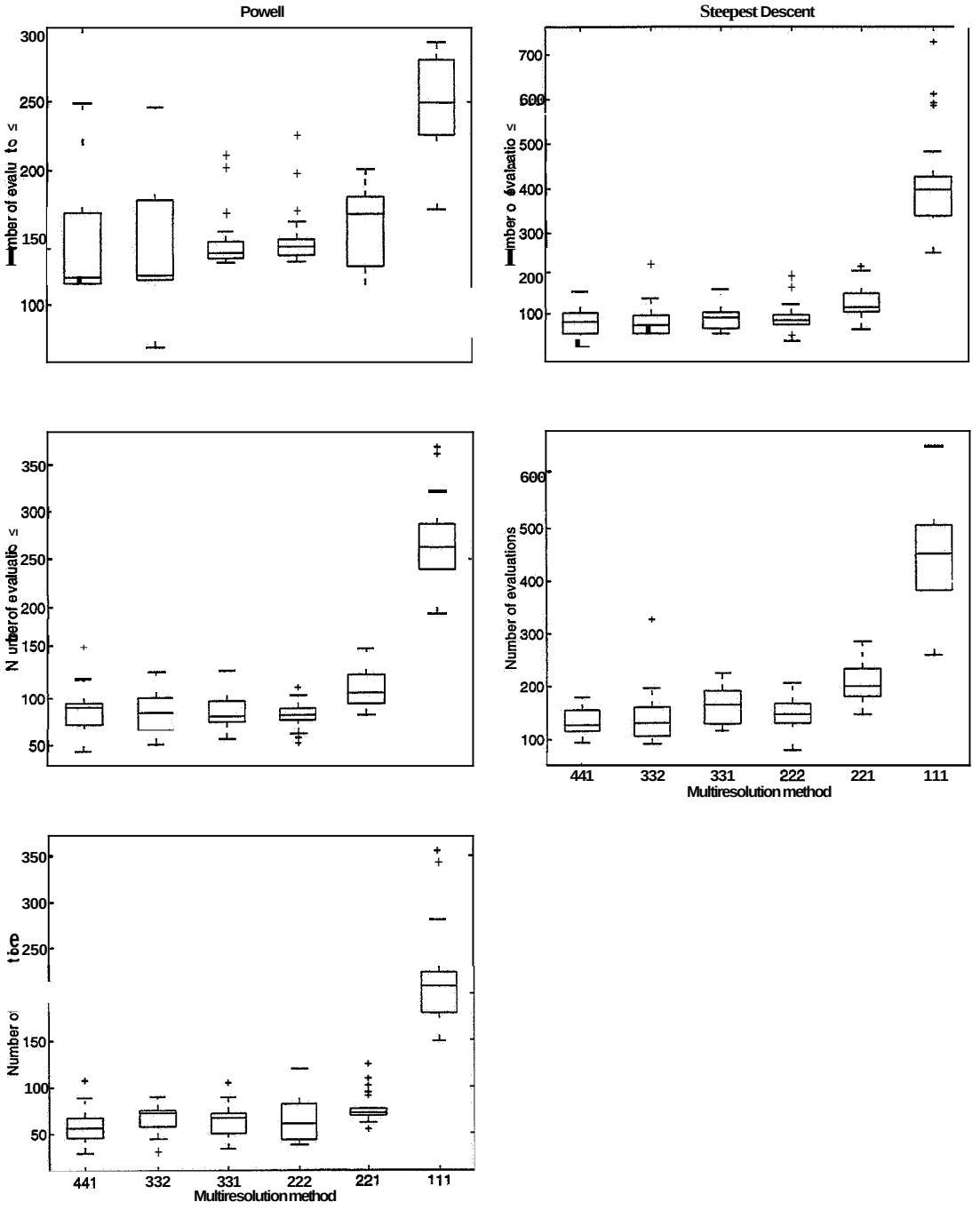


Figure 2: Number of equivalent **full** resolution function evaluations **for** each optimization method using various multiresolution strategies for all **30** CT/MR registration experiments. Each box has lines at the lower quartile, median and upper quartile values. The lines extending from each end of the box show the extent of the rest of the data. Outliers are indicated by '+'

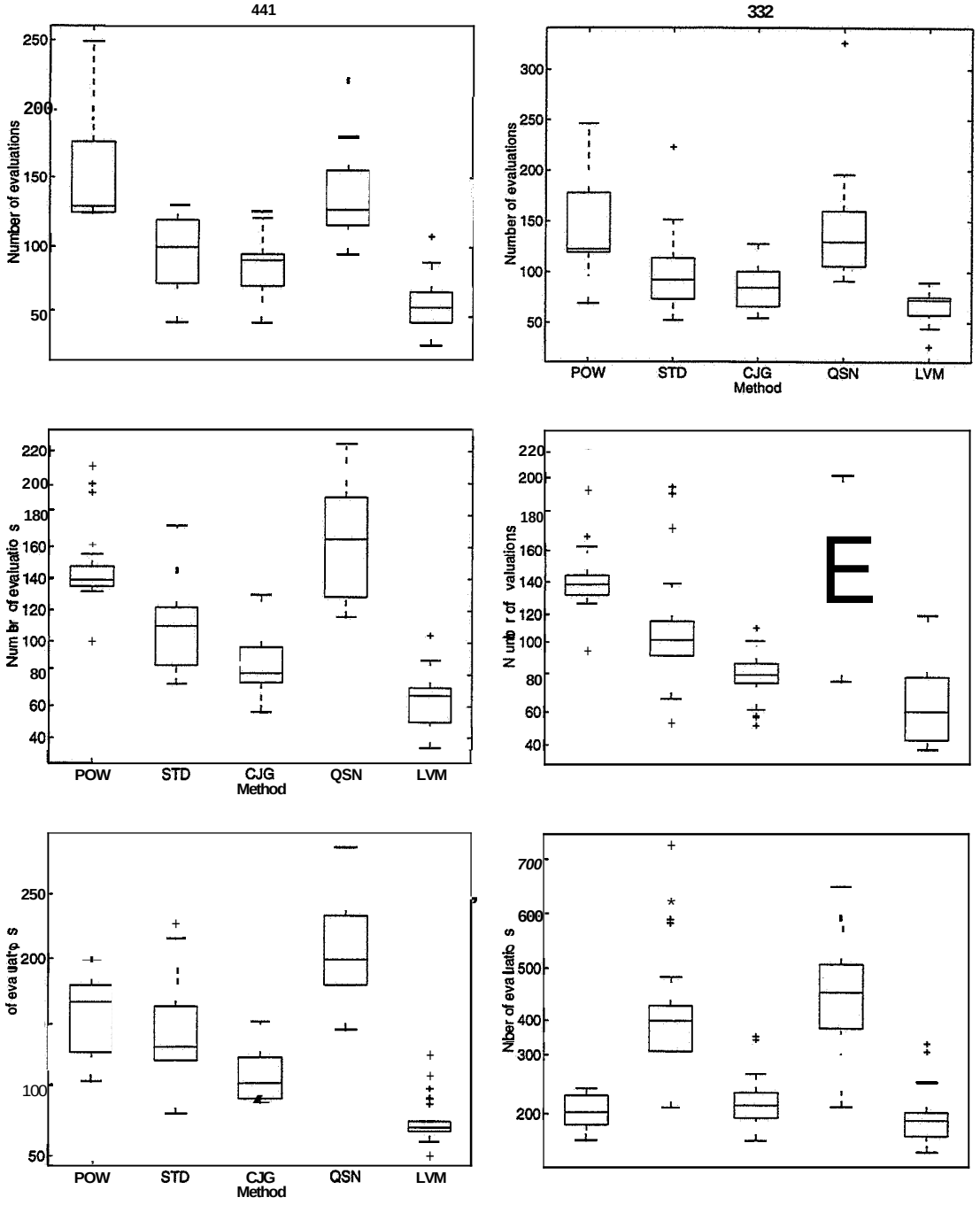


Figure 3: Number of equivalent **full** resolution function evaluations **for** each multiresolution strategy **using** various optimization methods for all **30** CT/MR registration experiments. Each box has lines at the lower quartile, median and upper quartile values. The limes extending from each end of the box show the extent of the rest of the data. Outliers axe indicated by '+'

	POW	STD	CJG	QSN	LVM
441	148 (116,249)	99* (45,169)	89* (43,150)	134 (94,180)	56* (29,108)
332	141 (69,246)	99* (52,223)	85* (54,128)	141 (91,326)	67* (26,90)
331	145 (98,209)	108* (74,173)	87* (56,129)	162 (116,224)	63* (34,105)
222	147 (96,223)	110* (51,198)	82* (52,111)	146 (79,205)	62* (38,119)
221	157 (107,201)	144 (82,225)	109* (82,152)	209 (146,285)	76* (54,124)
111	249 (171.294)	409 (258.728)	268 (190.369)	459 (260.656)	214* (149.355)

Table 2: Mean number of equivalent full resolution function evaluations for each method and multiresolution strategy over all 30 CT/MR registration experiments. The numbers between brackets are the corresponding minimal and maximal values for each case. Mean values that are significantly lower than the corresponding values obtained with POW using the same multiresolution strategy are indicated with * for each method.

	POW	STD	CJG	QSN	LVM
441	445 (249,795)	349 (134,574)	339 (220,573)	487 (271,752)	181 (106,273)
332	403 (200,607)	400 (282,713)	305 (220,374)	583 (316,1202)	255 (172,359)
331	460 (283,670)	409 (248,550)	300 (207,418)	597 (337,812)	237 (145,286)
222	465 (296,714)	397 (188,618)	287 (168,412)	505 (356,682)	231 (171,321)
221	481 (312,673)	538 (337,792)	374 (255,439)	769 (515,1017)	305 (223,469)
111	764 (514,1216)	1493 (910,2068)	957 (663,1120)	1617 (975,2146)	815 (554,1219)
441	255 (178,400)	231 (142,421)	195 (135,297)	315 (232,412)	156 (78,302)
332	241 (106,381)	210 (128,369)	188 (130,295)	310 (217,494)	173 (73,226)
331	248 (196,338)	249 (170,382)	202 (147,289)	379 (259,529)	166 (87,264)
222	241 (196,387)	249 (178,447)	185 (124,265)	330 (173,432)	162 (95,300)
221	273 (180,340)	322 (227,503)	245 (188,293)	480 (352,659)	195 (163,301)
111	367 (251,468)	800 (523,1286)	522 (385,705)	946 (553,1282)	511 (361,681)

Table 3: Mean CPU-time in seconds for each method and multiresolution strategy over all patient data sets for the 9 CT/MR-MPRAGE experiments (top) and the 21 CT/MR-(T1,T2,PD) experiments (bottom). The numbers between brackets are the corresponding minimal and maximal values for each case.

tigated for registration of CT and MR images of the brain and their performance compared with Powell's method used in [2]. Although the affine gradient computation is a computationally expensive operation, large speedups of the registration process can be realized without a loss of robustness by using a multiresolution optimization strategy. A 2-level multiresolution approach using Levenberg-Marquardt optimization method was found to be the most efficient method, outperforming Powell's method applied at full image resolution by a factor of 4. Further improvements in efficiency might be achieved by using different convergence criteria for the different resolution levels.

Subsampling of the images may occasionally introduce additional local optima in the registration criterion, especially when large subsampling factors are used. However, the results of section 4 show that large subsampling factors ([332], [441]) do not result in significantly better performance compared to moderate subsampling factors ([222], [331]) In all experiments, integral subsampling factors and nearest neighbour interpolation were used to subsample the image 1, such that no new intensity values are being introduced by the sampling process. It is unclear how the use of non-integral subsampling factors with trilinear interpolation, prior smoothing of the images or reducing the number of image histogram bins would influence the behaviour of the registration criterion at the lower resolution level.

The performance of the various optimization strategies was compared by the number of equivalent full resolution function evaluations required to reach convergence. This assumes that similar convergence criteria were specified for each method and that all methods have similar precision The variation in the optimum found by each method was evaluated by comparing for each of the 30 experiments the 6 registration parameters $\boldsymbol{a}$ for each of the 24 optimization results to the parameters found for the best optimization result $\boldsymbol{a}^*$ The parameter vectors were compared by the norm $|\delta\boldsymbol{\alpha}| = |\boldsymbol{\alpha} - \boldsymbol{\alpha}^*|$ of their difference vector The distribution of $\delta\alpha|$ for each method is shown in figure 4 All values of

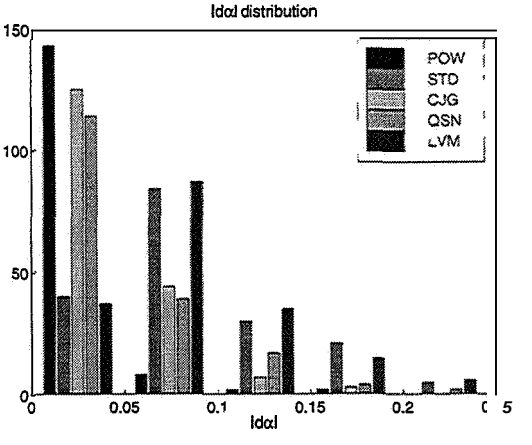


Figure 4. Evaluation of optimization precision for each method. This plot shows the distribution of the variation of the optimum for each method, computed for each experiment as the norm of the parameter difference vector $\delta\alpha|$ of each result $\boldsymbol{a}$ compared to the best one α^* The variation is smallest for POW and largest for STD and LVM, but the precision is better than 0.25 mm and 0.25 degrees for each method.

$|\delta\boldsymbol{\alpha}|$ are smaller than 0.25, meaning that for each experiment none of the optimization results differ by more than 0.25 mm for the translational and 0.25 degrees for the rotational parameters All differences between corresponding results are therefore smaller than 1 CT voxel everywhere in the image. The mean values of $|\delta\boldsymbol{\alpha}|$ are 0.03, 0.09, 0.04, 0.05 and 0.09

for POW, STD, CJG, QSN and LVM respectively The optima found using POW are thus most clustered, while those obtained using STD or LVM are least clustered. This indicates that more stringent convergence criteria should have been specified for STD and LVM to obtain the same precision as for the other methods. This should be taken into account when comparing the performance of the different optimization methods.

The optimization methods evaluated here are all local optimization methods. Global optimization methods, such as simulated annealing, have not been considered. These methods aim at finding the global optimum in the presence of many local optima of comparable strength. But the local optima observed in the mutual information registration criterion usually have very small attraction basins and are generally dominated by the much stronger global optimum with very large attraction basin Simulated annealing or other global optimization methods therefore do not offer specific advantages over the methods considered here.

All experiments discussed in section 4 involve matching of CT and MR images of the brain. Registration of PET to MR or CT images is less computationally expensive because of the lower resolution of the PET images and Powell's method has been demonstrated to be sufficiently performant [2, 8]. Experiments on CT and MR images of the prostate (512 x 512 matrix, 0.65 mm pixel size, 5 mm slice distance, 22 MR slices, 43 CT slices) have confirmed the superior performance of the Levenberg-Marquardt method using a 2-level multiresolution strategy

The results discussed above demonstrate that large image volumes as used in clinical practice can be robustly matched in less than 5 minutes on a state-of-the-art workstation. Thls is comparable to the time that is usually required to transfer the images over the local hospital network from the scanners to the workstation, to load the images from disk in the computer memory and to display them on the workstation screen Registration of the images is hence no longer a bottleneck for their integrated visualization and analysis. It is to be expected that this will further stimulate the use of multimodal data in routine clinical practice.

5 Conclusion

Mutual information of image intensitiesis a continuous function of the f f i e registration parameters when PV interpolation is being used. Expressions for its gradient can be derived analytically and exactly evaluated numerically Gradient-based optimization procedures can thus be considered for maximization of the mutual information of the two images to be registered. Evaluation of the gradient is about 10times as expensive as evaluation of the criterion itself, but large speedups can be achieved by using a multiresolution approach. The performance of various multiresolution optimization strategies has been evaluated for registration of CT and MR images of the brain Our results show that the conjugate gradient and Levenberg-Marquardt optimization methods using a two-level multiresolution strategy with subsampling factors [222] perform systematically 3 to 4 times better than Powell's method at full resolution, while the use of subsampling factors larger than [222] does not result in increased performance. High resolution CT and MR image volumes can be robustly matched in less than 5 minutes CPU-time on current workstations Registration is therefore no longer a bottleneck in the analysis of multimodal images in clinical routine.

6 Acknowledgements

This research was supported by IBM Belgium (Academic, Joint Study) and by the National Fund for Scientific Research, Belgium, under grant numbers FGWO 3.0115.92, 9.0033.93 and G.3115.92 This research is performed within the context of the EASI project "European Applications for Surgical Interventions", supported by the European Commission under contract HC1012 in their "4th Framework Telematics Applications for Health" RTD programme. The partners in the EASI consortium are Philips Medical Systems Nederland B.V , Philips Research Laboratory Hamburg, the Laboratory for Medical Imaging Research of the Katholieke Universiteit Leuven, the Image Sciences Institute of Utrecht University & University Hospital Utrecht, The National Hospital for Neurology and Neurosurgery in London, and the Computational Imaging Science group of Radiological Sciences at UMDS of Guy's and St. Thomas' Hospitals in London.

References

[1] T M Cover, and J.A Thomas, *Elements of Information Theory*, New York, N Y. John Wiley & Sons, 1991

[2] F Maes, A Collignon, D. Vandermeulen, G. Marchal, and P Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Bans. Medical Imaging*, Vol. 16, no. 2, pp. 187-198, April 1997

[3] F Maes, D. Vandermeulen, and P Suetens, "Multimodality Image Registration using Information Theory MIRIT Version 97/08 Manual Pages," *Technical report K U.Leuven, Dept. of Electrical Engineering, ESAT/MI2*, Nr KUL/MI2/9708, September 1997

[4] W H. Press, B.P Flannery, S.A Teukolsky, and W T Vetterling, "Numerical Recipes in C," 2nd Edition, Cambridge, UK. Cambridge University Press, 1992

[5] J.M Fitzpatrick, Principal Investigator, "Evaluation of Retrospective Image Registration," National Institutes of Health, Project Number 1R01 NS33926-01, Vanderbilt University, Nashville, TN, 1994.

[6] C. Studholme, D.L.G. Hill, D.J Hawkes, "Automated 3-D Registration of MR and CT Images of the Head," *Medical Image Analysts*, Vol. 1, no. 2, pp. 163-175, 1997

[7] W M Wells III, P Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Medical Image Analysis*, Vol. 1, no. 1, pp. 35-51, 1996.

[8] J West, J.M. Fitzpatrick, M Y Wang, B.M Dawant, C.R Maurer,Jr., R.M Kessler, R.J. Maciunas, C Barillot, D. Lemoine, A Collignon, F Maes, P Suetens, D. Vandermeulen, P.A. van den Elsen, S. Napel, T.S. Sumanaweera, B. Harkness, P.F Hemler, D.L.G Hill, D.J Hawkes, C Studholme, J.B.A. Maintz, M.A. Viergever, G. Malandain, X Pennec, M.E. Noz. G.Q. Maguire,Jr., M Pollack, C.A Pellizari, R.A. Rob, D. Hanson, R.P Woods, "Comparison and evaluation of retrospective intermodality brain image registration techniques," *Journal of Computer Assisted Tomography*, Vol. 21, no. 4, pp. 554566, 1997

Anomaly Detection through Registration

Mei Chen, Takeo Kanade, Dean Pomerleau, Henry A Rowley
({meichen, tk, pomerlea, har}@cs.cmu.edu)

Abstract

We study an application of image registration in the medical domain. Based on a 3-D hierarchical deformable registration algorithm, we developed a prototype for automatic registering a standard atlas to a patient's data to create a customized atlas. The registration algorithm can also be applied to detect asymmetry in the patient data to help indicate the existence and location of any pathology. We have conducted experiments on 11 MRI scans of normal brains, 3 MRI and 1 CT scan of brains with pathologies.

1. Motivation

Human anatomy presents a challenge to image registration algorithms. Because of genetic and environmental factors and because of diseases, biological structures have a large range of variation in appearance. In neuroscience, a major problem in the cross-patient analysis of brain images is morphological variability. Aside from the normal variations, various neurological conditions affect the gross anatomical shapes of the brain. For example, pathologies like bleeding or tumors can cause a shift of brain structures called mass effect. Traditionally, doctors have been using manual registration to detect lesions that make brain structures deviate from the norm, to analyze variations between normal brains, and to plan surgeries. Since the manual registration of anatomical structures is labor-intensive and prone to be inconsistent, many researchers are investigating automatic registration methods. While there have been some encouraging results, due to the complexity of the problem, it remains unsolved.

We present a 3-D *hierarchical deformable registration algorithm,* and a prototype for Anomaly Detection thrOugh REgistration, *ADORE*. **ADORE** is designed to automatically and accurately match an atlas (a hand-segmented image set of normal anatomy) to a patient's data to create a customized atlas, as well as to indicate any pathologies in the patient's data. It will facilitate image-based retrieval of similar cases in medical databases, assist doctors in detecting pathologies, comparing different pathologies' impact on brain morphology, observing the development of a pathology over time, and studying the functions of different parts of the brain.

2. Problem Definition

Considering the human head as a three dimensional volume, the task of registration is to extract and match the corresponding structures from different volumes. Registration may be performed on either a single imaging modality, or on multi-modal data, e.g. *computerized tomography* (CT) and *magnetic resonance imaging* (MRI). MRI is good at revealing soft tissue structures, while CT is good at uncovering bony structures.

In neurosurgery, the three principal axes of the head are called *axial, sagittal,* and *coronal* (see Figure 1). An MRI or CT scan consists of a series of parallel cross-sections along one of these axes. Figure 2 shows examples of a person's MRI scans along the three axes.

In practice, there is no enforced standard on the acquisition of the image data, so the axis along which the cross-sections are scanned may be at an angle to the principal axes, and the spacings between consecutive cross-sections may be non-uniform. Moreover, a data set can focus on a sub-section of the head if so desired. These complications are illustrated in Figure 3, which depicts an example scan pattern.

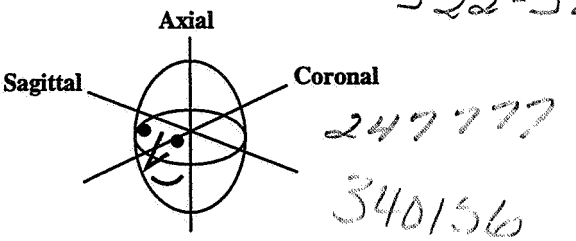


Figure 1. The three principal axes of a head volume.

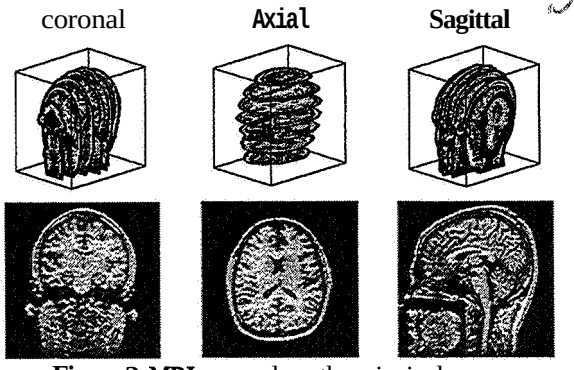


Figure 2. MRI scans along the principal axes.

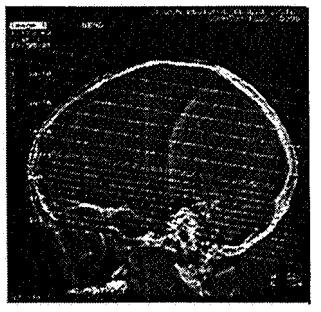


Figure 3. A typical axial scan pattern (indicated by the white lines). Note that the scanning direction is at an angle to the axial axis, the inter-scan spacings are non-uniform, and the scans do not cover the entire head volume.

One problem we will discuss is matching an atlas to a patient's data to create a customized atlas. The brain atlas we use is a set of 123 coronal T1 weighted MRI scans of a normal brain where each voxel measures *0.9375* x *0.9375* x 1.5 mm^3. 144 anatomical structures were hand segmented and labelled (courtesy of the Brigham and Women's Hospital of the Harvard Medical School) (see Figure 4). By matching the atlas to a patient's data to create a customized atlas, we can segment and label the anatomical structures in the patient's data. Figure 5 depicts the scenario in which the labels of anatomical structures in the customized atlas are used to segment the corresponding anatomical structures of the patient. Note that the original atlas (Figure 4, right) is warped into a customized atlas for the patient (Figure 5, left). We could choose to warp the patient data to match the atlas. However, since the patient's data will be of more interest for diagnosis and analysis, we prefer to keep it unchanged and warp the atlas instead.

Aside from creating the customized atlas, we will investigate the

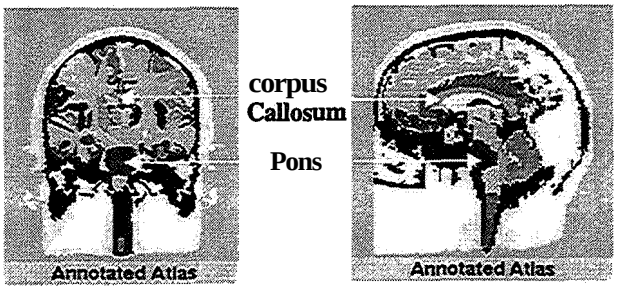


Figure 4. Coronal and sagittal cross-sections of the atlas. Two anatomical structures are annotated for illustration.

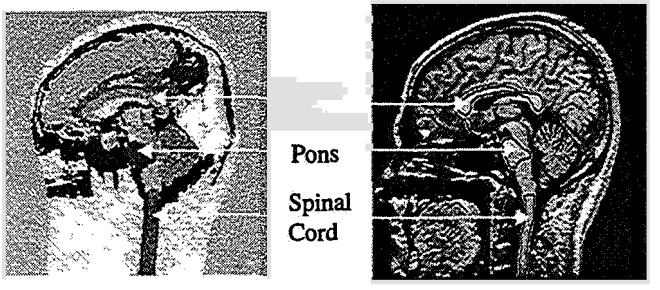


Figure 5. The scenario in which anatomical structures in a patient volume (right) are precisely segmented and labelled using information from the customized atlas (left).

problem of using registration to indicate any pathology in the patient's data. Figure 6 displays the corresponding axial cross-sections of the atlas and a brain with pathology. The atlas exhibits an approximate symmetry about the central line, whereas the symmetry is destroyed in the patient's data due to the existence of the pathology, Note that the shape, size, location, and intensity of the anatomical structures are different in the two data sets.

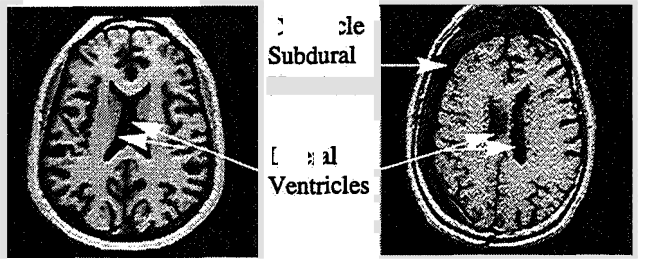


Figure 6. The corresponding axial cross-sections of the atlas (left) and a brain with pathology (right). Note the exhibition of symmetry in the atlas, and the loss of symmetry in the patient. Also note the difference in the shape, size, location, and intensity between the corresponding anatomical structures.

The variations between different data sets stem from two sources. The *extrinsic* source is the scanning process, which results in different scanning axes, different resolutions, or intensity inhomogeneities. The *intrinsic* sources are differences between different people's anatomical structures. or the existence of pathology. Variations from both sources need to be addressed to bring different image volumes into alignment.

3. Current Approach for Matching

Variations caused by extrinsic sources affect the orientation, scale, and intensity consistency of the volumes, while variations resulting from intrinsic sources are manifested as differences in the shape, size, texture, and location of the corresponding anatomical structures. Because of their different natures, we decompose the registration problem to address them separately.

We adopt a voxel-based approach which assumes no prior segmentation among voxels in a volume. An alternative is a *feature-based* approach, in which features, such as boundaries of the anatomical structures, are extracted and employed in the registration. Because the anatomical structures in the human brain have complex shapes, non-uniform textures, and ill-defined boundaries, we wish to avoid additional errors incurred by inaccurate feature detection. Alternative approaches to anatomical registration are reviewed and compared in section 7.

3.1. Highlights of the Approach

The major points of the current matching method are listed below. We will elaborate on them in the following sections.

- **Preprocessing**
  - Apply a maximum-connected-component method to extract the data of interest (the head volume), from the background to reduce distortions caused by background noise (4.1.1).
  - Normalize the intensities of the volumes to the same range so they are comparable (4.1.2).
- **Hierarchical deformable registration**
  - Apply a global transformation (3-D rotation, uniform scaling, and translation) to the atlas to grossly align it with the patient volume. An iterative optimization algorithm is used to determine the transformation parameters (3.2). This process adjusts for variations caused by all extrinsic factors except for intensity inhomogeneities (which was dealt within preprocessing).
  - Employ a smooth deformation represented by the warping of a grid of 3-D control points to approximately align the corresponding anatomical structures in the atlas to those of the patient. An iterative optimization algorithm is used to determine the deformation parameters (3.3.1). This process partially adjusts variations caused by intrinsic factors.
  - Employ a *fine-tuning* deformation in which each voxel moves independently to improve the alignment. An iterative optimization algorithm is used to determine the deformation parameters (3.3.2). This process further adjusts for variations caused by intrinsic factors.

3.2. Registration via Global Transformation

We address the variations from extrinsic sources by applying a global transformation to the atlas volume to grossly align it with the patient volume. This global *transformation* is composed of 3-D rotation, translation, and uniform scaling.

3.2.1 Representation of the Global Transformation

Figure 7 shows the coordinate systems employed in the global transformation. The origins of the coordinate systems in the atlas volume and the patient volume are placed at their centroids (center of mass). The Z axis coincides with the axis along which the volume was scanned.

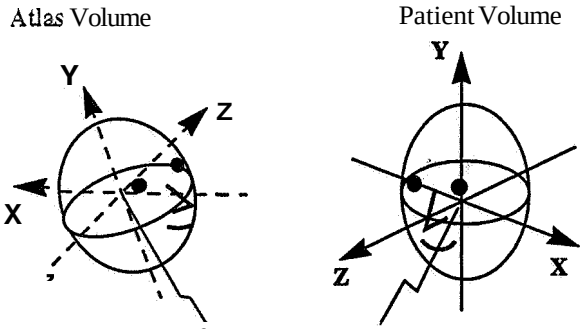


Figure 7. The coordinate system.

The three dimensional rotation is represented by a quaternion.

$$quaternion = q_0 + iq_x + jq_y + kq_z \quad (1)$$

A quaternion can be thought of as a complex number with three imaginary parts. A 3-D rotation by an angle θ about an axis defined by the unit vector $(\omega_x, \omega_y, \omega_z)$ can be represented by a unit quaternion.

$$unitquaternion = \cos\frac{\theta}{2} + \sin\frac{\theta}{2}\left(i\omega_x + j\omega_y + k\omega_z\right) \quad (2)$$

Thus the imaginary part of the unit quaternion gives the direction of the rotation axis in 3-D space, whereas the angle of rotation can be recovered from the real part or the magnitude of the imaginary part of the quaternion [3].

The three dimensional translation is represented by vector $(t_x, t_y, t_z)^T$, which represents the displacement of the origin of the atlas coordinate system with respect to the origin of the patient coordinate system.

The uniform scaling is denoted by a scalar s.

The order in which we apply these transformations to the atlas volume is first rotation about the centroid of the atlas, then scaling, then translation. Rotation aligns the atlas volume to the same orientation as the patient volume. Scaling adjusts the size of the atlas volume grossly to that of the patient. Translation removes any position difference not accounted for by the alignment of the centroids of the volumes.

3.2.2 Determining the Global Transformation

There are eight parameters to be determined for the *global transformation* T, $(q_0, q_x, q_y, q_z, t_x, t_y, t_z, s)$. Because the atlas and the patient volumes are innately different, we should not expect to find a transformation that exactly matches them. We can only pursue a transformation that minimizes the error,

We define the quality of a match in terms of the sum of squared differences (SSD) between the intensities of corresponding voxels in the two volumes:

$$SSD(T) = \sum_{(x,y,z)} \left(I_{Patient}(x,y,z) - I_{Atlas}(T(x,y,z)) \right)^2 \quad (3)$$

$I_{Patient}(x,y,z)$ is the intensity of voxel $[x,y,z]_{patient}$ in the patient volume, and $I_{Atlas}(T(x,y,z))$ is the intensity of voxel $[T(x,y,z)]_{atlas}$ in the atlas volume. $T(x,y,z)$ is (x,y,z) after applying the *global transformation* T The SSD is a function of this transformation T We apply the Levenberg-Marquardt non-linear optimization algorithm to iteratively adjust the transformation parameters to reduce the SSD. The iteration continues until the changes in the transformation parameters are below a user defined threshold, at which point the *global transformation* is considered to be recovered. We employ multi-resolution processing and stochastic sampling for efficiency and to help prevent the optimization process from being trapped in local minima (see section 4.2).

After applying the transformation, the voxel coordinates in the atlas may not have integral coordinates. Tri-linear interpolation is employed to determine a voxel's intensity from its eight bounding neighbors (see Figure 8). Notice that the result of the tri-linear interpolation is not affected by the order of the three linear interpolations along the three axes. If the voxel falls outside of the range of the volume, its effect is ignored.

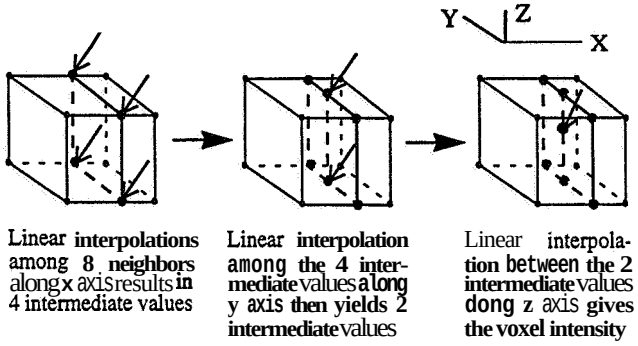


Figure 8. Tri-linear interpolation gives the intensity of a voxel by interpolating among its 8 bounding neighbors.

Figure 9 displays the gray level data of the atlas, a patient's data, and the result of using the global transformation to align the atlas with the patient. Note that the atlas volume is rotated (approximately 90 degrees, from the coronal view to the sagittal view), translated, and scaled to grossly match with the patient's volume. Labels of several anatomical structures in the transformed atlas are directly projected to the patient volume. They only roughly align with the corresponding anatomical structures of the patient, and considerable discrepancies remain in their shape, size, and location.

3.3. Registration via Deformation

After *the global transformation,* the extrinsic variations between the atlas and the patient volumes are reduced. However, the existence of intrinsic variations impedes accurate segmentation and correspondence of anatomical structures, as shown in Figure 9.

Because brain morphology is intricate, the discrepancy between the atlas and the patient cannot be fully addressed by applying one transformation to the whole atlas volume. Localized deformation is necessary.

3.3.1 Smooth Deformation

Although the variations across individuals *can* be large, the shape and density of corresponding anatomical structures are still distinct enough for them to be related. Therefore, the intensity difference between spatially corresponding voxels can act as the deforming force. The deformation process causes atlas voxels to be spatially shifted towards their counterparts in the patient volume.

The most intuitive way to represent the deformation would be a 3-D displacement for each voxel. This will allow each voxel to deform freely, but it can only succeed when the voxels' initial positions are close to their sought positions. The intrinsic variations

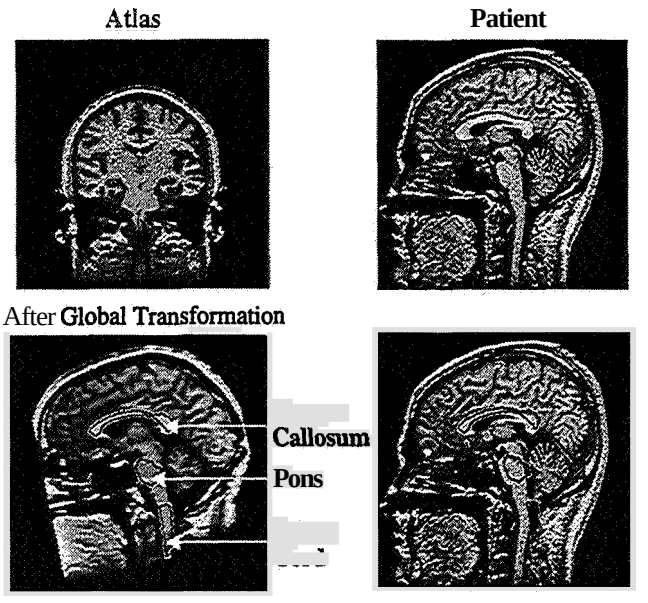


Figure 9. Cr sections of th s volume and a patient's volume before (top *row)* and after (bottom row) registration v: *glo l* nsfor i. Notice the atlas is rotated from a c or l v to align with the sagittal view of the patient. The outlines of anatomical structures in the transformed atlas are li l: proje d or the p i t ol N t th it they gf y lig th ir c p in the tie data. across individuals make ne *global transformation* unable to align the atlas to the patient precisely enough for a good initial alignment. W d a more robust r ati f intermediate l

- t the Smooth [ ]

Our solution is to ensure a *smooth deformation* by placing a 3-D c grid in e volume. The vertices of the nt i are control t e l sp c n s of the control p deter- r the deformation of e voxels y enclose. Th although the control points can deform freely, the voxels inside each control grid cell are forc to defc n ot Since the t of control points is orders of magnitude lower than the of voxels, th n e pa a to estimate, making l deformation process more stable

The simplest 3-D grid is composed of rectangular prisms, i n which case the control grid is a egul r sub-sampling f the voxel Figure 10 shows an example of a control grid cell and an rt y defor ti of it. Szeliski employed similar representation in 2-D registration [7 d 3 i of surfaces, [6]. Figure 11 is a 2-D illustration of the *smooth deformation* process.

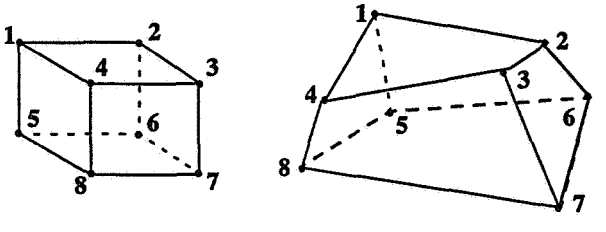


Figure 10. A control grid cell whose vertices are control points (left), and an arbitrary deformation of it due to movements of the control points (right).

- **Estimating the Smooth Deformation**

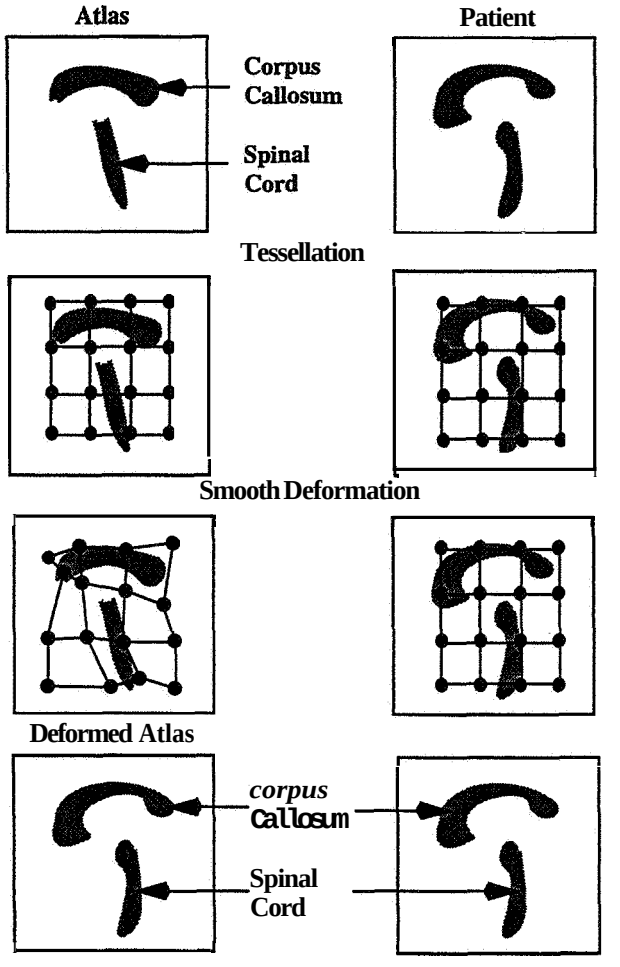


Figure 11. A 2-D illustration of the smooth *deformation.* The first row is the original data. The second row is the original data overlaid with the regular control grids. In the third row the control points of the atlas are shifted towards their counterparts in the patient. In the last row the atlas is deformed to match the patient.

Although we represent the *smooth deformation* using only the 3-D displacements of the control points, we estimate the deformation force using the intensity differences over the whole volumes. For a voxel $[x, y, z]_{patient}$ in the patient volume with intensity $I_{patient}(x, y, z)$, suppose the control grid cell it belongs to is the ith in the array of all control grid cells $Cell_{patient}[i]$. The corresponding grid cell for the atlas volume is $Cell_{atlas}[i]$. From the relative position of $[x, y, z]_{patient}$ with respect to the 8 control points of $Cell_{patient}[i]$ in the patient volume, we can hi-linearly interpolate the location of its counterpart in the atlas volume $[S(x, y, z)]_{atlas}$ using the 8 control points of $Cell_{atlas}[i]$ Should $[S(x, y, z)]_{atlas}$ have nor?-integral coordinates, hi-linear interpolation is applied to compute its intensity from its 8 neighboring voxels in the atlas (see Figure 8). In the case that $[S(x, y, z)]_{atlas}$ falls outside the atlas volume, it is ignored.

Under an ideal smooth *deformation,* the atlas volume should completely align with the patient volume, and $I_{atlas}(S(x, y, z))$ should equal to $I_{patient}(x, y, z)$. In practice, $I_{atlas}(S(x, y, z))$ and $I_{patient}(x, y, z)$ differ, the best deformation minimizes these differences. We take a similar approach to that of solving for the global transformation in 3.2.2. We use the sum of squared differences (SSD) between all voxels' intensities in the volumes as the

error function, and employ the Levenberg-Marquardt iterative nonlinear optimization method to determine the ***smooth deformation*** parameters that best match the two volumes.

The only difference between this optimization process and the one in **3.2.2** is that the parameters we compute are not the global transformation parameters, but the collection of local ***smooth deformation*** parameters for each control grid cell. The control grid cells do not deform independently, the deformations are linked by shared control points across the **3-D** control grid. They collectively define the ***smooth deformation.***

Figure **12** shows the effect of the ***smooth deformation*** on the previous example. Notice that the atlas volume is deformed in **3-D** to match the patient volume. The labels of the anatomical structures in the deformed atlas approximately align with the corresponding structures of the patient.

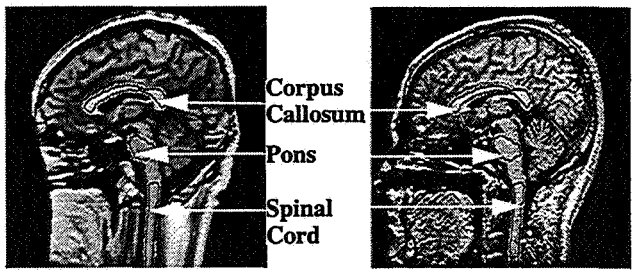


Figure 12. Corresponding cross-sections of the atlas (left) and the patient (right) after registration via ***smooth deformation.*** Note the atlas is deformed to match the patient. Labels of anatomical structures in the atlas approximately align with the corresponding structures in the patient.

3.3.2 Fine-Tuning Deformation

The smooth ***deformation*** is effective and robust at deforming the atlas to approximately match its anatomical structures to that of the patient, even when the global transformation gives poor initial alignment. However, since only the control points are allowed to deform independently, it can not account for any details smaller than the size of a control grid cell. To adjust to finer details, we apply a *fine-tuning* ***deformation.***

- **Representing the Fine-Tuning Deformation**

Similar to the smooth deformation, the intensity difference between spatially corresponding voxels acts as the deforming force. The difference is that the **3-D** displacement of each voxel in the atlas volume partakes in the representation of the fine-tuning ***deformation.*** Therefore each voxel can deform freely, allowing the deformation to attend to details. This procedure is similar to the approach discussed in [IO].

- **Estimating the Fine-Tuning Deformation**

For a voxel $[x, y, z]_{patient}$ in the patient volume with intensity $I_{patient}(x, y, z)$, its corresponding voxel $[D(x, y, z)]_{atlas}$ in the atlas volume has intensity $I_{Atlas}(D(x, y, z))$ $D_{(x,y,z)}$ is (x, y, z) after applying the current deformation D. Suppose $\delta D(x, y, z)$ is the difference between the current deformation D and the optimum deformation, then we have:

$$I_{Atlas}(D(x,y,z) + \delta D(x,y,z)) - I_{Patient}(x,y,z) = 0 \quad (4)$$

The first order Taylor expansion of the first term in (4) gives

$$I_{Atlas}(D(x,y,z)) + \delta D(x,y,z) \quad \nabla I_{Atlas}(D(x,y,z)) \quad (5)$$

From (4) and (5), one solution for $\delta D(x, y, z)$ is:

$$\delta D(x,y,z) = \frac{I_{Patient}(x,y,z) - I_{Atlas}(D(x,y,z))}{\|\nabla I_{Atlas}(D(x,y,z))\|^2} \nabla I_{Atlas}(D(x,y,z)) \quad (6)$$

To prevent the deformation parameters from going out of bounds when the atlas' gradient $\nabla I_{Atlas}(D(x,y,z))$ is close to zero, we add a stabilizing factor α

$$\delta D(x,y,z) = \frac{I_{Patient}(x,y,z) - I_{Atlas}(D(x,y,z))}{\|\nabla I_{Atlas}(D(x,y,z))\|^2 + \alpha} \nabla I_{Atlas}(D(x,y,z)) \quad (7)$$

The deformation D is recovered by iteratively solving for δD, until δD is smaller than a user defined threshold. Because the number of deformation parameters is **3** times the number of voxels, the problem is under-constrained. We apply **3-D** Gaussian smoothing to the **3-D** deformation parameters after each iteration to constrain the deformation. This process is similar to an optical flow algorithm. Trilinear interpolation is used to compute $I_{Atlas}(D(x, y, z))$ when the voxel coordinates after deformation are not integers.

Figure **13** is the result of the previous example after the fine-tuning deformation. Note that the atlas is further warped in **3-D** to better align with the patient. Labels of anatomical structures in the atlas match well with their counterparts in the patient.

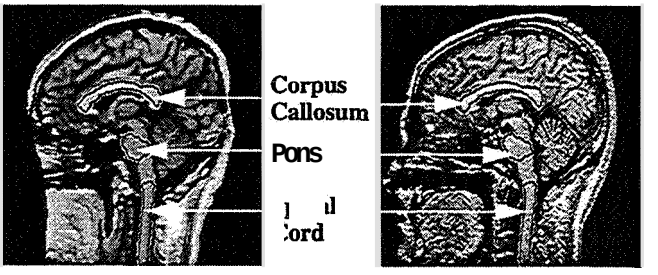


Figure 13. Corresponding cross-sections of the atlas (left) and the patient (right) after registration via *fine-tuning* ***deformation.*** Note the atlas is warped more to match the patient. Labels of anatomical structures in the atlas match well with their counterparts in the patient volume.

Once the atlas is deformed to align with the patient, we call it a *customized atlas.* Labels of anatomical structures in the *customized atlas* can be used to directly segment and label the patient volume.

4. Algorithm Implementation Details

4.1. Preprocessing

To improve the robustness of the algorithm, we apply two types of preprocessing: background separation and histogram normalization.

4.1.1 Background Separation

Since we take a voxel-based approach with no prior segmentation, each voxel in the data is assumed to play an equal role in the registration process. There is usually background noise in the data, which necessitates background separation to ensure only relevant data contributes to the registration procedure.

We first automatically threshold the volume to eliminate some of the dim background noise. The threshold is the first valley of the intensity histogram, after it is smoothed using a Gaussian filter (see

Figure 14). Since the region containing the head is connected, we apply a connected-component algorithm to the binary volume and find the largest connected component. Any holes caused by dark regions inside the head are filled in. We only consider data within that component in later processing.

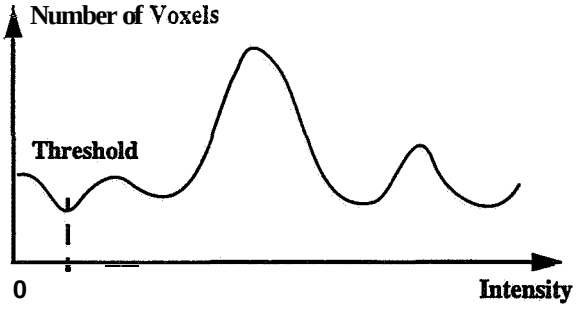


Figure 14. Thresholding based on the intensity histogram.

4.1.2 Histogram Normalization

We adopted an intensity-based method to avoid errors introduced by unreliable feature extraction processes, but inter-scan intensity variations make this approach problematic. The current solution is to perform histogram normalization on each volume prior to the registration process. To prevent outlier voxels with extreme intensities from stretching the histogram, we set the intensity of the darkest 2% of the voxels to 0, the intensity of the brightest 2% of the voxels to **255,** and then linearly scale voxel intensities in between to the range of 0 to **255.**

4.2. Efficient and Effective Processing

The volumes we deal with typically have **8** million voxels. Therefore it is imperative to carry out the registration in an efficient, yet accurate, manner,

4.2.1 Multi-resolution Processing

A multi-resolution strategy is used not only to improve efficiency, but also to help the optimization procedure to first focus on global patterns and gradually **shift to** the details. We employ two **kinds** of multi-resolution processing: an image pyramid (used in the ***global transformation*** and the ***smooth deformation),*** and a control grid pyramid (used in the ***smooth deformation).***

The image pyramid is a hierarchy of data volumes generated by successive subsampling. Currently it has 3 levels, each level being half the resolution of the next higher level. At the lower resolutions there **are** fewer details present, **so** the minimization process has less tendency **to** become trapped in local minima. By inheriting results from a lower resolution, the higher resolution registration can **start** closer to the global minimum. Optimization at the higher resolutions uses the finer details to refine the result.

The control grid pyramid is a hierarchy of different resolution control grids. Currently it has 5 levels, in which the numbers of control points along x, y, and z directions are **2x2x2, 3x3x3, 4x4x4, 5x5x5,** and 6x6x6. By starting with coarser control **grids** the deformation can focus on global patterns before plunging into details. The result from the lower resolution control grid **initializes** the registration with finer control **grids** close to the optimum. ***Smooth deformation*** with finer control grids allows for a more precise registration.

4.2.2 Stochastic Sampling

While estimating the ***global transformation*** and the ***smooth deformation,*** we do not process every voxel of the volumes during each iteration of the optimization process. Instead, we sample a random set of voxels at each iteration. **This** improvement in computation efficiency is possible because the optimizations **are** over-constrained. Moreover, the stochastic nature of the sampling helps the minimization process to escape from local minima [20]. Note **that** because the jine-tuning ***deformation*** **has** 3 times **as** many parameters **as** the number of voxels, the problem is highly underconstrained, **so** stochastic sampling is not appropriate.

4.23 Parallel Processing

In our approach, the computation at each voxel is identical, **so** the process is voxel-wise parallelizable. To reduce overhead, we employ parallel processing at a higher level than the voxel representation. During the registration via ***global transformation*** and during the fine-tuning ***deformation,*** we parallelize the processing of each cross-section of the volumes; whereas in the registration via ***smooth deformation,*** we parallelize the processing of each control grid cell. Because a voxel in the atlas volume can map to any position in the patient's volume, it is difficult to partition the volumes **so** each part can be processed independently, Therefore, we used a shared-memory multi-processor computer in which each processor has access to the full volumes.

5. ADORE

Based on the 3-D ***hierarchical deformable registration algorithm,*** we developed a prototype, ADORE (Anomaly Detection through REgistration). ADORE is designed **to** create a customized atlas for the patient, and use registration to indicate pathologies in the patient data. Experiments on actual medical data were carried out to evaluate the performance in matching and pathology indication.

5.1. Registration Performance

We conducted experiments on eleven MRI data sets. The performance was successful on five data sets, reasonable on four data sets, and poor on two data sets. Consistent satisfactory performance is impeded by two factors: one is that our approach is intensity-based, and suffers when the patient volume has a considerably different intensity distribution from the atlas; another factor is that if the patient data is coarsely sampled or only covers part of the head volume, it will not contain sufficient information for accurate registration.

The result of matching the atlas to one normal brain was presented in Section 3. Figure 15 shows the progressive results of the 3-D ***hierarchical deformable registration*** between the atlas and a brain with pathology. Note that the ***global transformation*** only addresses the extrinsic variations between the atlas and the patient using 3-D rotation, scaling, and translation. The atlas is rotated approximately **90** degrees from the coronal view to match the axial view of the patient. Labels of anatomical structures in the atlas can not align with the corresponding structures in the patient data. The intrinsic variations are adjusted roughly during the ***smooth deformation*** via 3-D warping of the control grid. Labels of anatomical structures in the atlas approximately align with the corresponding structures in the patient **data. The** deformation is refined using the jine-tuning ***defor-***

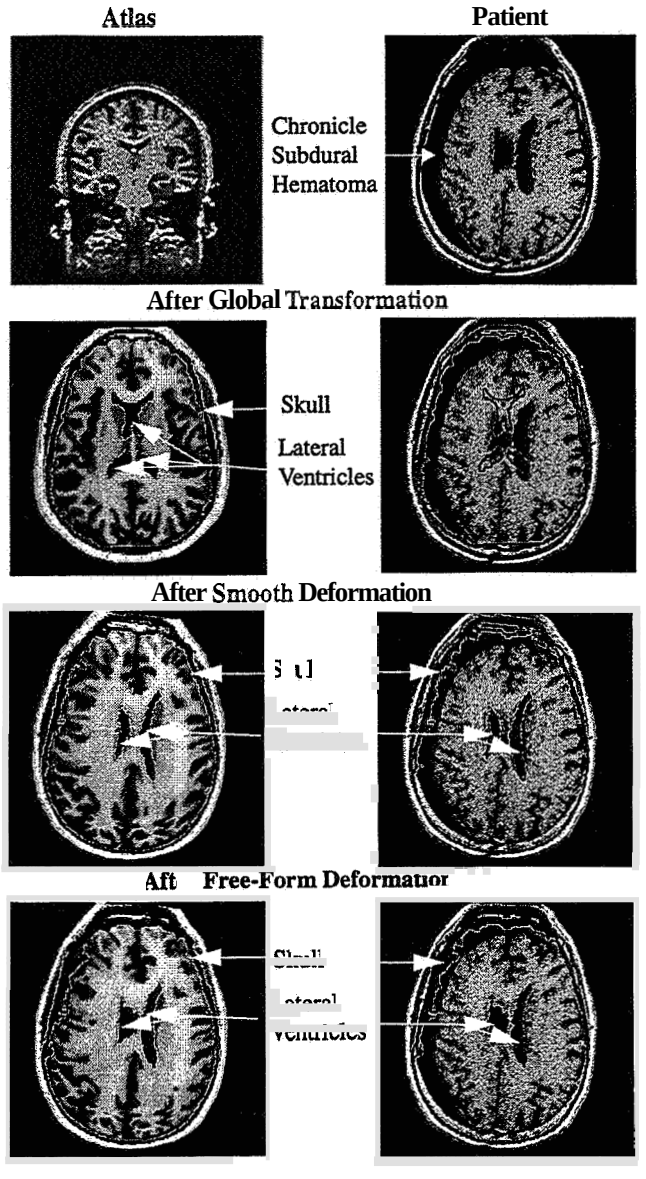


Figure 15. The progressive result of the **3-D** hierarchical deformable registration between the atlas and a patient with pathology. The atlas was warped in **3-D** to match the patient. The alignment of the corresponding anatomical structures improve along the registration hierarchy.

mation, where each voxel moves independently to align with its counterpart. Labels of anatomical structures in the atlas align well with the corresponding structures in the patient data. Figure 16 is a close-up on the registration of the lateral ventricles along the deformation hierarchy. The improvement of alignment accuracy is significant. The contours of the lateral ventricles appear to change during the registration process because the transformation and deformations move the **3-D** volumes with respect to the fixed cross-section we are observing.

5.2. b o d y Detection

It is of great clinical significance to create the customized atlases and to indicate pathologies for patients with neurological conditions. Certain pathologies cause mass effect, which shifts the brain

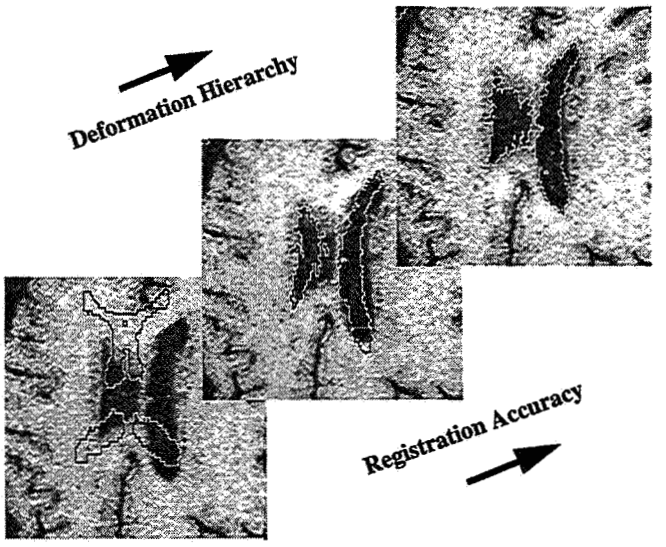


Figure 16. A close-up on the registration of the lateral ventricles at each stage in the registration hierarchy. The improvement in accuracy is significant.

structures, and causes anatomical features close by to deform significantly (see Figure 6). However, since our deformation process allows local deformations, it can still create the customized atlases with anatomical structures matching their counterparts in the patients' data, such as the case in Figure 15.

For the experiments on anomaly detection, we matched the patient's mirror volume (with left and right exchanged) to itself. Since a normal brain is approximately symmetric about the central line (visible from the axial and coronal view), not much deformation is needed to align its mirror volume to itself. A significant amount of deformation will indicate the absence of symmetry, and the possible existence of pathologies. In Figure **17,** the left column displays an axial cross-section of each patient's volume, the right column is the same cross-section overlaid with the deformation vectors of matching the patient's mirror volume to itself For the normal brain (first row), there is very little deformation, because it is approximately symmetric. For the brain with chronicle subdural hematoma (second row), there is significant and uniform deformation. This is because the pathology caused mass effect and destroyed the symmetry of the brain, forcing the anatomical structures (such as the lateral ventricles) to shift considerably The direction of the deformation vectors reveals the source of the mass effect, which is the possible location of the pathology For the brain with a tumor (third row), aside from the uniform deformation caused by the shifting of the lateral ventricles, there is a swirl pattern in the deformation vectors. This is because the tumor is so close to the central line that its "counterpart" in the mirror volume managed to shift back (partly) to match with itself. Note that unlike the other data sets, this is a CT scan. The characteristics of the deformation vectors combined with domain knowledge can provide indications of the existence and location of anomalies.

6. Evaluation and Analysis of Matching

It is difficult to quantitatively evaluate registration results in medical images, due to the lack of ground truth. We give a qualitative assessment of our algorithm in terms of simplicity, accuracy, and

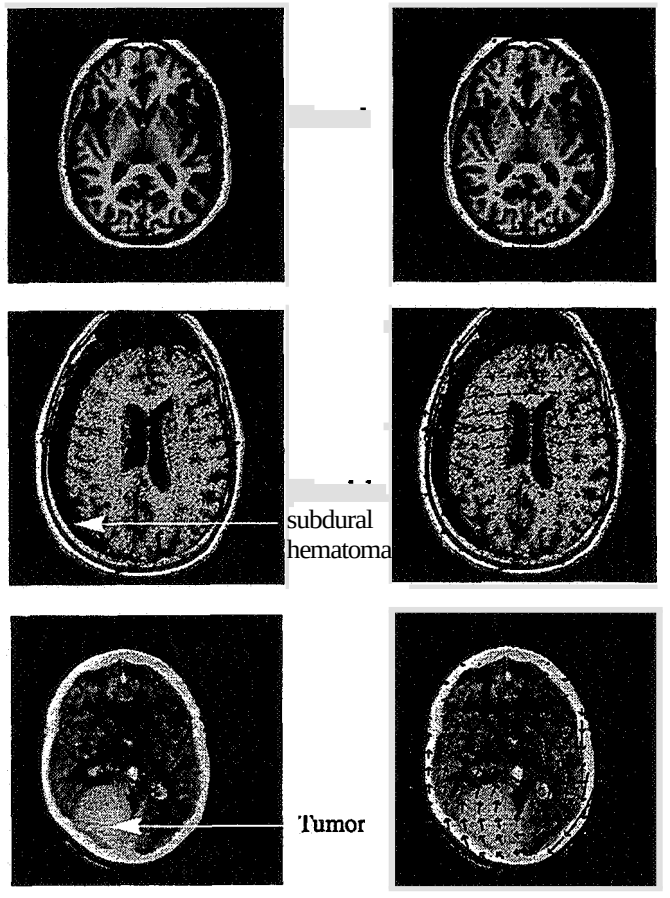


Figure 17 An axial cross-section of each patient volume (left), and the same cross-section overlaid with corresponding deformation vectors (right). The pattern, magnitude and direction of the deformation vectors indicate the possible existence and location of the pathologies.

speed.

6.1. Simplicity

Our current approach, from preprocessing, registration via ***global transformation,*** registration via ***smooth deformation,*** to registration via ***fine-tuning deformation,*** is completely automatic. The role of the user is to specify the principal scanning axis. This can be standardized so the user can simply select among different settings.

6.2. Speed

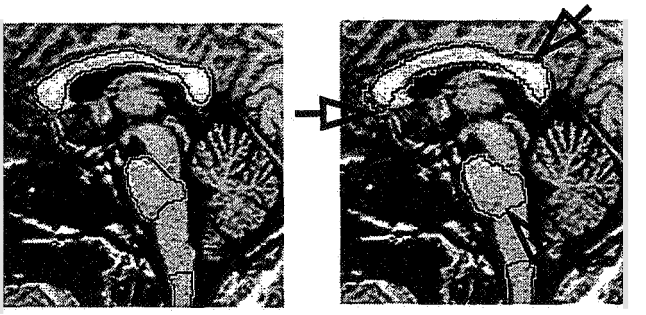
Currently it takes 12 minutes to match 256x256x124 volumes on an SGI computer with four 194 MHz R10000 processors. There are parameters that can be tuned to improve efficiency, such as the number of stochastic samples, the number of levels in the pyramids, and the number of control points.

6.3. Accuracy

Quantitative evaluation of a matching method is hard to perform due to the lack of ground truth. Figure 18 compares ADORE's segmentation and labelling of a patient volume with an expert's result, and it is noticeably less accurate at fine details. This is because ADORE adopts an intensity-based approach, but the same anatom-

ic str e in lif ent data sets can h te i

ADORE's Result



e .8. Labels from and gmentation
compared to the result of ADORE's

. Related Work

T e tr of medical images via optimization in transforma tion space has been an active research area--the comprehensive sur- article by van den Elsen et al [11] lists 5 citations The primary b of g c approaches is into ethod using e al and internal image properties matching, because di is important for the i t protocol. *l* e *properties* are introduced by artificial objects that are "added" to the patient, such as head frames or skin markers. *Internal image properties* are patient related characteristics of the image la e such as the intensity differences between anatomical t u s E ternal, g atio d have t u g e that any two modalities can be matched, as long as a marker can be constructed that is ta in ct o ies. They yield accuracy matching with respect to rigid transformations [4]. But to ensure accuracy, the markers must be g ly a to t e patient (by driving screws into the patient's skull), so as to indicate precise ly the patient's position and orientation. (viously, this is invasive and inconvenient.

g methods using internal image p p have the advantage of being non i and fully retrospective, which means that one does not need to know prior to the i of the g a will be required. They also have the potential to deal with non-rigid tr fo b e different image sets. This makes them more favorable for anatomical registration across individuals, where a rigid transformation ill no ffice.

Two u (e tr s i l q ti are *feature d* d *voxel-based. based* methods attempt to extra the anatomical structures in different data t , f the correspondences ween them. hey have the characteristic of l g i b s or and independent of imaging r t n -l t is critically l on the quality of the feature extraction, which is not trivial since an- l structures tend to have complex a and ill-defined d ri l a on is g erally necessary to lp select a extract features g l the t procedure. Consequently, it is bj to subjectivity, s and i

ar se use *mab models* for zsed *registration.* [gi c pl s r . he i l of the deformable must l ry close tc th ought feature to g a a successful s for l ly de- able els, therefore human intervention is generally neces

sary. The approach in [16] employs a 3-D elastic warping transformation to register 3-D images. The transformation is driven by an external force field defined on a number of distinct anatomical surfaces, which were acquired in an interactive manner.

As an alternative, *voxel-based* algorithms obviate the need for an explicit segmentation, although the representation is not as concise. The most intuitive voxel-based approach is based on voxel *intensities*. Bajcsy et al. develop a system that elastically deforms a 3D atlas to match anatomical brain images [9],[8]. The atlas is modelled as a physical object and is given elastic properties. Although their approach is similar to our current one, they assume the intrinsic variations between people can be modelled by an elastic deformation whereas we only enforce smooth deformation in an intermediate stage. Without user interaction, their atlas *can* have difficulty converging to complicated object boundaries. *Also* their method is more computationally expensive and requires interactive and time-consuming preprocessing.

Christensen et al. presented a method very close to ours, but they used a fluid dynamic model for the deformation [1], [2]. It constrains neighboring voxels to have similar deformations, while allowing large deformation for small sub-volumes. It takes 1.8 hours to match 128x128x100 volumes on a 16384-processor MasPar, while our algorithm takes 12 minutes to match 256x256x124 volumes on an SGI with four 194MHz R10000 processors.

In [10], Thirion takes a similar approach as ours, except that he assumes the volumes are already globally aligned, and he applies optical flow from the beginning. To reduce computation time, he used the gradient of the patient volume instead of the deformed atlas, because the computation of the latter is more expensive, requiring trilinear interpolation of each voxel's gradient. However, this quicker method can cause errors when the deformed atlas does not resemble the patient closely Because optical flow relies heavily on the constant brightness assumption, it is prone to failure when there are large intensity variations between different image sets.

Although voxel intensity-based approaches have shown encouraging results, they are problematic when there are intensity inhomogeneities. Moreover, they only work for multi-modal data if there exists a linear mapping between intensity values, which is unfortunately almost never the case. Viola and other researchers have investigated registration based on mutual information (MI) [18], [19], [20], [21]. MI is a basic concept from information theory, which measures the statistical dependence between two random variables, or the amount of information that one variable contains about the other, The MI registration criterion assumes that the statistical dependence between corresponding voxel intensities is maximal if the images are geometrically aligned. Because no assumptions are made regarding the nature of this dependence, the MI criterion is highly data independent and allows for robust and completely automatic registration. Current applications of MI to registration only perform rigid transformations to register image data of the same person from different modalities. The possibility of applying the MI criterion in deformable registration remains to be studied.

To date, most efforts are focused on exploring the information content in the images to achieve registration. Little work has been done in using domain knowledge *to* guide the process, and to tackle pathological cases which are of more clinical importance.

8. Research Directions

Our experiments demonstrated the strength and promise of the hierarchical deformable registration algorithm, and the automatic registration and pathology detection prototype, ADORE. However, for them to be applied to real medical practice, much research r e mains to be done.

8.1. Improve Registration Performance

Although the histogram normalization step can address considerable intensity discrepancies between data sets, the performance of our approach still decreases if the discrepancy is large. It may also fail to align data from different imaging modalities. Since mutual information can discern similar patterns despite differences in intensities, and has been successfully applied to multi-modality registration via rigid transformations, it is promising to apply this criterion to defcrmable registration.

8.2. Use Domain Knowledge as Guidance

Our current approach has no constraints on the deformations. If the deformation could be limited to shapes within the normal anatomical variation of structures, or the characteristic deformation caused by pathologies, potential local minima could be avoided

Knowledge of natural shape variability and pathology-afflicted deformation can be used to define allowable deformations, as well as to assist pathology detection. Figure 19 compares the corresponding cross-sections of the atlas after global transformation (left) and then local deformation (right) to match a patient. After deformation, the anatomical structure labelled as skull becomes thicker and uneven. Since this is out of the normal range of variation, it implies the possible existence of pathology. This applies to pathologies that only cause intensity variations as well. After matching the atlas to the data, the area of pathology will have an anatomical label. If the intensity distribution of the pathology does not conform to that of the labelled anatomical feature, it suggests an abnormality.

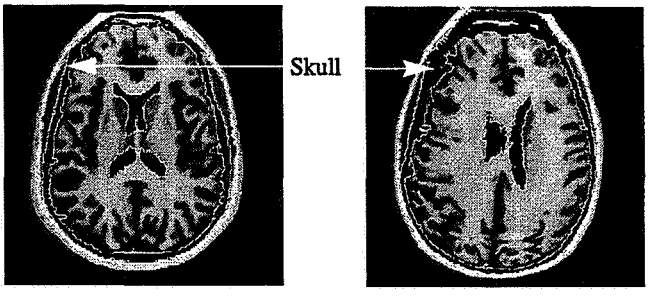


Figure 19. The corresponding cross-sections from the atlas after global transformation (left) and local deformation (right) to match a patient volume. Note that after deformation the region labelled as the skull becomes thicker and uneven.

Domain knowledge *can* be acquired from the statistics of training samples. The challenge is how to formulate a representation of the knowledge that can be incorporated into the registration procedure. For information described by a large number of possibly highly correlated parameters, principal component analysis (PCA) may offer a promising solution by characterizing the few largest eigendeformations, reducing the dimensionality substantially. To represent the information in a way that is appropriate for applying PCA, Fourier descriptors or modal representations have been examined in previous work [17], [19], [20].

Knowledge of anatomy, such as symmetry and relative positions of structures, can also provide guidance. Moreover, some anatomical

structures are more distinctive and therefore more easily registered. These structures should be given more weight in the estimation of the deformation.

83. Quantitative Evaluation

Quantitative evaluation of registration results is important because that will allow researchers to compare different approaches, and to assess the trade-off between computational efficiency and accuracy. One way to quantitatively evaluate the performance of ADORE would be to employ hand-labelled multiple atlases. Using one of the atlases as the patient, the registration process creates its customized atlas. The registration accuracy could be quantified by the fraction of voxels whose labels given by an expert agree with the labels in the customized atlas. We are currently investigating potential sources for additional atlases to make this type of evaluation possible.

9. Conclusion

The goal of this research is to perform automatic, fast and accurate registration of volumetric data in 3-D space, as well as anomaly detection using registration and domain knowledge. One important application domain is medical image registration. Anatomical structures vary considerably in appearance across individuals or within one individual over time, and any pathology may aggravate these variations.

We adopted an intensity-based approach for registration, which assumes no prior segmentation of the voxels in a volume. To address the appearance variations of anatomical structures, we have developed a three-level hierarchical deformable registration algorithm. First, a 3-D global *transformation* (rotation, uniform scaling, and translation) aligns the different image volumes. This procedure corrects for variations induced during the image acquisition process. Secondly, a smooth ***deformation,*** represented by the warping of a 3-D grid of control points, approximately matches the corresponding anatomical structures in the image volumes. This step partially adjusts for the inherent variations in the appearances of the anatomical structures across individuals. Several grid resolutions are used to progressively change the emphasis from global alignment to alignment of specific anatomical structures. Finally, a *fine-tuning &formation,* which allows each voxel to move independently, refines the matching of corresponding anatomical features. For computational efficiency, we use a hierarchy of image resolutions, stochastic sampling, and parallel processing.

Based on this algorithm, we have developed a prototype for automatic registration and pathology detection, ADORE (Anomaly Detection through REgistration). We conducted experiments on 12 sets of real image data of the brain. ADORE can match two 16MByte volumes in 12 minutes on an SGI workstation with four 194 MHz R10000 processors, with an accuracy qualitatively comparable to manual registration. By matching an expert-segmented atlas to a patient's data, ADORE can automatically build a customized atlas for different individuals. Moreover, ADORE is able to indicate the possible existence and location of pathologies by measuring the asymmetry of the anatomical features.

Acknowledgments

Many thanks to Shumeet Baluja, Carlos Ernesto Guestrin, John Hancock, Prem Janardhan, Andrew Johnson, Eric Rollins, and David Simon for their generous help and precious comments. The atlas data was obtained by Yanki Liu from the Brigham and Women's hospital of Harvard Medical School.

References

[1] Christensen *et al.*, "Individualizing Neuroanatomical Atlases Using A Massively Parallel Computer", IEEE Computer, pp. 32-38, January 1996.

[2] *Christensen et al.*. "Deformable Templates Using Large Deformation Kinematics", IEEE Transactions on Image Processing, September 1996.

[3] Horn, "Closed-form solution of absolute orientation using unit quaternions", Journal of the Optical Society of *America*, Vol. 4, No. 4, pp 629-642, April, 1987.

[4] Lavallee *et* al, "Image-Guided operating Robot: A Clinical Application in Stereotactic Neurosurgery", Computer-Integrated Surgery--Technology and Clinical Applications, Taylor *et* al., pp. 346-351.

[5] McInerney and Terzopoulos, "Deformable models in medical image analysis: a survey", Medical Image Analysis, Vol. 1, No. 2, pp 91-108, 1996.

[6] Szeliski, "Matching 3-D Anatomical Surfaces with Non-Rigid Deformations using Octree-Splines", International Journal of Computer Vision, Vol. 18, No. 2 pp 171-186.1996.

[7] Szeliski and Coughlan, "Spline-Based Image Registration", Digital Equipmat *Corporation*, Cambridge Research Lab, Technical Report 94/1.

[8] Bajcsy and Kovacic, "Multiresolution Elastic Matching", Computer Vision, Graphics, and Image Processing, Vol. *46,* pp 1-21, 1989.

[9] *Gee.* Reivich and Bajcsy, "Elastically Deforming 3D Atlas to Match *Ana*tomical Brain Images", Journal of Computer Assisted Tomography, Vol. 17, No. 2, pp 225-236, 1993.

[10] Jean-Philippe Thirion, "Fast Non-Rigid Matching of 3D Medical Images" INRIA, Technical Report No. 2547, May, 1995.

[11] van den Elsen, Pol, and Viergever, "Medical image Matching--A Review with Classification" IEEE Engineering in Medicine and Biology, Vol. 12, No. 1, pp. 26-39,1993.

[12] Antoine Maintz, van den Elsen, and Viergever, "Evaluation *of* Ridge Seeking Operators for Multimodality Medical Image Matching", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 4, April, 19%.

[13] Szekely *et* al., "Segmentation Of 2-D and 3-D Objects from MRI Volume Data Using Constrained Elastic Deformations of Flexible Fourier Contour and Surface Models" Medical Image Analysis, Vol. 1, No. 1, pp. 19-34, 19%.

[14] Sclaroff and Pentland, "On Modal Modelling for Medical Images: Underconstrained Shape Description and Data Compression" M.I.T. Media Laboratory Perceptual Computing Section Technical Report No. 275.

[15] Sclaroff and Pentland, "Modal Matching for Correspondence and Recognition" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 6, June, 1995.

[16] Moghaddam Nastar and Pentlaud, "A Bayesian Similarity Measure for Direct Image Matching". M.I.T Media *Laboratory* Perceptual Computing Section Technical Report No. 393.

[17] Pokrandt. "Fast Non-Supervised Matching: A probabilistic Approach"

[18] Pokrandt. Stein and Hassfeld, "Medical Image Registration: Fast Feature Space Evaluation with Gaussian Entropy Function"

[19] Wells III et al. "Multi-Modal Volume Registration by Maximization of Mutual Information", Medical Image Analysis, Vol. 1, No. 1, pp 35-51, 1996.

[20] Viola and Wells III, "Alignment by Maximization of Mutual Information", Proceedings of the fifth International Conference on Computer Vision, pp 16-23,1995.

[21] Maes *et* al.. "Multimodality Image Registration by Maximization of Mutual Information". IEEE Transactions on Medical Imaging, Vol. 16, No. 2, April, 1997.

[22] Press *et* al., "Numerical Recipes in C", Cambridge University Press, 1992.

On Matching Brain Volumes*

James C. Gee
Department of Neurology
University of Pennsylvania Medical Center
Philadelphia 19104, USA
gee@grip.cis.upenn.edu

Abstract

To characterize the complex morphological variations that occur naturally in human neuroanatomy so that their confounding effect can be minimized in the identification of brain structures in medical images, a computational framework has evolved in which individual anatomies are modeled as warped versions of a canonical representation of the anatomy, known as an atlas. To realize this framework, the method of elastic matching was invented for determining the spatial mapping between a 3-D image pair in which one image volume is modeled as an elastic continuum that is deformed to match the appearance of the second volume. In this paper, we review the primary concepts underlying the elastic matching, consider the implications of an integral formulation of the problem on its solution, and explore a more general Bayesian interpretation of the method in order to address issues that are otherwise difficult or unnatural to resolve within a continuum mechanical setting, including the examination of a solution's reliability or the incorporation of empirical information that may be available about the spatial mappings into an analysis.

1 Introduction

Relying on the knowledge that different instances of the same anatomy share a common topological scheme, Chaim Broit proposed in his 1981 dissertation that the various problems of localizing, measuring, or visualizing anatomic structures in a computed tomographic study could be reduced to the single problem of determining the spatial mapping between the anatomy depicted in the study and a reference, or *atlas,* image of the same anatomy [1]. Using this mapping, information encoded in the atlas—such as anatomic labels and structural boundaries—is transferred to the image study. The transformed atlas labels automatically localize their corresponding structures in the tomographic images. Visualization is achieved by deforming, according to the mapping, the structural boundaries contained in the atlas and then displaying the result. Quantitative measurements of any structure in the tomographic stiidy can be inferred from the morphometric information associated with its warped atlas version. If functional scans have also been acquired, the imaged activity can be related to anatomy, through the same mapping, once the subject's pose in these scans is made the same as in the computed tomographic study.

To realize this method of analyzing tomographic images, Broit invented a general registration technique in which one image volume is modeled as an elastic continuum that is deformed to match the appearance of the second volume. In this paper', we review the seminal concepts underlying Broit's elastic matching and then discuss their generalization within a probabilistic framework We introduce the probabilistic approach by first studying the integral formulation of elastic matching. The integral form reveals important possibilities, including the use of powerful numerical solution methods and the existence of a Bayesian decision-theoretic interpretation of the registration problem.

Bayesian analysis allows a fully model-based description of the elastic matching to be constructed and, as a result, can potentially resolve a number of issues that are otherwise difficult or unnatural to address within a continuum mechanical setting. These issues include, for example, the exploration of a solution's reliability or the incorporation of empirical information that may be available about the spatial transformations into an analysis. For computational anatomy, these aspects of the probabilistic approach will figure importantly in the development of a comprehensive methodology

2 Elastic Matching

In conceiving the elastic matching, Broit sought a registration method that would optimally balance the similarity it induced between an image pair with the amount of deformation, to the transformed image, it required to achieve the result Specifically, the solutions optimize the objective function $cost = deformation - similarity$. The method, therefore, allows for the fact that erroneous estimates can be produced by the procedure by which correspondences are established between points in the two images, as a result of,

*This work was supported by the U.S.P.H.S. under grant RO1-NS-33662

[1] In an effort to present in this paper a coherent picture of the various aspects of our work, the text draws from a great deal of previously reported material Also, to meet the manuscript length requirement, the paper focuses on our own work and we are forced, unfortunately, to make little reference to the excellent research in computational neuroanatomy that is ongoing at other institutions.

for example, noise in the images or because certain portions of an image lack sufficient features to distinguish them from neighboring regions. To help compensate for these errors, the mappings are presumed to be smooth, which explains the role of the deformation penalty in the problem formulation.

Broit drew an analogy between the elastic matching of two images and the physical process of applying forces to an elastic version of the object depicted in one of the images so that its deformed configuration resembles the target object in the other image. The appropriate forces are derived from a potential function that provides, at each point of the object under deformation, a measure of the similarity between that point and any of the points in the target object. The first object is deformed in this way until an equilibrium configuration is reached, where the total potential energy of the system will be at a local minimum.

Now if the strain energy were used as the measure of deformation in Broit's objective function, the cost would be identical to the forementioned total potential energy. Therefore, the equilibrium configurations, which locally minimize this cost, represent solutions to the corresponding elastic matching problem

2.1 Integral Formulation[1]

A body experiences strain when the relative position of its points are changed. The change in position of the points or *deformation*, effected by some motion $x = x(X, t)$, results in a new configuration of the body at time t, where $X = x(X, t_0)$ describes the original undeformed configuration. Points within the body may also be displaced without deformation: they are said to undergo *rigid* motion.

Consider the displacement of the material point P to position $x = X + u(X, t)$ in Figure 1. Continuum theory requires every mapping $x = x(X, t)$ to be one-to-one and its Jacobian determinant, $\det[\partial x_i / \partial X_j]$, everywhere greater than zero. These conditions ensure that the continuum is both indestructible and impenetrable—the topology of our atlas is therefore always preserved after deformation. Suppose the displacements also take a neighboring point Q at $X + dX$ to $x + dx = X + dX + u(X + dX, t)$. The relative position of P and Q becomes

$$dx = dX + u(X + dX, t) - u(X, t),$$

or

$$dx = dX + (u \overleftarrow{\nabla})\ dX,$$

where $u \overleftarrow{\nabla}$ is the displacement gradient. Note that $u \overleftarrow{\nabla} = 0$ implies $dx = dX$ or a rigid translation. $u \overleftarrow{\nabla}$ therefore incorporates both the rotational component of the rigid transformation and the pure deformation in the motion To factor out the deformation component, consider two infinitesimal fibers emanating from the same point P such that

$$dx^1 = dX^1 + (u \overleftarrow{\nabla})\ dX^1$$

and

$$dx^2 = dX^2 + (u \overleftarrow{\nabla})\ dX^2$$

Their scalar product provides the deformation measure we are after.

$$dx^1\ dx^2 = dX^1\ dX^2 + 2\, dX^1\ E\, dX^2,$$

[1] Although the notation is from [2, chapter 2], our overview of the kinematics closely follows the treatment found in [3].

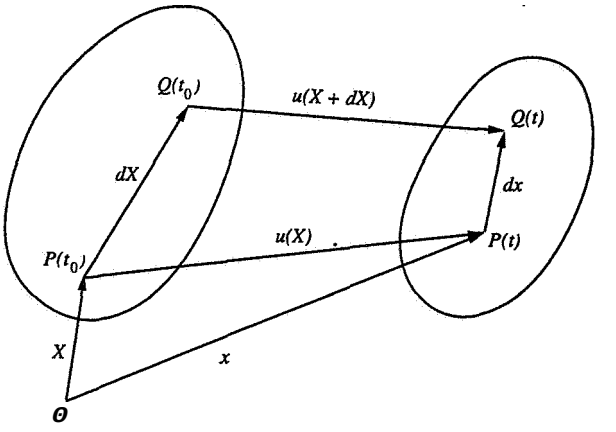


Figure 1 Deformation and motion of a continuum A particle P is identified with respect to the fixed origin O by its material coordinates $X = (X_1, X_2, X_3)$. At time t, the displaced position of P is specified by the equation $x = x(X, t)$ More generally, by varying X and therefore the particle under consideration, the equation defines the motion of the continuum The corresponding displacement is $u(X, t) = x(X, t) - X$, and its gradient incorporates information about the pure deformation in the motion

where $E = 1/2\{u \overleftarrow{\nabla} + (u \overleftarrow{\nabla})^T + (u \overleftarrow{\nabla})^T\ (u \overleftarrow{\nabla})\}$ is the Green or Lagrangian strain tensor and represents pure deformation. in the absence of strain, $dx^1\ dx^2$ is equal to $dX^1\ dX^2$, and the motion is rigid.

In the linearized theory adopted by Broit, the displacement gradients are assumed small. We drop the product $(u \overleftarrow{\nabla})^T (u \overleftarrow{\nabla})$ in E and obtain the infinitesimal strain tensor ε, whose components are given by the following expression;

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial X_j} + \frac{\partial u_j}{\partial X_i}\right)$$

The internal strain energy is the work performed by the conjugate stress σ or internal force that arises in response to the strain For isotropic isothermal elastic solids under small strain, Hooke's law reduces to the following linear relation, in indicial notation.

$$\sigma_{ij} = \lambda \varepsilon_{kk} \delta_{ij} + 2\mu \varepsilon_{ij}, \tag{1}$$

where λ and μ are known as the Lam6 constants [2]. The solution $\tilde{u}$ to the integral or variational formulation of the elastostatics problem is the deformation which induces a configuration with minimum potential energy $\pi(\tilde{u})$, where

$$\pi(u) = \frac{1}{2}\int_V \sigma : \varepsilon\, dV - \int_V b\ u\, dV - \int_S t\ u\, dS. \tag{2}$$

The first term measures the internal strain energy induced in the body volume V, the latter two terms, derived from the similarity potential, represent the external work performed by the body forces b and surface tractions, where the tractions t are distributed over surface area S.

Note that the strain energy compares with the first-order quadratic stabilizers used in standard regularization Moreover, the elastic constants are related to the regularization parameter by varying their values, we can modulate the stiffness of the body and thus the degree of smoothness in the deformations

2.2 Finite Element Analysis

To solve for $\tilde{u}$, only displacements of the form $u_h(x) = \alpha_i \phi_i(x)$ are considered from the set of kinematically admissible displacement fields. The use of a finite number n of trial functions ϕ_i to discretize the variational problem characterizes the Rayleigh-Ritz method. The next step is to substitute u_h into (2) The new potential expression is a function only of α_i and its minimization,

$$\frac{\partial \pi}{\partial \alpha_i} = 0, \quad i = 1, \quad , n,$$

yields a system of n equations for the n unknowns. Instead of solving the field equations at a finite set of points, the Rayleigh-Ritz method returns a continuous solution based on a finite set of functions.

In the finite element method [4], the ϕ_i are defined piecewise according to subdivisions of the problem domain (called finite elements): in addition, they assume simple forms-usually low-degree polynomials–over any element; must be sufficiently smooth to ensure the relevant integrands are computable; and satisfy the relation $u_h(x_i) = \alpha_i$. The unknowns a_i are thus associated with discrete nodal points x_i, at which the elements are connected.

A primary feature of the finite element bases is that the unknowns at points within any element are interpolated using only the nodal values of that element:

$$u = Nu^e, \tag{3}$$

where u^e is the set of nodal displacements of the element and the entries of N correspond to component parts of those basis functions with non-zero extent over the element. We can therefore write $\varepsilon = Bu^e$ within a given element, where B is composed of derivatives of N_i. The expression (1) for the stress tensor becomes, in matrix form, $\sigma = D\varepsilon = DBu^e$, over element e, D is the elasticity matrix. The total potential energy can then be rewritten in terms of its elemental contributions and its minimization calculated on an element-by-element basis:

$$\pi = \sum_e \pi^e = \sum_e \left\{ \frac{1}{2} \int_{V^e} [u^e]^T [B]^T DBu^e \, dV - \int_{V^e} [u^e]^T [N]^T b \, dV - \int_{S^e} [u^e]^T [N]^T t \, dS \right\},$$

$$\begin{aligned} \frac{\partial \pi^e}{\partial u^e} &= \frac{1}{2} \int_{V^e} [B]^T DBu^e \, dV - \int_{V^e} [N]^T b \, dV - \int_{S^e} [N]^T t \, dS \\ &= K^e u^e - f^e, \end{aligned} \tag{4}$$

where the stiffness and load characteristics of the continuous element have been 'lumped" at the nodal points through K^e and f'', respectively. Note that the nodal loads in f'' are unlike the pointwise forces used in finite difference methods. The decomposition into finite elements is motivated by the efficiency with which the element stiffness and load terms can be constructed and evaluated in practice. To solve for the nodal displacements, the expressions in (4) are assembled into a large sparse system of equations. Additional details about the finite element implementation can be found in [5]. In the next section, we consider a probabilistic interpretation of elastic matching by deriving Gibbs distributions from the individual terms in the expression (2) for the total potential energy.

3 Bayesian Generalization

Broit's elastic matching method innovated the physics-based approach to shape modeling that is now widely applied in image analysis. Its appeal and success derive from the available theory and numerical techniques for the continuum mechanics, as well as from the intuitive nature of the method. On the other hand, the analysis in terms of mechanical systems is not the most general In this section, we reformulate the image matching problem within a decision-theoretic framework based on Bayesian modeling and demonstrate some of its important features, including the computation of reliability and interval estimates and the incorporation of empirical information about the spatial mappings into the analysis.

The Bayesian paradigm associates with each model component of the problem a probability distribution for the model parameters, which quantifies the reliability of their values, and propagates this information to bear on subsequent analyses. To illustrate the basic ideas, we begin with a general definition of the image matching problem: find x and f such that

$$I_A(X) = f(I_S(x(X)), \tag{5}$$

where the spatial transformation x specifies for each point X in I_A its corresponding point $x(X)$ in I_S, and f models the compound effect of processes which modify the values of the original intensity signal [6].

In our application, mappings are sought between images that arise from distinct but topologically similar sources, so (5) does not strictly apply—the success of atlas matching therefore depends both on the degree to which two anatomies actually bear -topological resemblance and on the level of localization accuracy required by the analysis. Nevertheless, it is possible to identify in the images corresponding features whose registration will yield the desired mappings. By definition, each particular form of correspondence has associated with it a metric for quantifying the degree of similarity between two feature instances.

As an example, the following is a simple instance of (5) that is commonly assumed for magnetic resonance imaging (MRI) scans, after the original images have been suitably preprocessed.

$$I_A(X) = I_S(x(X)) + \mathcal{N}(\mu, \sigma^2), \tag{6}$$

where the noise process is additive and ruled by the same known Gaussian distribution at each pixel. In classical maximum likelihood (ML) estimation, (6) leads to the following *likelihood function:*

$$p(I_S, I_A, x) \propto \exp - \left\{ \frac{1}{2\sigma^2} \sum (I_A(X) - I_S(x(X)) - \mu)^2 \right\},$$

where the summation is performed over the image domain; I_S and I_A are realizations of the random fields modeling the image pair; and the spatial mapping x between them is treated as the unknown vector parameter of the distribution[3] For the imaging situation specified by (6), the (exponent of the) likelihood function above defines the pertinent criteria for rating the similarity induced between the observed images by any value of x In particular, the mapping x^* which maximizes $p(I_S, I_A, x)$ is the the ML estimate.

Note that the estimate in effect represents a *measurement* of the true disparity between the image pair By examining the shape of the likelihood function, an error may be

[3] The observations at the pixels are assumed independent

associated with each disparity estimate. The error is a consequence of the uncertainty in the observations on which the results are based.

3.1 Prior Information

For atlas matching, the estimated mapping between the images should not only have large probability as measured by the likelihood but must at some level be homeomorphic so that topological information about one image, the atlas, is preserved when transferred to interpret the other Even so, these requirements insufficiently constrain the problem because the image features on which matching is based are usually sparse, as, for instance, in anatomic scans where the strongest features lie primarily along the tissue interfaces. By viewing the images as physical continua in the elastic matching, the associated strain energy or deformation measure essentially favors smooth mappings and so acts to *regularize* the matching problem.

The use of such *prior* information is a fundamental element of Bayesian methodology, and is introduced into the analysis by modeling the unknowns as random variables—the particular mapping that we are after is considered to be a realization of the random field $\boldsymbol{x}$or, more conveniently, $\boldsymbol{u}(\boldsymbol{X}) = \boldsymbol{x}(\boldsymbol{X}) - \boldsymbol{X}$

The prior knowledge is then expressed as a probability distribution on the space of admissible mappings. The distribution describes our certainty about the unknown relation between any image pair before any data has been collected

For example, the presumption in elastic matching that the mappings should satisfy the governing equations of a linear elastic body can be encoded through a Gibbs prior, where the potential is just the internal strain energy of the system. More significant, the prior allows domain knowledge specific to the problem to inform its analysis. In computational anatomy, instead of or in addition to assuming certain general mechanical properties, the analysis can be performed using actual statistics of the anatomic variation observed in a population.

To illustrate, consider the plots, depicted in figure 2, of the maximum eigenvalue of the pointwise displacement variance for the displacements, $\boldsymbol{u}^{(i)}, i = 1, \quad ,k$, from a referential anatomy to a group of 8 subjects. The variance

$$var(k) = 1/k[\boldsymbol{u}(1) \quad \boldsymbol{u}(k)][\boldsymbol{u}(1)^{\mathrm{T}} \quad \boldsymbol{u}(k)^{\mathrm{T}}]^{\mathrm{T}} - \bar{\boldsymbol{u}}(k)\bar{\boldsymbol{u}}(k)^{\mathrm{T}}$$

computed in the Cartesian space will be singular because the number of observed mappings will be small in comparison with the dimensionality of the mappings. In order to build distributions based on empirically derived variance information, the statistics must be represented in a subspace of the original Cartesian space. In particular, the subspace should grow in dimension as additional observations are introduced, and its basis must permit recursive updating. Assume such a basis is available—details of its construction can be found in [7, 8]. $var_V^{(k)}$ remains positive semidefinite. Consider instead the eigendecomposition.

$$var_V^{(k)} = \mathcal{G}^{(k)}\Lambda^{(k)}\mathcal{G}^{(k)-1},$$

where $\mathcal{G}^{(k)} = [\boldsymbol{g}_1^{(k)} \quad \boldsymbol{g}_r^{(k)}]$ are the eigenvectors of $var_V^{(k)}$ expressed in the orthonormal basis $V = \{\boldsymbol{v}_1, \quad ,\boldsymbol{v}_r\}$, and $\Lambda^{(k)}$ denotes the diagonal matrix of eigenvalues. Figure 3 shows the first two principal modes of the eigendecomposition of the displacements obtained for the subject group referenced in figure 2.

If there exists an null eigenvalue in $\Lambda^{(k)}$, it is removed; the corresponding eigenvector is similarly deleted from $\mathcal{G}^{(k)}$ In Cartesian space, $\{\boldsymbol{g}_1^{(k)} \quad \boldsymbol{g}_r^{(k)}\}$ forms a new orthonormal basis of the subspace spanning the observed mappings:

$$[\boldsymbol{e}_1^{(k)} \quad \boldsymbol{e}_r^{(k)}] = [\boldsymbol{v}_1 \quad \boldsymbol{v}_r][\boldsymbol{g}_1^{(k)} \quad \boldsymbol{g}_r^{(k)}],$$

and its subset $\{\boldsymbol{e}_1^{(k)} \quad \boldsymbol{e}_{r-1}^{(k)}\}$ spans the space of the displacement variance:

$$var_{\mathcal{E}}^{(k)} = [\boldsymbol{e}_1^{(k)} \quad \boldsymbol{e}_{r-1}^{(k)}]\Lambda^{(k)}[\boldsymbol{e}_1^{(k)\mathrm{T}} \quad \boldsymbol{e}_{r-1}^{(k)\ \mathrm{T}}]^{\mathrm{T}}$$

The mean $\bar{\boldsymbol{w}}^{(k)}$, expressed in the basis $\{\boldsymbol{g}_1^{(k)} \quad \boldsymbol{g}_r^{(k)}\}$, becomes

$$\bar{\boldsymbol{z}}^{(k)} = [\boldsymbol{g}_1^{(k)\mathrm{T}} \quad \boldsymbol{g}_r^{(k)\mathrm{T}}]^{\mathrm{T}}\bar{\boldsymbol{w}}^{(k)},$$

or, in the Cartesian space, is derived as follows:

$$\bar{\boldsymbol{u}}^{(k)} = [\boldsymbol{e}_1^{(k)} \quad \boldsymbol{e}_r^{(k)}]\bar{\boldsymbol{z}}^{(k)}$$

The mean $\bar{\boldsymbol{z}}$ and variance $\mathbf{A}$ derived from the observations specify completely a Gaussian prior distribution for the unknown displacement field:

$$p(\bar{\boldsymbol{z}}) \propto \exp -\left\{\sum_{i=1}^{r-1}\frac{1}{2\lambda_i}(z_i - \bar{z}_i)^2\right\} \propto \exp -\left\{\frac{1}{2}\boldsymbol{\eta}^{\mathrm{T}}\Lambda^{-1}\boldsymbol{\eta}\right\},$$

where $\boldsymbol{\eta} = z - \bar{z}$, and $\boldsymbol{u} = \bar{\boldsymbol{u}} + [\boldsymbol{e}\mathbf{l} \quad \boldsymbol{e}_{r-1}]\boldsymbol{\eta}$

3.2 Likelihood Model

The Bayesian view of $\boldsymbol{u}$ as a random field alters the observation model accordingly, the likelihood is now formulated as the conditional probability $p(I_S, I_A \mid u)$ of observing the images given any particular value of the unknown mapping between them.

3.3 Posterior Analysis

Our prior beliefs about the unknown mapping are revised when observations are available From Bayes's law, the conditional density

$$p(\boldsymbol{u} \mid I_S, I_A) \propto p(I_S, I_A \mid \boldsymbol{u})\, p(\boldsymbol{u})$$

describes the ***posterior*** *distributzon* for u Bayesian analysis proceeds from the posterior and thus utilizes all the information that is available about the unknown mapping.

To perform matching, point estimates of u are inferred by minimizing an expected *loss* with respect to the posterior distribution. In fact, solutions obtained with Broit's elastic matching are estimates of the posterior mode and hence correspond to optimal actions under a zero-one loss function Another commonly cited point summary of the posterior is its mean, which optimizes the quadratic **loss.** The specification of the loss function is a formal aspect of Bayesian decision theory and requires consideration of the cost incurred for making a particular decision given some value of the unknown parameters. Some of the practical implications of computing Bayes actions for the zero-one and quadratic loss functions in brain image matching have been explored in [9].

In Bayesian analysis, a solution can be further characterized by its reliability so as to reveal the influence of the prior on the solution as well as the uncertainty in the observations. The calculation of the second order statistics is

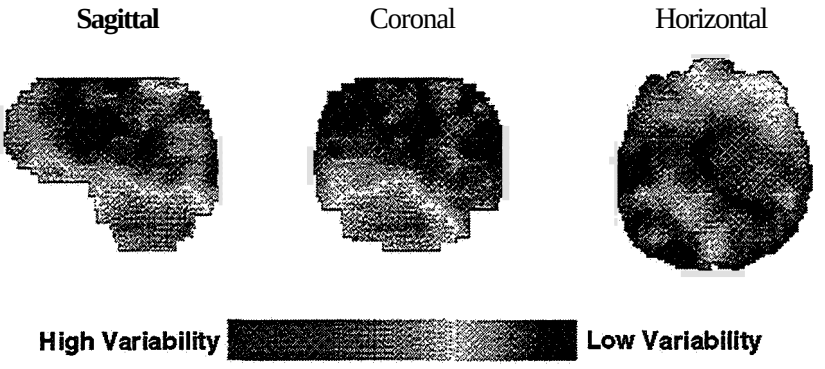


Figure **2:** Anatomic variation in a group of 8 subjects, measured by warping each individual into alignment with a referential anatomy after an initial affine normalization was performed. The variability is depicted using three orthogonal sections through the reference space, where the intensity value at a point corresponds to the maximum eigenvalue of the **3-D** displacement variance of that point from the reference to the subject population.

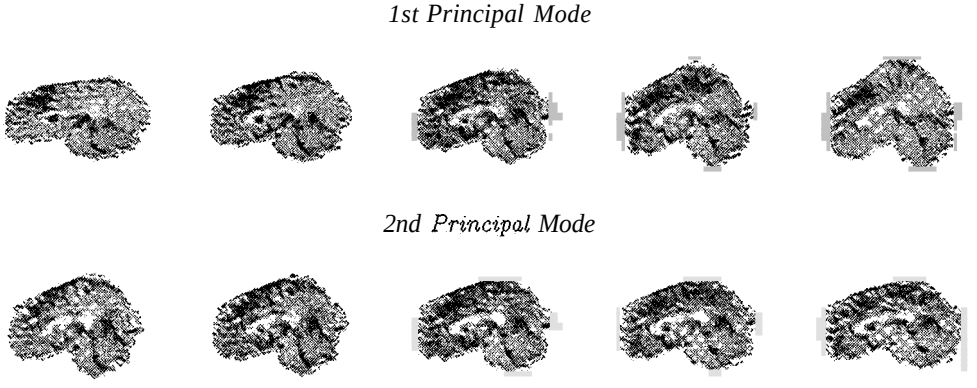


Figure **3:** Eigendecompositions of the observed mappings derived from normalizing a group of subjects with respect to a referential anatomy. The morphological variation captured by each principal mode is visualized by deforming the referenti anatomy using the mappings specified by the mode and its magnitude-scaled versions.

made tractable in our implementation by considering only the covariance between the components of the estimated displacement vector at each point and by sampling a Gaussian approximation to the posterior at the solution **[10]**

The existence **of** the covariance information admits another possibility a confidence region or *Bayesian credible interval* may be developed about an estimate. The result is an interval of possible displacement values for each point. **In** atlas-based localiiation, each structure can therefore **assume** a range of plausible shapes. An example is shown **in** figure **4,** where **we** have additionally visualized the covariance on which the credible interval is based. Note the underlying tissue interfaces in the covariance plot and the narrowing of the interval estimate along the edges of the subcortical structures that line the ventricles.

4 Applications

4.1 Atlas Matching

From the outset our interest in image matching **has** been in its use to localize anatomy in cerebral scans of a subject by aligning an atlas with the subject's **brain** volume. Figure 5 shows one such example. **In** this case, the atlas **was** based on the MRI scan of another individual, and its expert-defined anatomic labels were mapped to the subject's MRI image, using the transformation obtained by matching the two MRI scans. The volume dimensions were **112 x 96 x 88** for the highest spatial resolution at which matching was performed In comparison, only **14,794** nodes were needed in the corresponding finite element discretization to obtain the results shown

4.2 Spatial Normalization for Functional Analysis

The capability of quantifying anatomic variations with respect to some reference space implies that these same m i - ations may be removed in the analysis of group data. **In**deed, spatial normalization of this kind is prerequisite to most imaging studies of brain activation, in which the task or stimulus induced activity are presumed to be localized to specific regions of the brain. The spatial resolution at which the brain structures can be aligned limits the highest resolution at which comparison can be made across a group

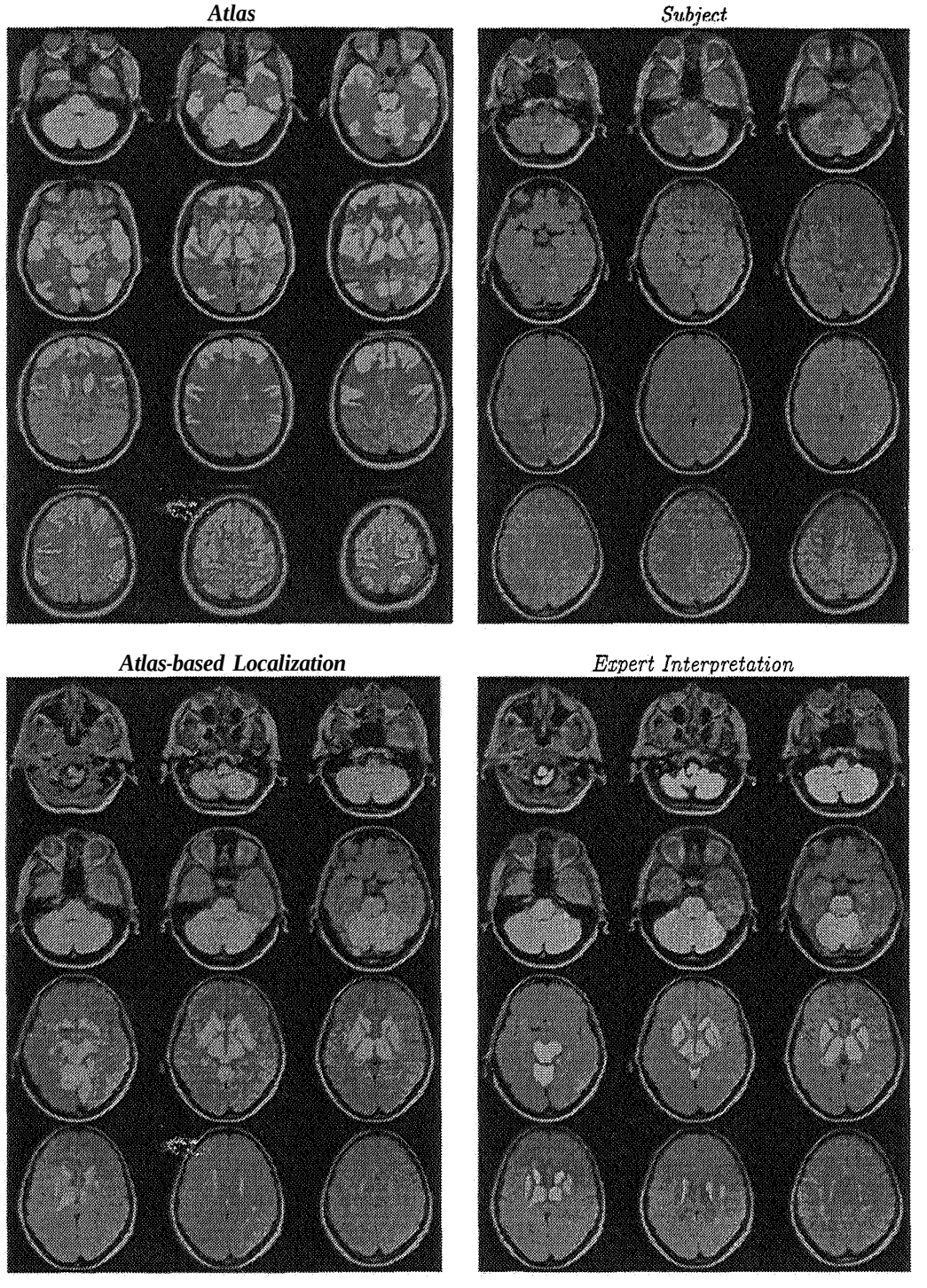


Figure *5* MRI-based atlas localization of neuroanatomy. Anatomic labels, superimposed in **green** in the figure, encoded in the atlas (top left) axe mapped to a subject's MRI study (top right), using the transformation obtained by deforming the underlying MRI volume of the atlas to match the appearance of the subject's MRI The resultant anatomic localization (bottom left), superimposed in **green** over the subject, is compared with an expert's interpretation, shown in yellow, (bottom right)

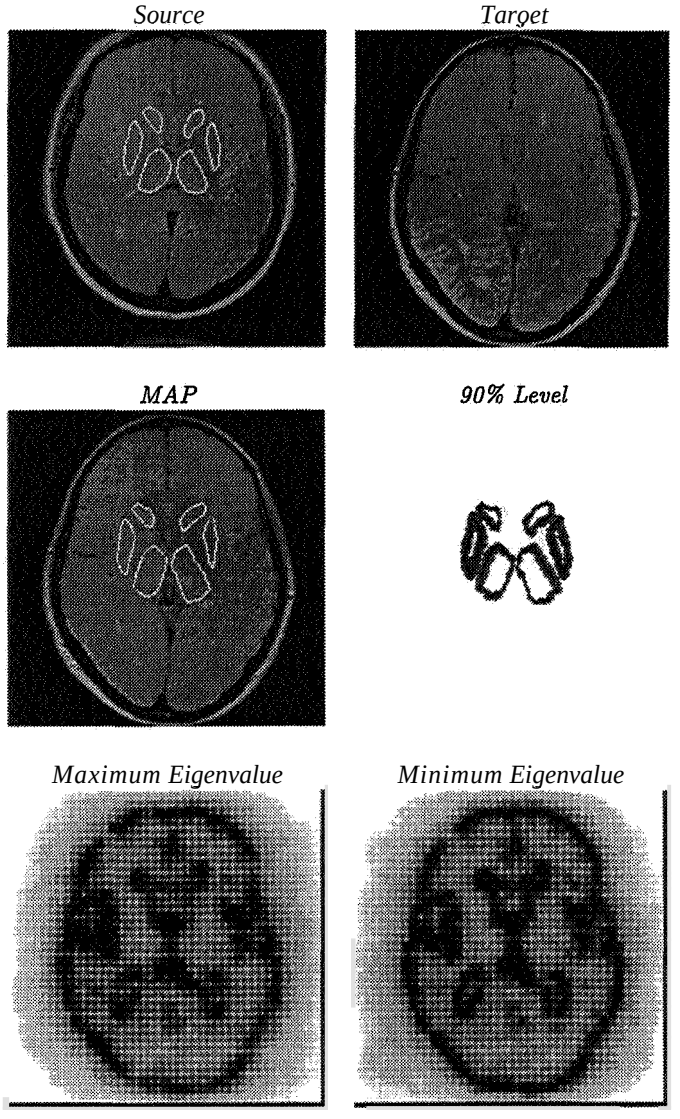


Figure **4:** Reliability and interval estimates about a **MAP** solution. Top row Source image was matched to the target; expert-defined contours of several subcortical structures on the source image were then transferred using the maximum a posteriori **(MAP)** estimate of the mapping. Middle row: 90% confidence interval (right) about the **MAP** segmentation (left); the contours outline the transformed structures of the source, while the expert-defined interpretation is shown **as** solid regions. Bottom **row:** Covariance of the estimated mapping.

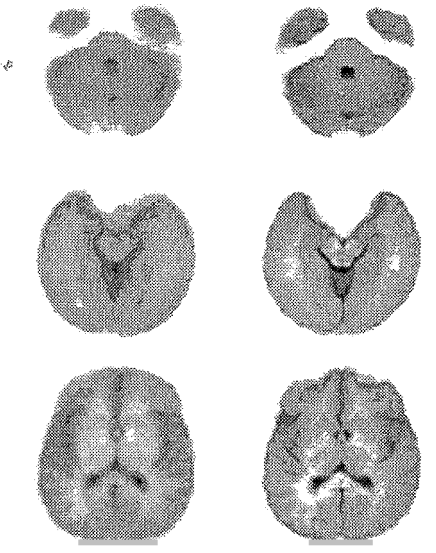


Figure *6:* Comparison of the spatial normalizations, averaged over a subject group. For each alignment method, the normalized results for the entire group were averaged together Each row in the figure depicts a different section from the mean images. The high dimensional warps estimated by elastic matching produce a uniformly sharper mean image, reflecting a more accurate alignment of the tissue interfaces within the subject anatomies.

of subjects to determine average functional responses or the variation of response in the population. The comparison, shown in figures **6–7** and detailed in [11], of mean activation **as** a function of the alignment method highlights the necessity of transformations that are very high in dimension so **as** to accommodate the complex anatomic variations that naturally exist among individuals, **as** exemplified in figure **2.**

5 Summary

We examined the problem of image matching and its application to anatomic localization in images of the brain, introducing in the process an improved finite element implementation and more general Bayesian interpretation of elastic matching.

In the finite element analysis of elastic matching, the solution is approximated by a combination ***of*** functions defined piecewise to fit the geometry of the problem domain. For anatomy, parsimonious model descriptions are therefore possible, which would substantially simplify the matching calculation

We illustrated the role of elasticity in expressing our prior expectations about the nature of the anatomic variations among the subjects, and constructed a posterior model for elastic matching, from which estimates other than the **MAP** solution, including their reliability, can be derived. Furthermore, a method was described by which Bayesian analysis can potentially **address** the empirical modeling of anatomic variation and its application to guide machine analyses of individual data sets.

Finally, a word is required about the case when the local

Homogeneous Affine *Elastic Matching*

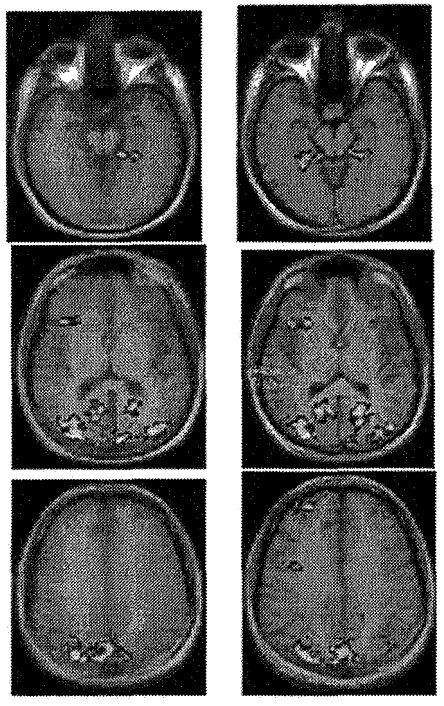


Figure 7 Comparison of the effect of spatial normalization on the functional MRI activation images. Mean activation was higher in amplitude and qualitatively more spatially compact with the high dimensional elastic matching than with the conventional 9-parameter affine registration

variations in anatomy are very large in magnitude. In this situation, plastic or even fluid characterizations of the images may be advantageous, as large deformations are made more probable than in the elastic theory [12]. On the other hand, their implementation requires extraordinary computational resources because of the nonlinearity introduced through the kinematic variables and the constitutive relations. A more fundamental difficulty with large displacements is that the likelihood of false matches increases. A standard approach is to solve the problem at different spatial scales, as in the multi-resolution version of elastic matching described in [13]: large scale displacements are first determined by matching the lower spatial frequencies in the images and these are then used to remove their confounding effect in the alignment of the higher frequency content. When even the multi-resolution search fails, additional external knowledge must be introduced into the problem formulation.

References

[1] C. Broit, *Optimal Regzstratzon of Deformed Images*, Ph.D. thesis, Department of Computer and Information Science, University of Pennsylvania, Philadelphia, 1981

[2] L. E. Malvern, *Introduction to the Mechanics of a Continuous Medium*, Prentice-Hall, Englewood Cliffs, 1969

[3] W M Lai, D. Rubin, and E. Krempl, *Introduction to Continuum Mechanics*, Pergamon Press, New York, 1978.

[4] K.-J Bathe, *Finite Element Procedures in Engineering Analysis*, Prentice-Hall, Englewood Cliffs, 1982.

[5] J C. Gee, D R. Haynor, M Reivich, and R. Bajcsy, "Finite element approach to warping of brain images," in *Medical Imaging 1994 Image Processzng*, M H Loew, Ed., Bellingham, 1994, SPIE.

[6] L. Brown, "A survey of image registration techniques," *ACM Comput. Surveys*, vol. 24, pp. 325-376, 1992.

[7] L. Le Briquer and J C Gee, "Design of a statistical model of brain shape," in *Information Processing in Medical Imaging*, J S. Duncan and G Gindi, Eds., pp. 477–482. Springer-Verlag, Heidelberg, 1997

[8] J. C Gee and L. Le Briquer, "An empirical model of brain shape," in *Maximum Entropy and Bayeszan Methods*, G. Erickson, Ed. Kluwer Academic, Dordrecht, 1997

[9] J C. Gee, L. Le Briquer, C Barillot, D. R Haynor, and R. Bajcsy, "Bayesian approach to the brain image matching problem," in *Medical Imaging 1995. Image Processing*, M H. Loew, Ed., Bellingham, 1995, SPIE, pp. 145-156.

[10] J C Gee, L. Le Briquer, C. Barillot, and D R. Haynor, "Probabilistic matching of brain images," in *Informatzon Processzng in Medical Imaging*, Y Bizais, C. Barillot, and R Di Paola, Eds., pp. 113-125. Kluwer Academic, Dordrecht, 1995.

[11] J C Gee, D C Alsop, and G. K Aguirre, "Effect of spatial normalization on analysis of functional data," in *Medical Imaging 1997 Image Processzng*, K M Hanson, Ed., Bellingham, 1997, SPIE, pp. 550–560.

[12] G. E. Christensen, R. D. Rabbitt, and M. I Miller, "Deformable templates using large deformation kinematics," *IEEE Trans. Image Process.*, vol. 5, no. 9, 1996.

[13] R. Bajcsy and S. Kovačič, "Multiresolution elastic matching," *Comput. Vision, Graphics, Image Process.*, vol. 46, pp. 1–21, 1989.

Session VII

Theory and General Methods

REGISTRATION OF DEFORMED IMAGES

Martin Satter
Nuclear Medicine/PET
Kettering Medical Center
Kettering, OH 45429

Ardeshir Goshtasby
Computer Science and Engineering
Wright State University
Dayton, OH 45435

Abstract

Determination of transformataon functions for registration of deformed amages as discussed. Each component of a transformation as viewed as a surface, and by surface-fittang, transformations are determaned. Viewing the components of a transformation as surfaces enables detection of wrong homologous points and makes it possible to estzmate the geometry of the scene. Based on the ideas presented an this paper, a coarse-to-fine method as developed that can register images wath local geometrac differences, At the coarsest resolution, the images are aligned using planes as the components of the transformation. At mad resolutions, surfaces with radial basis functaons are used to register the images, and at the highest resolution, elastic surfaces are used to accommodate large local geometrac differences between the amages. The results of the proposed registrataon on a variety of amages are presented.

1 Introduction

Image registration is the process of establishmg correspondence between all points in two image of the same scene. This process is required in many image analysis problems that involve multiple images of the same scene. When images of a scene obtained at different times are registered, it becomes possible to determine changes in the scene between the times the images are obtained This process is called change detection When images from different sensors are combined, it becomes possible to fuse multimodality images and reduce image ambiguities. This process is called sensor fusion When images from different views of a scene are registered, it becomes possible to estimate the geometry of the scene. This process is called depth perception

To register two images, first a number of homologous points are determined in the images. Then, using the homologous points, a transformation function is determined to establish correspondence between all points in the images, The most difficult step in image registration is the determination of the initial homologous points. Difficulty arises when different sensors are used to obtain the images, when the images are not captured from the same viewpoint of the scene, and when changes occur in the scene between the times the images are obtained. If the images do not have intensity and/or geometric differences, they can be registered without much difficulty [3] Image registration, however, is not of much value if the images do not have intensity and/or geometric differences. It is these differences that make change detection, sensor fusion, and depth perception possible.

When the images have geometric and/or intensity differences, automatic determination of homologous points becomes difficult Therefore, often user interactions are allowed to either initialize the homologous points or evaluate obtained homologous points and eliminate the wrong ones.

The second step in image registration is the determination of a transformation function that will take the coordinates of a point in one image and produce the coordinates of its conjugate point in the other image If geometric difference between the images are known, then the right transformation functions for image registration can be selected. For instance, satellite images of a relatively flat area can be registered with an affine transformation If the images have nonlinear geometric differences due to lens imperfection, and the characteristics of the lens are known, the images can be geometrically corrected before registration [7], or the lens nonlinearity can be incorporated into the registration process. Most often, however, geometric differences between the images are not known, and one has to select a transformation function that can adapt to local geometric differences between the images. When nonlinear geometric differences between images does not exist, the transformation function should not introduce non-

linear distortions to the resampled image; and when there is a geometric distortion in one area in an image, it should not propagate the distortion to other areas in the resampled image, Ideally, a transformation function should be able to represent local as well as global geometric differences between the images.

In this paper, our focus will be on the second step of image registration; that is, the determination of a transformation function that will resample one of the images to overlay with the other We will also discuss the use of properties of transformation functions in locating and removing the wrong homologous points.

We will call one of the images the ***reference*** and the other image the ***sensed.*** The reference image will be our base image and will be kept unchanged. The sensed image will be the one that will be resampled using the obtained transformation function to overlay with the reference image. Coordinates of points in the reference image will be denoted by (x, y), and coordinates of points in the sensed image will be denoted by (X, Y). We will first show how a locally sensitive transformation function can be determined from a number of homologous points in the images. Then, we will show how the incorrect homologous points can be identified and how the geometry of the scene can be guessed using an obtained transformation function. Next, we will describe a coarse-to-fine image registration approach that uses some of the ideas presented in the paper Finally, we will present registration results on a variety of images.

2 Locally Adaptive Transformation Functions

Given a set of n homologous points in the images, $\{(x_i, y_i), (X_i, Y_i) \quad i = 1, \quad , n\}$, we would like to determine a transformation function $\mathbf{f}$ with components f_x and f_y that satisfies

$$X_i = f_x(x_i, y_i) \tag{1}$$
$$Y_i = f_y(x_i, y_i), \qquad i = 1, \quad , n \tag{2}$$

Once f_x and f_y axe determined, for each point (x, y) in the reference image, we can determine its conjugate point (X, Y) in the sensed image, and the intensity in the sensed image can be resampled and saved in the reference image (or in an image with the same geometry as the reference image). For images with only translational (t_x, t_y), rotational (e), and scaling (s) differences, f_x and f_y can be written as

$$X = sx \cos\theta - sy \sin\theta + t_x \tag{3}$$

$$Y = sx \sin\theta + sy \cos\theta + t_y \tag{4}$$

For simplicity, equations (3) and (4) are often replaced by an affine transformation-

$$X = ax + by + c \tag{5}$$
$$Y = dx + ey + f \tag{6}$$

These functions may be used to register images obtained from about the same viewpoint of a scene with the distance of the camera to the scene being much larger than variations in scene elevation. In such images, if three or more homologous points are determined, parameters $a - f$ of the transformation can be determined and the images can be registered The affine transformation, however, can be used only when parallel lines in the scene remain parallel in both images.

If the images are from a flat area but there is a large viewpoint difference between the cameras, a projective transformation should be used to register the images [15] A projective transformation is defined by

$$X = \frac{ax+by+c}{dx+ey+1} \tag{7}$$
$$Y = \frac{fx+gy+h}{dx+ey+1} \tag{8}$$

Knowing the coordinates of at least four homologous points in the images, we can determine parameters $a - h$ The projective transformation requires that straight lines in the scene remain straight in both images. In other words, the scene should be flat and image distortions due to lens nonlinearities should not exist. In practice, however, cameras are not ideal and scenes are often nonplanar In addition, lens nonlinearities and geometry of the scene may not be known To register such images, a transformation function that uses information hidden in the homologous points is needed to properly model geometric differences between the images.

To find geometric differences between two images, we will view image registration as a surface-fitting problem. If we rearrange the coordinates of the given n homologous points, $\{(x_i, y_i), (X_i, Y_i) \quad i = 1, \quad , n\}$, into two sets:

$$\{(x_i, y_i, X_i) \quad i = 1, \quad , n\} \tag{9}$$
$$\{(x_i, y_i, Y_i) \quad i = 1, \quad , n\}, \tag{10}$$

we can then consider f_x and f_y as single-valued surfaces interpolating the 3-D point sets given in (9) and (10), respectively From equations (5) and (6), we see that

if the images have only linear geometric differences, f_x and f_y will be planes Otherwise, they will be surfaces containing hills and valleys that correspond to local geometric differences between the images. We would like to determine surfaces that can accurately represent local geometric differences between images and would not introduce unnecessary distortions to the resampled image. Surface fitting is an area that has been studied considerably in Computer-Aided Geometric Design as well as in Approximation Theory Excellent reviews of existing techniques can be found in [5, 6, 11]

We have studied the use of surface splines [12], multiquadrics [13], and sum of Gaussians [16] in representing the components of transformation functions in image registration Surface splines are defined by

$$f(x,y) = a_0 + a_1 x + a_2 y + \sum_{i=1}^{n} F_i d_i^2 \ln d_i^2, \qquad (11)$$

where d_i is the distance of an arbitrary point (x,y) in the reference image to point (x_i,y_i). The unknown parameters a_0–a_2 and F_0–F_n are determined by applying the coordinates of homologous points in the images into $\{f(x_i,y_i) = Z_i \quad a = 1, \quad ,n\}$ and solving the obtained linear equations. When f_x is to be determined $Z_i = X_i$ and when f_y is to be determined $Z_i = Y_i$ The parameters are also determined by satisfying the following boundary conditions:

$$\sum_{i=1}^{n} F_i = 0 \qquad (12)$$

$$\sum_{i=1}^{n} x_i F_i = 0 \qquad (13)$$

$$\sum_{i=1}^{n} y_i F_i = 0. \qquad (14)$$

$$(15)$$

Multiquadrics are defined by

$$f(x,y) = \sum_{i=1}^{n} \alpha_i \sqrt{(x-x_i)^2 + (y-y_i)^2 + r^2} \qquad (16)$$

Parameters $\{\alpha_i \quad 2 = 1, \quad ,n\}$ are determined by substituting coordinates of homologous points in the images into (16) and solving equations $\{f(x_i,y_i) = Z_i \quad a = 1, \quad ,n\}$, where $Z_i = X_i$ when f_x is determined and $Z_i = Y_i$ when f_y is determined. Parameter r defines the smoothness of an obtained surface and is typically between 0 and 1 Carlson and Foley [4] discuss selection of parameter r using the density of interpolating pomts.

Sum of Gaussian surfaces are defined by

$$f(x,y) = \sum_{i=1}^{n} V_i G_i(x,y), \qquad (17)$$

where $G_i(x,y)$ is a 2-D Gaussian centered at (x_i,y_i)

$$G_i(x,y) = \exp\{-[(x-x_i)^2 + (y-y_i)^2]/2\sigma^2\} \qquad (18)$$

Parameters $\{V_i \quad i = 1, \quad ,n\}$ are determined by applying the coordinates of homologous points in the images into (17) and solving $\{f(x_i,y_i) = Z_i \quad a = 1, \quad ,n\}$, where $Z_i = Xi$ when f_x is determined and $Z_i = Y_i$ when f_y is determined Parameter σ shows the smoothness of an obtained surface. Goshtasby and O'Neill [10] have characterized the smoothnesses of an obtained surface as a function of parameter σ

Appropriate σ can be chosen if some information about the images is known For instance, if it is known that the images to be registered do not have high local geometric differences, then σ can be taken rather large But, if the images to be registered are stereo, which may contain sharp local geometric differences, a rather small σ is needed Most often, however, such information is not available Therefore, σ should be selected taking into consideration the density of the homologous points. If a large number of homologous points is given, distances between adjacent points decrease and surfaces with relatively small smoothness parameters will be able to interpolate the points. However, if a smaller number of homologous points is given, since distances between adjacent points increase, a larger smoothness parameter is required to produce a smooth interpolation to the points. If a large number of homologous points is given and sharp local geometric differences exist between the images, when surfaces with large smoothness parameters are used, the interpolating surfaces may produce unnecessary fluctuations. Our experimental results show that if the average distance between adjacent points is D, parameter σ should be between $2/D$ and $5/D$

Note that a surface obtained by equation (11), (16), or (17) smoothly interpolates a set of 3-D points. If a surface smoothly interpolates the points and does not produce fluctuations, it closely follows the points. This means that if there is a large geometric difference between the images in a small area, the transformation will represent that difference well and will not propagate the deformation to other image areas Therefore, the local shape of the surface will closely reflect geometric differences between the images,

The multiquadrics, surface splines, and the sum of Gaussians are interpolating surfaces and require that a system of equations be solved to determine them. If

the number of homologous points is small ($n \leq 50$), the equations may be solved to register the images However, if a large number of homologous points ($n > 50$) is given, solution of a large number of equations is not practical, and a surface formulation that does not require solving a system of equations must be used

If a large number of homologous points is given and/or the distances between adjacent points in the images vary considerably, approximating rather than interpolating surfaces should be used. When a large number of homologous points is given, we will use a rational Gaussian (RaG) surface [9] to represent each component of a transformation function. RaG surfaces are similar to the sum of Gaussian surfaces except that, instead of pure Gaussians, rational Gaussians are used so that at each point in a surface, the sum of basis functions becomes equal to 1.0. This property is called the convex-hull property. Since a RaG surface approximates a set of points, it does not produce unnecessary fluctuations; and because it provides the convex-hull property, even when distances between adjacent points vary considerably, an obtained surface will closely follow the points. This is in contrast to multiquadrics, surface splines, and the sum of Gaussians, where areas that contain densely spaced points produce fluctuations in the surface, and areas where sparsely spaced points are available, the surface poorly interpolates the points

A RaG surface is defined by

$$f(x,y) = \sum_{i=0}^{n} Z_i g_i(x,y), \tag{19}$$

where $g_i(x,y)$ is the ith basis function of the surface defined by

$$g_i(x,y) = \frac{G_i(x,y)}{\sum_{i=1}^{n} G_i(x,y)}, \tag{20}$$

and $G_i(x,y)$ is a 2-D Gaussian centered at (x_i, y_i) (see (18)) When the 2-component transformation is determined, $Z_i = X_i$, while when the y-component transformation is determined, $Z_i = Y_i$ Note that, because of the way this surface is formulated, there is no need to solve a system of equations. A surface will be immediately obtained after substituting the coordinates of given points into the equation of the surface. The surface will not interpolate the points but will approximate them When the number of given points is large, this approximation will be sufficiently close to the points and will accurately register the images. As parameter σ is decreased, the surface will follow the points more closely In image registration, this means that the components of a transformation will more closely reflect local geometric differences between the images, and they will not propagate a local deformation to the areas in the resampled image. Note that a RaG surface also can be made to interpolate a set of points just like multiquadrics, surface splines, and the sum of Gaussians by solving two systems of linear equations. RaG surfaces are, however, computationally more intensive than the other three surface formulations, and the additional computations may not be justifiable when a small number of homologous points is available. However, when a large number of homologous points is given or when distances between adjacent points vary considerably in the images, since solving large systems of equations becomes impractical and sometimes ill-conditioned, one should use approximating surfaces as opposed to interpolating surfaces to represent the components of the transformation

Viewing image registration as a surface fitting problem has various advantages. It enables estimation of the geometry of the scene and provides the means to identify the wrong homologous points.

3 Identifying Wrong Homologous Points

Treating image registration as a surface-fitting problem enables us to guess the geometry of the scene and to detect the wrong homologous points. Suppose homologous points in two images are as shown in Fig. 1 The two components of the transformation function obtained from surface splines are shown in Fig. 2. There seems to be a relatively large geometric difference between the central portion of the images, while there is relatively little geometric differences between other areas in the images. This can be due to different factors. The elevation of the scene near the center of the images could be varying sharply In that case, from the surfaces in Fig. 2, we can guess the geometry of the scene. However, if it is known that the images are from about the same viewpoint of the scene, or if it is known that the scene is rather flat, then from Fig. 2 we can conclude that homologous points near the center of the images are wrong. Since nonlinearity in the y-component is more pronounced than in the 2-component of the transformation, we can conclude that homologous points near the center of the images are shifted more along the y-axis than along the x-axis. If we remove point 7 from the two images, we will obtain the surfaces shown in Fig. 3, which are nearly planar

High variations in a surface can be determined by discretizing the surface into a digital image (by sub-

stituting the rows and columns of the reference image as the x's and y's into the equation of the surface and recording the surface values in an array) and examining the array. The arrays obtained in this manner can be considered digital images and image processing techniques can be used to characterize them Figures 4a and 4b show digital images corresponding to the surfaces shown in Figs. 2a and 2b, respectively The heights of points in f_x and f_y increase as x and y, respectively, increase. The angles of f_x and f_y with the x-axis and the y-axis are both equal to 45 degrees when the images do not have any geometric differences. Rotating f_x and f_y by 45 degrees about y-axis and x-axis, respectively, we can make the surfaces horizontal Quantizing the surfaces in horizontal state, we will obtain the surfaces shown in Figs. 4c and 4d We can now use the peaks and valleys in these surfaces to characterize geometric differences between the images, If after rotating the surfaces, we negate surface values that fall below the xy plane, we then need only to look for the peaks in the obtained images. Figs. 4c and 4d actually show images corresponding to the horizontal surfaces after values less than zero are being negated Intensities of points in these images are proportional to the magnitudes of the surface values.

If we locate locally peaks in images obtained in this manner, we can determine the wrong homologous points. By overlaying Fig 1a with Figs. 4c and 4d, we obtain images 4e and 4f, respectively. We observe that locally peak values of Figs. 4c and 4d fall on top of point 7 Both components of the transformation function suggest that points labeled 7 in the images do not correspond to each other There are a couple of other smaller peaks in Fig. 4d, but they are much smaller and do not fall on top of any points, and, therefore, can be safely ignored These smaller peaks are artificial peaks generated by the surface splines and are not a part of the given data.

Some of the geometric differences between the images could be due to camera distortions. To determine the camera distortions, we can obtain the image of a scene whose geometry is known and compare it with a synthetically generated image of the scene. For instance, a regular grid may be placed in front of the camera and the image of the grid may be obtained. Since the geometry of the grid is known, correspondence can be established between grid points in an obtained image and grid points in the ideal image. The surfaces interpolating the corresponding grid points in the obtained and ideal images can then be used to characterize the camera distortions. Since the distortions could vary with the focal length of the camera, several images should be obtained with different focal lengths of the camera. Then using distortions at known focal lengths, distortions at an arbitrary focal length can be determined by interpolation [17].

If given images are from a 3-D scene, the scene structure will influence the shapes of obtained surfaces. Assuming that camera distortions are locally negligible, we can separate image geometric differences due to scene structure and due to wrong homologous points. To achieve this, we will split the given set of homologous points into two (even and odd) sets and use each set separately to obtain the transformation function. Since the two sets represent the same scene, if all homologous points are correct, similar transformations will be obtained. If the transformations are not similar, by subtracting them we can locate the wrong homologous points by locating areas where large differences exist.

Figures 6a and 6b show a pair of terrain stereo images from the CMU collection Homologous points determined by a corner detector are also shown The components of the transformation after being rotated to lie horizontal are shown in Figs. 6c and 6d These surfaces show disparities along x and y directions, respectively (In all images shown in this paper, x increases from top to bottom and y increases from left to right.) From Figs. 6c and 6d, a new image (Fig. 6e) is generated, showing disparities along both x and y directions. Elevation of a surface point in Fig 6e is equal to the square root of the sum of squares of corresponding surface elevations in Figs. 6c and 6d. We observe that this image has correctly predicted the geometry of the scene, except at an area in the lower right corner of the scene because of a wrong homologous points.

The transformation functions obtained using the odd homologous points and even homologous points are shown in Figs. 7a,b and Figs. 7c,d, respectively Since these images are stereo, disparities along the x axis are much smaller than those along the y axis. Disparities along the y axis describe the scene elevation Subtracting corresponding components of the transformations, we obtain Figs. 7e and 7f, both showing a potential wrong homologous point in the lower right corner of the images. By combining Figs. 7e and 7f we obtain Fig. 7g, showing differences along both x and y directions. Image value at a point in Fig. 7g is equal to the square root of the sum of squares of values at the same point in Figs. 7e and 7f. By overlaying the points in the reference image (Fig. 6a) with Fig 7g, we obtain Fig 7h, which clearly identifies the wrong homologous point. In this manner, surface variations from wrong homologous points can be separated from surface variations from scene structure and, thus, wrong homologous points can be identified even when

images contain variations due to scene structure.

4 Results

Difficulties arise in image registration when the images to be registered have intensity and/or geometric differences. To achieve a more successful registration, the intensity and geometric differences between the images should be reduced, To reduce intensity differences between the images, we use image gradients or image edges rather than image intensities. Image edges are stable and often appear across modalities, For instance, the boundary between the land and water in satellite images is distinguishable in both thermal and visible images, and the boundary between the skin and air is visible across modalities in medical images.

To reduce geometric differences between images, a coarse-to-fine strategy is adopted. At the coarsest level where the images have very little geometric differences, the registration is carried out using a linear transformation Thls step will basically align the images. The result of registration at the coarsest resolution is then used to resample the sensed image to align with the reference image at one level higher resolution By image resampling, the geometry of the sensed image is continuously kept close to that of the reference image. Since local geometric differences may exist between images at higher resolutions, a nonlinear transformation is used to register them. Again, the result of the registration at one resolution is used to resample the sensed image to overlay the reference image at one level higher resolution. This process is repeated until highest resolution images are registered Since at each resolution, the sensed image is resampled to have similar geometries as the reference image, geometric differences between the images will be kept small, facilitating the registration process. As the resolution of the images is increased, since geometric differences between the images become mostly local, the elasticity of the surfaces used to represent the components of the transformation function is also increased to accommodate larger local geometric differences between images A surface with a larger elasticity also ensures that a local deformation is not spread all over an image.

Results of our coarse-to-fine image registration method on a variety of images are presented below First, we register images 6a and 6b. Overlaying these images, we obtain image 8a. Disparities between the images are most clearly observable near the roads. Since the given images are stereo, there is no need to align them. Therefore, the image alignment step is skipped. After removing the wrong homologous point found in Fig. 6e or Fig. 7h from the set of homologous points, the remaining homologous points are used to determine two surface splines that overlay the images. Resampling image 6b according to the obtained surface splines and overlaying with image 6a, we obtain image 8b We see that most areas in the images register well However, there still remain some misalignments due to sharp changes in terrain elevations, Registration at the highest resolution will establish correspondence between high-gradient pixels in the images and fit elastic (RaG) surfaces to the points. The surfaces are then used as the components of the transformation function to register the images (Fig. 8c). By combining the components of the transformation, we obtain disparities along both x and y directions (Fig. 8d) Higher intensities in image 8d show larger disparities and thus correspond to higher elevations in the scene.

Another pair of stereo images is shown in Figs. 9a and 9b These images are from a recent Mars Pathfinder mission planned by NASA Figure 9c shows these images after being overlaid Again, the image alignment step is skipped because the images are already aligned We move to the mid resolution to find two surface splines that can register the images. Resampling image 9b according to the obtained surface splines and overlaying the resampled image with image 9a, we obtain image 9d. Although large features in the images overlay, the registration is not accurate as witnessed by blurred edges obtained in image 9d After completing registration at the highest resolution, we obtain the result shown in Fig. 9e. Object boundaries in this overlaid image are sharper, showing an accurate registration, The disparity map obtained as a result of this registration is shown in Fig. 9f, which reflects depths of points in the scene.

Registration of shuttle images from NASA are shown in Fig. 10. Figures 10a–10d are four views of a coastline. Geometric differences between the images are mainly global because of the rather flat coastal line. There are some local geometric differences between the image at areas away from the coastline Registering images 10a and 10b according to our coarse-to-fine method, we obtain image 10e. Here the first step of the process, whch aligns the images, is very crucial After aligning the images, registration at higher resolutions becomes rather trivial because of the small geometric differences between the images. In all examples demonstrated in this paper, registration was carried out at three resolutions: low resolution, mid resolution, and high resolution. Low resolution was the coarsest resolution and was obtained by replacing each block of 4×4 pixels with a single pixel by averaging intensities in the block and using that as the intensity of the corresponding pixel Mid resolution was obtamed

by replacing blocks of 2 x 2 pixels with a single pixel. High resolution was the finest resolution, corresponding to the original images. Registration of image 10a with images 10c and 10d are shown in images 10f and 10g, respectively. Since images 10b, 10c, and 10d are all resampled to register with image 10a, they can all be overlaid with image 10a to obtain image 10h. Image 10h, therefore, shows regstration of all four images (10a–10d).

Figures 11a and 11b are two cross sectional images of a human chest. Image 11a is a color image showing all chest anatomies very well, while image 11b is an x-ray computed tomography (CT) image showing the bone structures well. These images are from the Visible Human Data set provided by the National Library of Medicine. The CT image was obtamed when the cadaver was fresh (not frozen), while the color image was obtained after the cadaver was frozen It is anticipated that some local geometric differences will occur during the freezing process. The mid resolution and high resolution registration results are shown in Figs. 11c and 11d, respectively Although both images show similar results, by a closer inspection we see that along region boundaries a sharper registration is obtain in image 11d compared to image 11c

The last example demonstrates registration of two brain magnetic resonance (MR) images of the same patient taken at different times (Figs 5a and 5b) These images are provided by the Kettering Medical Center After aligning the images at the coarsest resolution, image 5c is obtained Although the skulls in the images align well, there is considerable misalignment between the brains. Most areas of the brains are similar in images 5a and 5b and can be registered with a rigid-body transformation Near the tumor area, however, there is a considerable local deformation. We would need a technique that can register the left half of the brain with an approximately linear transformation and register the right half of the brain with a nonlinear transformation function Figure 5d shows the result of our registration, adapting well to the different levels of geometric differences between the images. In the overlaid image we can observe the tumor growth between the times the images were obtained.

5 Conclusions

Selection of the right transformation function is very crucial in image registration A transformation function should be able to adapt well to local geometric differences between the images, and should not propagate a local deformation to other areas in an image. Transformation functions contain information about the homologous points, and the properties of the transformation functions can be used to eliminate the wrong homologous points. In this paper, preliminary results of a coarse-to-fine algorithm to register images with local deformations were presented The results are encouraging and we are now in the process of extending the technique into 3-D to register whole-body images.

The originals of figures presented in this paper may be viewed by browsing:
http://www.cs.wright.edu/people/faculty/agoshtas/irw97.html

References

[1] H G Barrow, J M Tenenbaum, R. C. Bolles, and H C Wolf, "Parametric correspondence and chamfer matching: Two new techniques for image matching," *Proc. J Conf Artificial Intelligence,* 1977, pp. 659–663

[2] G Borgefors, "Hierarchical chamfer matching: A parametric edge matching technique," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 10, no. 6, 1988, pp 849–865.

[3] L G Brown, "A survey of image registration techniques," *ACM Computing Surveys,* vol. 24, no. 4, 1992, pp. 325–376.

[4] R. E. Carlson and T A Foley, "The parameter R^2 in multiquadric interpolation," *Computers Math. Applic.,* vol 21, no. 9, 1991, pp. 29–42.

[5] R. Franke, "Recent advances in the approximation of surfaces from scattered data," *Topics on Multivariate Approximation,* Academic Press, 1987, pp. 79–98.

[6] R. Franke and L. Schumaker, "A bibliography of multivariate approximation," *Topics in Multivarzate Approximation,* Academic Press, 1987, pp 275–335

[7] J G Fryer and D C Brown, "Lens distortion for close-range photogrammetry," *Photogrammetnc Engzneenng and Remote Senszng,* vol. 52, no. 1, 1986, pp 51–58.

[8] A Goshtasby, "Design and recovery of 2-D and 3-D shapes using rational Gaussian curves and surfaces," *Int. J Computer Vision,* vol 10, no. 3, 1993, pp. 233–256.

[9] A. Goshtasby, "Geometricmodeling using rational Gaussian curves and surfaces," *Computer-Azded Design,* vol 27, no 5, 1995, pp. 363–375

[10] A Goshtasby and W D O'Neill, "Surface fitting to scattered data by a sum of Gaussians," *Computer Azded Geometric Deszgn,* vol **10, 1993,** pp. **143–156.**

[11] **E. Gross, "A** catalogue of algorithms for approximation," *Algorithms for Approtimation 11,* J C. Mason and M. G Cox (Eds.), Chapman and Hall, New **York, 1990,** pp. **479–514.**

[12] R. **L.** Harder and R. N Desmarais, "Interpolation using surface splines," *J Aircraft,* vol. **9, no. 2,** pp. **189–191, 1972.**

[13] R. **L.** Hardy, 'Theory and applications of the multiquadnc-biharmonic method," ***Comput. Math. Appl.,*** vol **19,** no. **8/9, 1990,** pp **905–1915**

[14] **A** K. Jain ***Fundamentals*** of ***Digital Image*** *Processing,* Prentice Hall, **1989,** pp **384–385.**

[15] F G Peet, "The use of invariants in the transformation and registration of images," ***The Canadian Surveyer,*** vol **29,** no. **5, 1975,** pp. **501–506.**

[16] **I** P Schagen, "The use of stochastic processes in interpolation and approximation," ***Intern. J Comput. Math.,*** vol **B-8, 1980,** pp. **63–76.**

[17] R. G Willson, ***Modekng and Calibratzon of Automated Zoom Lenses,*** Ph.D Dissertation, Carnegie Mellon University, January **1994,** Technical Report # CMU-RI-TR-94-03.

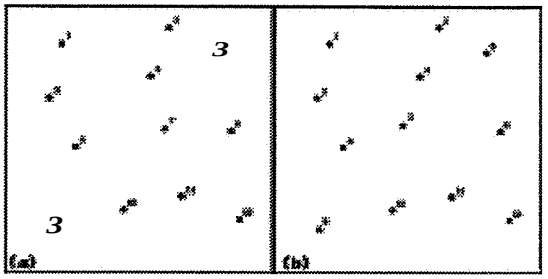


Figure 1. Corresponding points in two images.

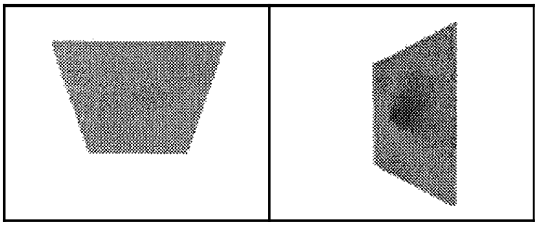


Figure 2. x and y components of a transformation.

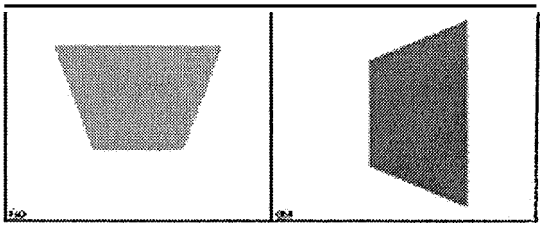


Figure 3. Nearly planar surfaces.

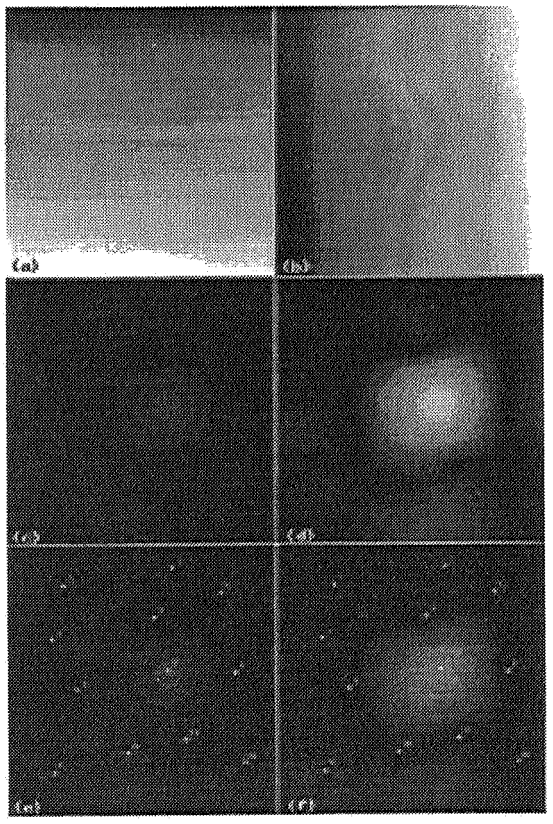


Figure 4. Detecting wrong homologous points.

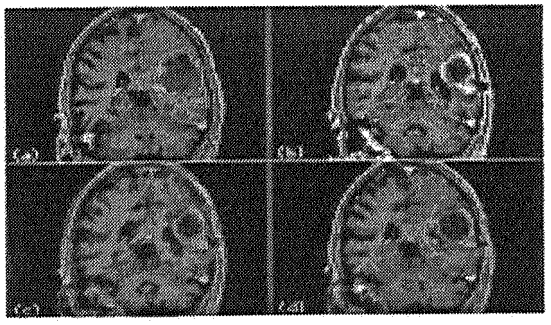


Figure 5. Registering brain MR images.

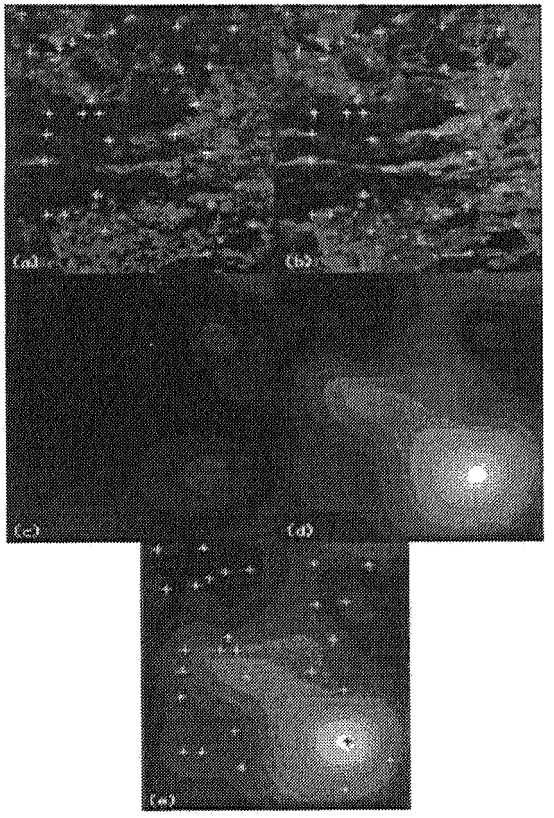


Figure 6. Detecting wrong homologous points in images with local geometric differences.

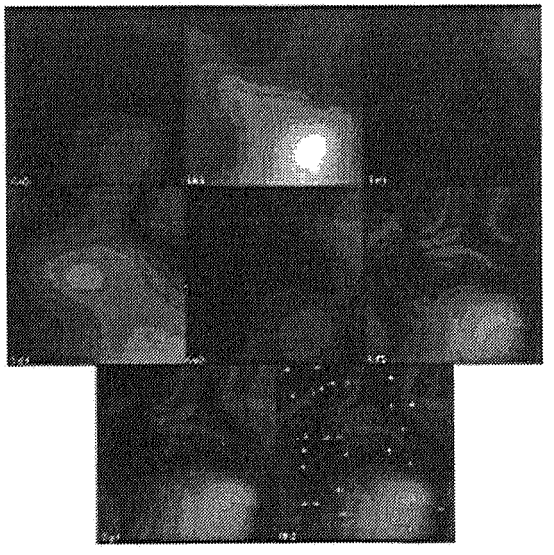


Figure 7. Separating geometric differences due to scene structure and wrong homologous points.

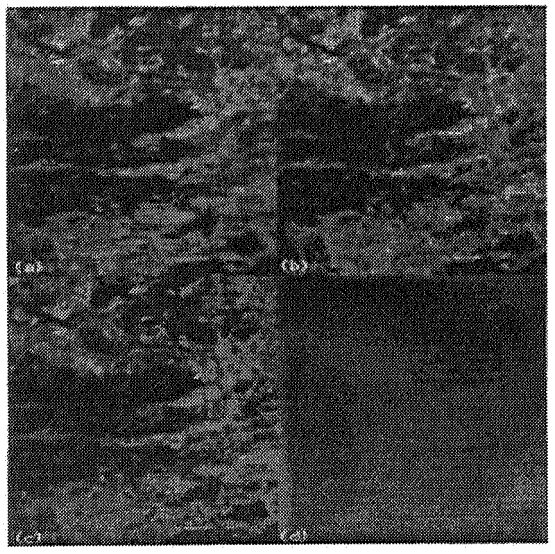


Figure 8. Registering images 5a and 5b.

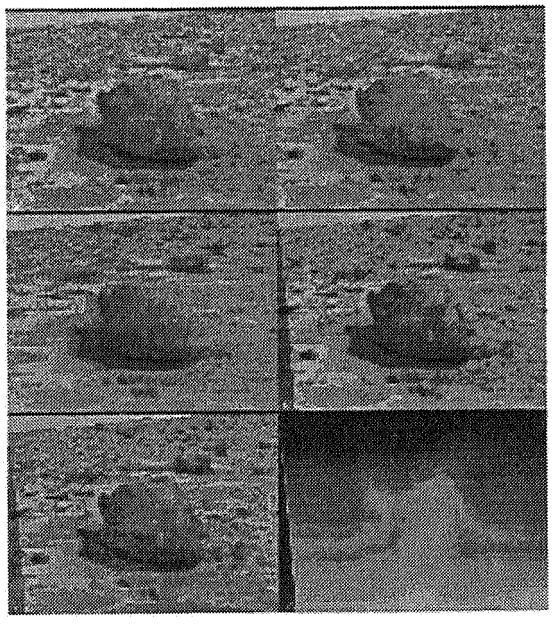


Figure 9. Registering a pair of stereo images.

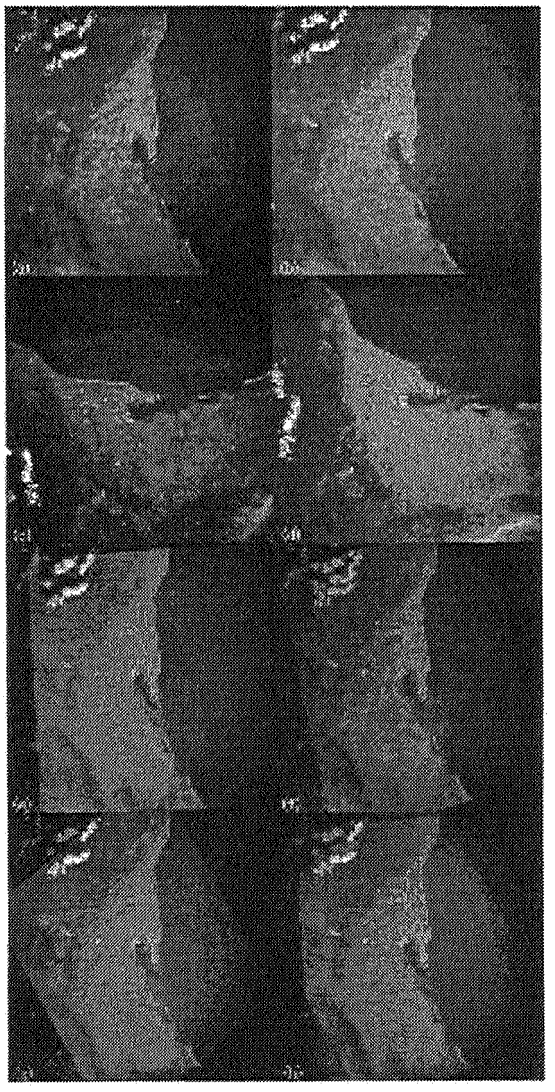


Figure 10. Registering shuttle images.

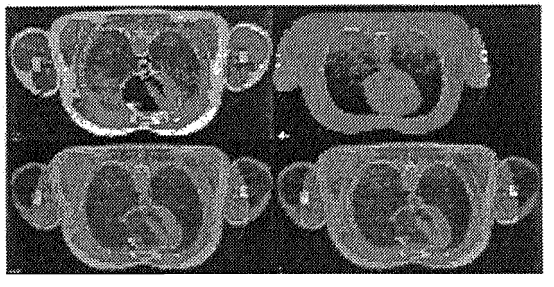


Figure 11. Registering CT and Visible human images.

Correspondence-less Image Alignment using a Geometric Framework

Venu Govindu Chandra Shekhar Rama Chellappa

Center for Automation Research, University of Maryland

College Park, MD 20742

Email : {venu,shekhar,rama}@cfar.umd.edu

Abstract

In this paper we describe a framework for ***aligning images without using explicit feature correspondences. We show how certain geometrzc propertzes of image contours are related to the parameters of the geometric transformation between the images. For a given transformation model, we show how the transformation parameters can be recovered using simple statistical distributions of these geometrzc properties. The use of these statistical descriptzons eliminates the need*** for ***establishing explicit feature correspondence. Further, the proposed method is robust to problems of occlusion, clutter and errors in low-level processing. We demonstrate the effectiveness of our method on various real tmages.***

1 Introduction

Image alignment (variously known as registration, positioning etc.) amounts to establishing a common frame of reference for a set of images of a common scene. Alignment is an essential step for data fusion, change detection, pose recovery etc. and has been widely investigated in various contexts. A good survey of the exisiting literature is available in [2]. Traditionally, the transformation required to achieve image alignment is computed using either feature matching [4] which usually results in a set of equations that can be solved to obtain the transformation parameters, or a search strategy that optimises a meaningful criterion for achieving image alignment ([1], [5]). However both these approaches have their drawbacks. While feature matching methods can give very accurate solutions, obtaining correct matches of features is a hard problem, especially in the case of images acquired using different sensors or widely different viewing positions. On the other hand, search strategies are not always helpful since they are computationally very expensive and typically need good initial guesses to ensure convergence to the correct solution.

We alleviate some of these problems in our approach to image alignment. While images may appear differently in different sensors, the relative qualitative differences are typically preserved, ie. two regions with different properties would still look different. As a result, the boundaries between them would be preserved. Therefore we use image discontinuities (ie. contours) for our approach and show how these contours are sufficient to determine the transformation between the images. In the rest of this paper we assume that an image consists of a set of contours. In Section 2 we develop our framework and show how simple geometric properties of points on contours in different images are re-

lated through the transformation between them. However since we do not want to establish explicit correspondence between feature points, we use distributions of the geometric properties instead and show how the transformation parameters can be recovered using these distributions. In Section 3 we develop the solutions for specific transformation models and we present our results in Section 4. In Section 5 we discuss our method and some implementation issues.

2 Geometric Framework

We consider each image to have been obtained by applying a projection T_i on a scene Ω which can be modelled as

$$I_i = T_i(\Omega_i), \Omega_i \subset \Omega, \bigcap_i \Omega_i \neq \emptyset$$

In other words, the images are views of scenes that have areas in common that have to be registered. Therefore, we can state our problem as follows: **Given images I and $\tilde{I}$ of Ω, compute the composite tmnsformation $T = T_1 \, o \, T_2^{-1}$ such that $T : T_2(\Omega_1 \cap_{Q2}) \rightarrow T_1(\Omega_1 \cap \Omega_2)$.**

To develop an intuitive understanding of our method, we consider the following simple example. Consider the contours shown in Fig. 1. The contour on the right is a rotated version of the one on the left. Let the rotation angle be θ. Now consider a given point (p) on the first contour image and its corresponding point on the second contour ($\tilde{p}$) (indicated by the circles). The slope angles of the tangents to points (p) and ($\tilde{p}$) are denoted by Ψ and $\tilde{\Psi}$ respectively. Therefore, we can observe that $\tilde{\Psi} = \Psi + \theta$. Now this relationship holds for every corresponding point pair on the contours and the rotation angle can be easily recovered from such measurements on given point pairs. However we do not need to establish point correspondences to extract the rota-

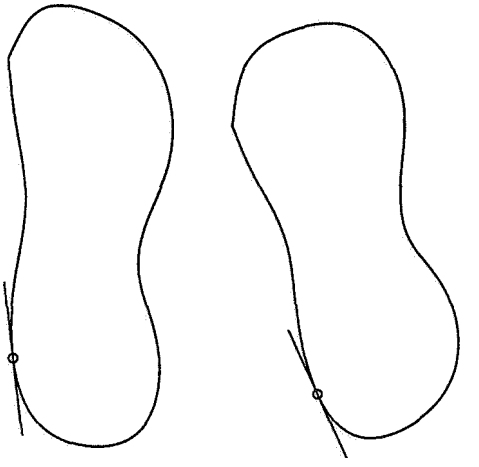


Figure 1: The contour on the right is a rotated version of the one on the left.

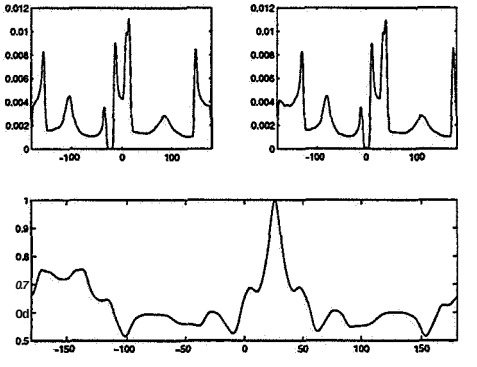


Figure 2: The plots on the top are the distributions of slope angles for the two contours. The plot on the bottom shows the circular cross-correlation of the distributions. The peak is the estimated angle of rotation in degrees.

tion angle. Instead, we note that since the above relation holds for ***every*** point pair, it ***also*** holds for the distribution of the measures on each image. In other words if we denote by $P(.)$ the distribution of slope angles of a contour, we have the simple relationship,

$$P(\tilde{\Psi}) = P(\Psi + \theta)$$

and hence given the two distributions of the slope angles in the two images, we can recover the rotation angle θ in a simple fashion (This can be done by **minimising** the norm $||P(\tilde{\Psi}) - P(\Psi + \theta|\theta)||$ which is equivalent to maximising the cross-correlation between the two distributions. See Fig. 2). Note that the probability distributions can be computed independently on each image without any need to establish feature correspondences.

The above idea can be extended to other transformation parameters. For a transformation model $T = T_{a_1,a_2,\cdots,a_n}$ where $\{a_1, a_2, \quad, a_n\}$ are n independent parameters, we can choose a reparametrisation of T such that $T = T_{g_1,g_2,\cdots,g_n}$, where $\{g_1, g_2, \quad, g_n\}$ are n independent functions of the original parameters $\{a_1, a_2, \quad, a_n\}$[1] Thereafter each of these new parameters can be estimated in a manner similar to that described above for the case of the rotation angle. These notions are formally developed below and the different transformation models are described in Section 3. We consider a point p on curve C in image I_1 and its corresponding point $\tilde{\mathbf{p}}$ OR curve $\tilde{C}$ on image I_2. The local geometric properties measured at points p and $\tilde{\mathbf{p}}$ are called geometric descriptors and are denoted **as** D and $\tilde{D}$ respectively. Therefore

$$D(\mathbf{p}) = \tilde{D}(\tilde{\mathbf{p}}, g(T)), \forall\ (\mathbf{p}, \tilde{\mathbf{p}}),$$

where $g(T)$ is a function of the transformation T

We denote a function $g_i(T)$ to be ***"observable",*** if its value can be recovered from the descriptor pair' Thus for an n parameter transformation model, we can recover the transformation given n independent descriptors that make the parameters observable.

Given a descriptor D, we can easily determine its probability distribution $\mathcal{P}(D)$ in an image by computing its value on every point along the contours and computing the distribution of the computed values. We denote the probability functions thus obtained from images $\mathbf{I}$ and $\tilde{I}$ **as** $\mathcal{P}(D)$ and $\mathcal{P}(\tilde{D})$ respectively. We can now describe the estimator for $g_i(T)$ given. the pdf's **as** follows:

Theorem 1 *Given an image pair I and $\tilde{I}$, and descriptors $D(\mathbf{p})$ and $\tilde{D}(\tilde{\mathbf{p}}|g(T))$, the Maximum Likelihood Estimator (MLE) for $g(T)$ is*

$$\arg \min_{g(T)} \| \mathcal{P}(D) - \hat{\mathcal{P}}(\tilde{D}|g(T)) \|_n \qquad (1)$$

where P and $\hat{\mathcal{P}}$ are the observed probability measures of D and $\tilde{D}$ respectively and $\| \ \|_n$ is the L_n norm for some $n > 0$. The probability distribution function (pdf) of the obseruatton noise is assumed to be monotonically decreasrng with a mode at zero.

For convenience, we shall henceforth denote $g(T)$ by g.

Proof :

[1]For example, under a Euclidean transformation, the new parameters could be the rotation angle, translations in x and y directions.

[1]In the case of rotation, $D = \Psi$ and $\tilde{D} = \tilde{\Psi}$ Hence $g_i(T) = \theta$ is observable, since its value can be obtained from those of D and $\tilde{D}$

We denote the true probability distributions of D and $\tilde{D}$ as P and $\tilde{P}$ respectively. The observed probability distribution of D is $\mathcal{P}$ and that of $\tilde{D}$ is modeled as .

$$\hat{\mathcal{P}}(\tilde{D}) = \tilde{\mathcal{P}} + \mathcal{N},$$

where $\mathcal{N}$ is an independent, stationary random noise process with a pdf that is monotonically decreasing with a mode at $\mathcal{N} = 0$.

Now we know that for every possible distribution pair $\{ P , \tilde{P} \}$ the relationship $\mathcal{P}(D) = \tilde{\mathcal{P}}(\tilde{D}|g)$ holds.

In the following we denote the conditional probability measures $\tilde{\mathcal{P}}(\tilde{D}|g)$ and $\hat{\mathcal{P}}(\tilde{D}|g)$ as $\tilde{\mathcal{P}}_g$ and $\hat{\mathcal{P}}_g$ respectively.

Therefore, using the observation model for estimating g

$$\hat{\mathcal{P}}_g = \tilde{\mathcal{P}}_g + \mathcal{N} \tag{2}$$

the MLE is defined as

$$\arg\max_g \mathbb{P}(\hat{\mathcal{P}}|\mathcal{P}, g),$$

where $\mathbb{P}$ denotes the probability of the observations $\hat{\mathcal{P}}$[3] In other words, we maximise the probability of jointly observing the two probability distributions P and $\hat{\mathcal{P}}$ on the images I and $\tilde{I}$ respectively.

For the given observation model, we get :

$$\max_g \mathbb{P}(\hat{\mathcal{P}}|\mathcal{P}, g) = \max_g \mathbb{P}(\hat{\mathcal{P}} - \mathcal{P}|g)$$

From the noise model we observe that:

$$\arg\max_g \mathbb{P}(\hat{\mathcal{P}} - \mathcal{P}|g) = \arg\min_g \| \hat{\mathcal{P}}_g - P \|$$

[3]Note that P and 6 are the observations we extract from the images.

By the triangle inequality and Eqn. (2), we know that:

$$\| \hat{\mathcal{P}}_g - \mathcal{P} \| \leq \| \tilde{\mathcal{P}}_g - \mathcal{P} \| + \| \mathcal{N} \| \tag{3}$$

Therefore,

$$\min_g \| \hat{\mathcal{P}}_g - P \| \leq \min_g \max_{\{\hat{\mathcal{P}}_g, \mathcal{P}\}} \| \hat{\mathcal{P}}_g - P \| \tag{4}$$

Since the right hand side of equation (4) is equal to $\| \tilde{\mathcal{P}}_g - \mathcal{P} \| + \| \mathcal{N} \|$, by the assumption of independence of noise, we have

$$\min_g \| \hat{\mathcal{P}}_g - \mathcal{P} \| = \min_g \| \tilde{\mathcal{P}}_g - \mathcal{P} \| \tag{5}$$

Hence the Maximum Likelihood Estimator for the parameter g is given by Eqn. (1)

3 Transformation models

In the following subsections, we examine specific transformation models and show how we can estimate the relevant parameters. We shall denote the points in the first image by $\mathrm{p} = (x, y)^T$ and those in the second image by $\tilde{\mathrm{p}} = (\tilde{x}, \tilde{y})^T$ The transformation model adopted is

$$\tilde{\mathrm{p}} = \mathrm{T}\mathrm{p} + \mathrm{t} \tag{6}$$

where $\mathrm{t} = (t_x, t_y)^T$ denotes the 2-D translation vector and the matrix T denotes a 2 x 2 invertible matrix. Henceforth we shall refer to matrix T as the transformation matrix.

3.1 Euclidean

The Euclidean transformation is parametrised by three parameters, the rotation angle θ and the two translation pa-

rameters t_x and t_y. Here we have

$$\tilde{\mathbf{p}} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \mathbf{p} + \mathbf{t} \qquad (7)$$

To compute rotation we choose the descriptors $D(\mathbf{p})$ and $\tilde{D}(\tilde{\mathbf{p}})$ to be the slope angles of points on the curves in the two images. For the example shown in Fig. 1, the rotation value can be computed using the MLE defined above. Having compensated for the rotation between the images, we can compute the x-direction translation t_x between the two images using $D(\mathbf{p}) = x(\mathbf{p})$ and $\tilde{D}(\tilde{\mathbf{p}}) = x(\tilde{\mathbf{p}})$, where $x(\mathbf{p})$ is the x-coordinate of point $\mathbf{p}$. The y-component of translation can also be computed in a similar fashion.

3.2 Similarity

To compute the similarity transformation, we need to compute an additional scale parameter s. Recalling that the radius of curvature (R) of a point is directly proportional to the scaling parameter, we have $R(\mathbf{p}) = sR(\tilde{\mathbf{p}})$. From this we can deduce that $D(\mathbf{p}) = \ln(R(\mathbf{p}))$ and $\tilde{D}(\tilde{\mathbf{p}}) = \ln(R(\tilde{\mathbf{p}}))$. Hence we have the simple additive relationship

$$D(\mathbf{p}) = \tilde{D}(\tilde{\mathbf{p}}) + \ln(s)$$

Also since the radius of curvature is independent of the slope angle, we can compute the scaling and rotation parameters independently

3.3 Affine

In the more general case, we have an affine transformation between two images which can be described by the equation

$$\tilde{\mathbf{p}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mathbf{p} + \mathbf{t} \qquad (8)$$

where the transformation matrix T is a non-singular 2 x 2 matrix. In this case, we need to compute four independent parameters of the transformation matrix T, which can be accomplished in the following manner.

To compute the determinant of the transformation ($|T|$), we define the matrices P and $\tilde{P}$ as follows:

If the curves are parametrised according to their arc-length representations (See [3]), then we denote s and $\tilde{s}$ as the arc-length indices of the curves C and $\tilde{C}$ respectively. We define the 2 x 2 matrices P and $\tilde{P}$ as $P = [\dot{\mathbf{p}}(s), \tilde{\mathbf{p}}(s)]$ and $\tilde{P} = [\dot{\tilde{\mathbf{p}}}(\tilde{s}), \ddot{\tilde{\mathbf{p}}}(\tilde{s})]$ for point pairs $(\mathbf{p}, \tilde{\mathbf{p}})$, where the dots denote derivatives with respect to the arc-length parametrisation (s) of the curve.

From the definition of the transformation, we get

$$\tilde{P} = TP \Rightarrow \ln|\tilde{P}| = \ln|T| + \ln|P|$$

Therefore we use $D(\mathbf{p}) = \ln|[\dot{\mathbf{p}}, \tilde{\mathbf{p}}]|$ and $\tilde{D}(\tilde{\mathbf{p}}) = \ln|[\dot{\tilde{\mathbf{p}}}, \ddot{\tilde{\mathbf{p}}}]|$ to compute the determinant of the transformation T It is important to observe that unlike in the case of invariants, we do not actually need to compute an "arc-length" reparametrisation of the curve that is invariant to the transformation. It suffices to compute the descriptors for a sufficiently dense number of points on the curves. This can be achieved by using local fitting of curves (See [3] for example). Thereafter, we need three more independent equations to solve for the

transformation.

We can derive the simple relationship

$$\frac{\ddot{\tilde{x}}}{\dot{\tilde{x}}} = \frac{a\ddot{x} + b\ddot{y}}{a\dot{x} + b\dot{y}}$$

By a simple reparametrisation of the parameters $(a, b) = \sqrt{a^2 + b^2}(\sin(\phi), \cos(\phi))$ in the above equation, we get

$$\frac{\ddot{\tilde{x}}}{\dot{\tilde{x}}} = \frac{\sin(\phi)\ddot{x} + \cos(\phi)\ddot{y}}{\sin(\phi)\dot{x} + \cos(\phi)\dot{y}} \tag{9}$$

from which we can solve for the ratio $(\frac{a}{b})$ using the left-hand side of Eqn (9) for $D(\mathbf{p})$ and the right-hand side for $\tilde{D}(\tilde{\mathbf{p}}, g(T))$.

By similar analysis, we can recover the ratio $(\frac{c}{d})$ using the relationship for $\frac{\ddot{\tilde{y}}}{\dot{\tilde{y}}}$ Finally, since we now know the ratios $(\frac{a}{b})$ and $(\frac{c}{d})$, we can observe that

$$\frac{\dot{\tilde{x}}}{\dot{\tilde{y}}} = \left(\frac{b}{d}\right)\frac{\frac{a}{b}\dot{x} + \dot{y}}{\frac{c}{d}\dot{x} + \dot{y}}$$

from which the ratio $(\frac{b}{d})$ can be computed by taking the logarithm on both sides of the above equation. It may be noted that the reparametrisation of (a, b) and (c, d) results in a finite range of possible estimate values of the new parameters (ie, range of ϕ is $[-\pi, \pi]$). The translational component can be recovered in a manner identical to those in the previous cases.

4 Results

In this section we present the results obtained by applying our method to a variety of images. The same technique is used for all the examples. First, a set of salient contours are extracted from the images by applying a Canny-like edge detector followed by a contour extractor. In general, the contours would be fragmented due to the presence of noise in the images. Contours fragments that are small (less than 20 pixels in length) are discarded. The relevant geometric properties are then computed for points on the contours and the probability distributions estimated Finally the transformation parameters are estimated as described in the previous sections.

Figure 3: Mosaic created out of a video sequence.

In Fig. 3, we show a mosaic constructed by aligning images from a video sequence in a common frame of reference using a Euclidean model,

Fig. 4 shows the alignment of aerial images of the Mojave desert obtained from a balloon flying over the area. The alignment achieved is accurate as is evident from the alignment of the image features like the roads, rock outcrops etc. In Fig. 5 we show the results for aligning a pair of images from the Landsat Thematic Mapper (TM). The images are from different bands of the electro-magnetic spectrum and therefore have different radiometric properties. The images are correctly aligned as can be observed from the continuity of the coastline and the alignment of features that run across the two images. Fig. 6 shows the results obtained for another pair of Landsat TM images.

One of the underlying assumptions for our method is that

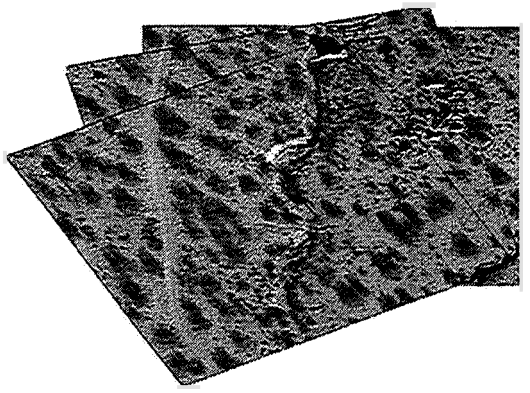


Figure 4: Alignment of a set of aerial images

the scenes being imaged are the same, ie. the contours in the two images arise from the same area and hence the metric minimised would result in the correct estimate for the transformation parameter. However, in the case where the overlapping areas of the image are small (ie. say 30 % or less), this assumption cannot be true anymore and hence the parameter estimates may not be correct. However by applying the estimated transformation to the images, we can increase the percentage of the overlap and hence better approximate the assumptions of a common scene. We have found that by extracting the overlapping areas after applying this initial estimate and recomputing the transformation using these areas we can get the correct transformation estimate. We illustrate this in the example in Fig. 7 The two images on the left were extracted from a larger image. The images are of size 300x300 and the relative translation between the images is 150 pixels in both the x and y directions (overlap is 25%). The initial estimate for the translation was (151,124) pixels. The estimated translation was applied to the images and the overlapping areas of the two images were extracted (The common area is now about 80%). Hence the new images are a better approximation to the underlying assumption of a common scene.

The required transformation is recomputed and the new estimate is found to be correct, ie (150,150) pixels to within subpixel accuracy. The registered images are shown in the right half of Fig. 7 We have found that in cases with little overlap, correct registration can be achieved by applying one or two iterations of the above method.

Finally, as an illustration of our algorithm for affine transformation estimation, we show the registration achieved for two MRI images of different modalities. (Fig.8) The image on the left in Fig. 8 is a proton density MRI image and the next one is the corresponding T2 weighted image that has been subjected to an arbitrary affine transformation. The image on the right shows the alignment of one contour from the different modalities. It may be noted that the correct alignment is achieved inspite of the difference in the photometric properties of the images. The alignment can be further refined if neccessary using any of the standard energy minimisation techniques.

All the above examples illustrate the ability of our algorithm in achieving the correct alignment for a variety of possible scenarios.

5 Discussion

An advantage of our method is the fact that we use the same method for computing the alignment for all cases. This is possible since we use the same distributional framework of local geometric properties to estimate the transformation parameters. As a result, the method works just as well for images with a large transformation between them as with images with small transformations. This is not always the case with energy minimisation methods which typically re-

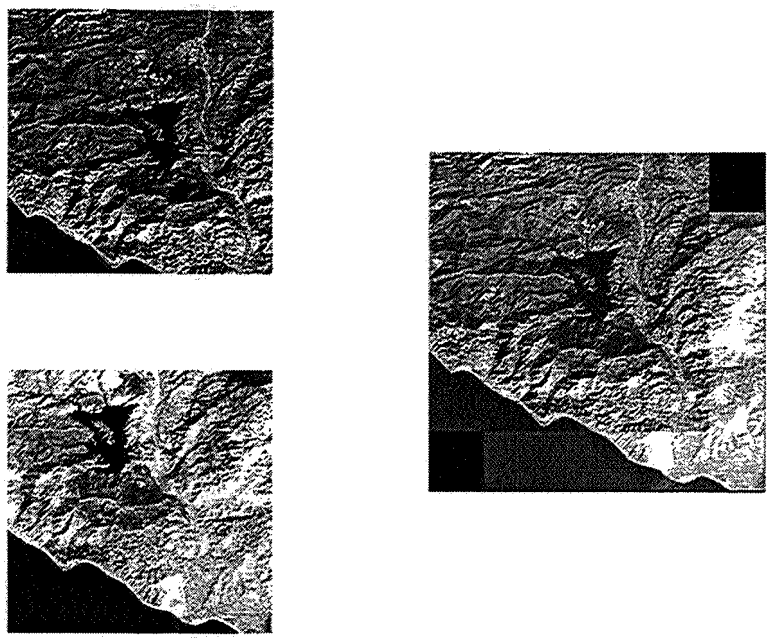


Figure *5:* Alignment of LANDSAT TM images from different bands.

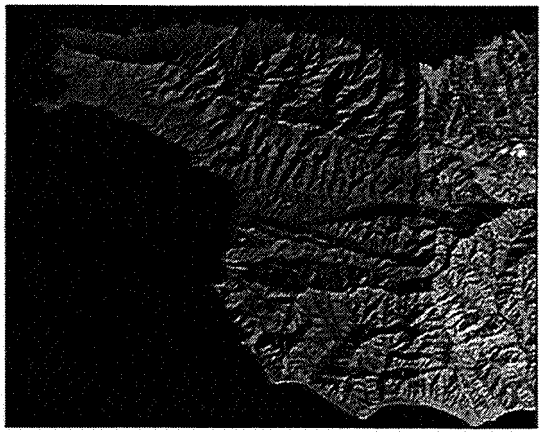


Figure **6:** Alignment of two LANDSAT TM images

quire the solution to be started fairly close to the true solution to avoid being trapped in local minima. *Also* any explicit need for domain dependent knowledge is avoided since the image primitives we use are contours that are usually extractable using standard edge-detection techniques. The use of a distributional method has other significant advantages too. Fragmentation of contours is easily handled since the local geometric properties can still be computed on the fragments without any significant loss of information. This is possible since if we have a contour, say, of length 200 pixels, fragmented into many segments that add up to say 180 pixels in total length, then we still have 180 points on which we can compute the geometric properties. The probability distribution that we estimate is not significantly changed. Therefore, we can observe a graceful degradation of the method with increasing loss of information due to occlusion or fragmentation. Such robustness would not be possible with the use of traditional global techniques like moments unless special care is taken to handle the change in shape due to occlusion or fragmentation. It is easy to note that the results of a moment-based method would be perturbed significantly due to small amounts of fragmentation of contours, while our method would still give the correct estimates.

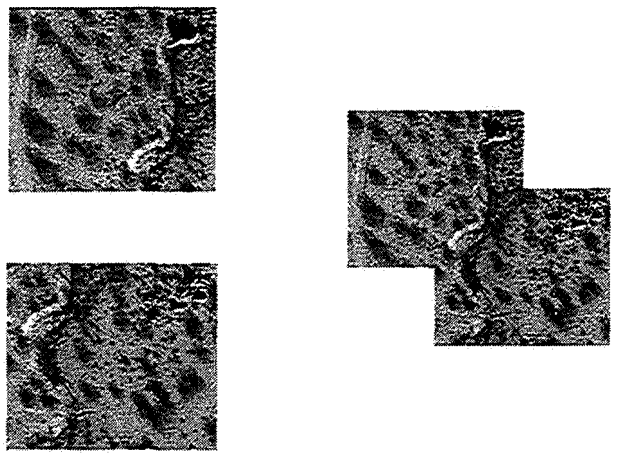


Figure 7: The images at the top have 25% overlap. The correctly registered images are shown in the lower half of the image.

In most voting based schemes, there is a combinatorial explosion of possible solutions that need to be checked as we increase the dimensionality of the search space or more significantly, as the number of observations are increased. This is so since most voting schemes, use all possible combinations of "hypothesised" feature matches to populate the space of possible solutions and then searches for a maxima in this space. In our case, we have two advantages in this regard. Firstly, all computations are carried out independently on each image and its only in the final stage that the two distributions are compared. Secondly, each stage of parameter estimation is one-dimensional since we parametrise each transformation into a set of independently estimated parameters. As a result, for a given number of contour points and for a given desired accuracy of estimation, the computational load is fixed.

While we have a simple theoretical framework for solving the problem of alignment, certain practical issues of implementation exist. As discussed in Section 4, our method is based on the underlying assumption that the images are views of the same scene. However in any practical scenario (except possibly video sequences), the amount of overlap between images to be aligned can vary a lot and can be small. As described earlier, we handle this problem by estimating the parameters, applying the estimated transformation to the images and recomputing the transformation by windowing the data. We have found that this issue is not a problem since we are able to converge to the true solution in at most two iterations of our method. Also while in most cases the MLE is unique, there are situations where there are multiple possible solutions. For example, in the case of urban aerial images, for the rotation angle there are typically two peaks that are separated by 90 degrees. This is due to the presence of contours that are oriented at 90 degrees to each other (since buildings are typically rectangular). Multiple peaks may also arise due to the presence of significant amounts of clutter in the images that corrupt the estimated probability distributions. However the number of such possible solutions is small (typically two or three instead of one) and it is easy to disambiguate between them by applying a measure of quality that is independent of the parameter being estimated.

As demonstrated by our examples, the theoretical framework we have developed works well for many real-life examples. While the existence of clutter, occlusion or the fragmentation of image contours does violate the underlying assumptions, our method is robust enough to be able to easily handle these problems. However it would be advantageous to modify the metric that we minimise so as to take into account knowledge about the scene. If we can determine which parts of the image contain areas that are

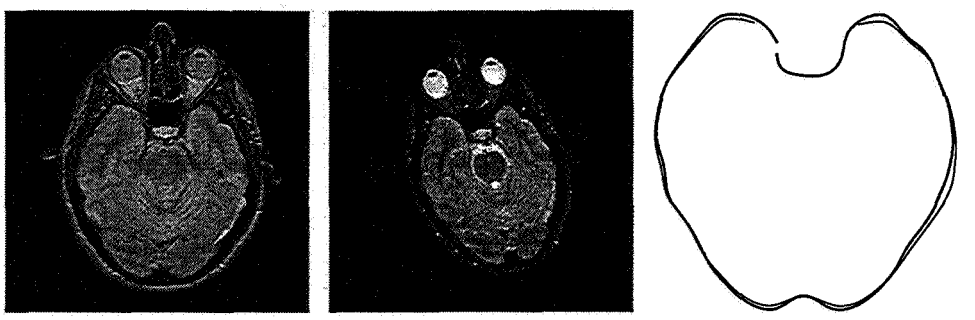


Figure 8: MRI image registration : The image on the left is a proton density image, the one in the middle is T2 weighted and the image on the right shows the registration of a single contour of the two modalities

visible in both images, which parts are occluded etc. then we should modify the metric to exploit this information to make our method further robust. Ideally, the distributions should be that of only the overlapping areas that are visible in both images without occlusion, ie. the descriptors should be "maximally common" One limitation of our framework is that the scene is assumed to be planar which may not be the case for certain scenarios. In such cases our method may not be applicable to achieve some initial estimate of the true projective transformation required to align such images.

6 Conclusion

We have described a framework for image alignment that does not use explicit feature correspondences. We have demonstrated the effectiveness and explained the advantages of using a distributional framework for computation of the parameters required for image alignment. This framework is robust and is more general than most existing methods.

7 Acknowledgments

We would like to acknowledge Eric Rignot (JPL), Prof B. S. Manjunath and maintainers of the UCSB Image Registration web site for providing the data sets used in this paper.

References

[1] P Viola and W Wells, '(Alignment by Maximization of Mutual Information", *Proceedings of ICCV,* pp. 16-23, 1995.

[2] L. G. Brown, "A Survey of Image Registration Techniques", *ACM Computing Surveys*, Vol. 24, No. 4, pp. 325-376, 1992.

[3] A. Bruckstein, R. Holt, A. Netravali, and T Richardson, "Invariant Signatures for Planar Shape Recognition Under Partial Occlusion", *CVGIP. Image Understanding,* Vol. 58, No. 1, pp. 49-65, July, 1993.

[4] H Li, B. Manjunath and S. Mitra, "A Contour-Based Approach to Multisensor Image Registration", *IEEE Transactions on Image Processing,* Vol. 4., pp. 320-334, No. 3, 1995.

[5] P Fua and Y Leclerc, "Image Registration Without Explicit Point Correspondences", *Proceedings of DARPA IUW*, 1994.

Finding Corner Point Correspondence from Wavelet Decomposition of Image Data

Mareboyana Manohar* and Jacqueline Le Moigne**

Abstract

A time efficient algorithm for image registration between two images that differ in translation is discussed. The algorithm is based on coarse-fine strategy using wavelet decomposition of both the images. The wavelet decomposition serves two different purposes: (i) its high frequency components are used to detect feature points (corner points here) and (ii) it provides coarse-to-fine structure for making the algorithm time efficient. The algorithm is based on detecting the corner points from one of the images called reference image and computing corresponding points from the other image called test image by using local correlations using 7x7 windows centered around the corner points. The corresponding points are detected at the lowest decomposition level in a search area of about 11x11 (depending on the translation) and potential points of correspondence are projected onto higher levels. In the subsequent levels the local correlations are computed in a search area of no more than 3x3 for refinement of the correspondence.

Introduction

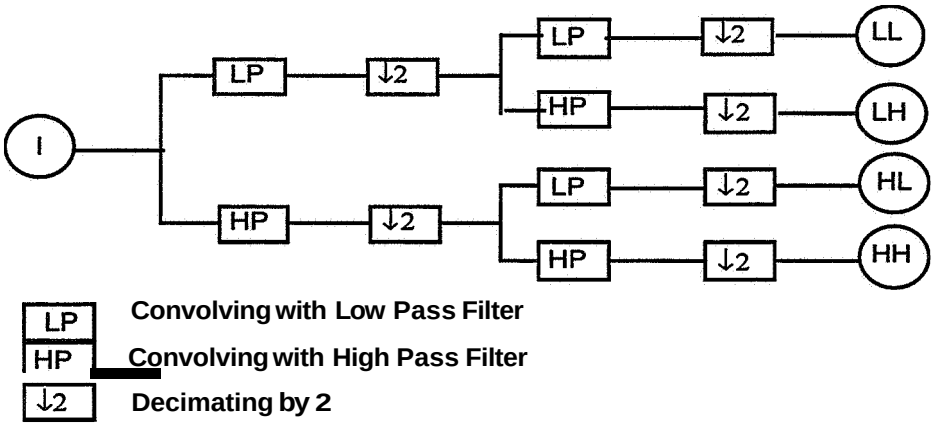
Image registration is an important task in spatial and temporal analysis, data fusion, medical diagnosis. Image registration involves computing a linear spatial transform of corresponding pixels for a given image pair that differ in translation, rotation and scale. Various image registration techniques are surveyed in [1]; these techniques are broadly classified into: correlation and sequential methods which are based on statistical properties of images, Fourier methods that use the properties of the Fourier transform, point mapping method based on feature detection and elastic model-based matching which utilize the characteristics of elastic models. Our work is based on the point mapping method, which involves the identification of distinct non-changing feature points in a reference image; for these points in the reference image the corresponding points in the test image are computed, and finally a linear transformation is computed from these correspondencies. Feature extraction is one of the critical components of image registration, because the success of the point mapping approach depends greatly on accurately identifying the feature points. The number of feature points obtained is also an important issue because it presents a trade-off between the accuracy and efficiency Feature points in the context of image registration are the ones that correspond to localized object characteristics such as corners, or vertices in the case of geometrical objects, vertices of the island boundaries in the case of remotely sensed images. Typically corners are considered to be feature points.

We use wavelet decomposition for corner detection that is time efficient as well as noise immune. The wavelet decomposition of image data is a very effective tool in many image processing techniques such as image compression, edge and corner detection, object recognition, image registration to name a few. However, very little work is available in the literature, on how one can use the wavelet decomposition for feature detection that can be an important step in image registration. In [2], features are modeled and an image dictionary is used to store some basic features of an image, and feature detection is performed by matching against the components of the image dictionary Yet another treatment of feature detection for matching pursuits can be found in [3]. In this paper we propose a computationally feasible, noise insensitive algorithm for corner detection using the wavelet decomposition method. We then describe a coarse to fine computational structure for computing corresponding pixels to all corner points from the reference image.

Multiresolution wavelet decomposition

The multiresolution approach of wavelet decomposition [4], involves the decomposition of a given image into four images named LL(Low-Low), LH(Low-High), HL(High-Low) and HH(High-High). The four images are obtained in the following manner: LL is obtained by convolving the row elements of the original image by a low pass filter, decimating by two; convolving the column elements of the obtained image, and decimating by two. LH is obtained by a similar procedure; a low pass filter and a high pass filter are used for convolution with the row elements and column elements respectively The other images are obtained in a similar manner. The complete process of one level wavelet decomposition is illustrated in Figure 1 The LL component is one-fourth the size of the original image and the rest of the components LH, HL and HH also one-fourth the size of the original image, enhance the horizontal, vertical edges and corners respectively The LL component is further decomposed into four compoents. Thus a given image can be recursively decomposed into LL, LH, HL and HH obtaining a multiresolution representation of the image also known as the pyramidal representation.

* Department of Computer Science, Bowie State University, Bowie, MD 20715, email manohar@cs.bowiestate.edu
** USRA/CESDIS; NASA/ GSFC- Code 930.5, Greenbelt, MD 20771, email: lemoigne@nibbles.gsfc.nasa.gov



Fig, 1 One Level Wavelet Decomposition

Algorithm

The multiresolution approach of wavelet decomposition [4], involves the recursive decomposition of a given image into four **quadrants:** LL(Low-Low), LH(Low-High), HL(High-Low) and HH(High-High). Since HL and LH decomposition **are** sensitive to horizontal and vertical edge features, they can be used to make inferences about the comer points in the image **data.** We use comer metric on HH, LH, and HL decomposition by using a simple **3x3** operators. Depending on the response of these operators on HH, LH and HH, we determine if the given point is a comer or not.

The two images to be spatially registered are wavelet decomposed to kth level. Comer points are detected from the reference image at kth level using its kth level HH, LH, HL components. Strong comer points are used to eliminate other weak comers within a window of 11x11 to make sure that too many comer points are not detected. For every comer point detected from reference image a window (typically 11x11 pixels) is selected centered around the comer point from LL component. This window is used to compute the corresponding point from the test image by spatial correlations on the HH component of test image. The comer points from reference image and corresponding points from test image are projected to the next higher level of wavelet pyramid. The comer points and corresponding points are refined in **3x3** windows. **This is** continued to full resolution of the image.

Results

The algorithm is demonstrated using a pair of images of synthetic images of hammer (Figures *2* and 4). Comer points detected from the reference image are shown in Figure **3** at original resolution. The test image (Figure 4) is generated by translating the reference image by a certain value in X- and Y- directions. The translation parameters are then found **by** running the algorithm on these images. The results are given in Table 1 at the 3^{rd} level of decomposition. The comer points of the reference image and corresponding points from the test image are projected on the next higher level and accurate comer points and corresponding points are computed by searching in **3x3** window. Projection, fine comer locations and corresponding points are performed recursively up to the original resolution. The results of the corner points and corresponding matching point locations from the test image are shown in Table **2** at the original resolution. Note that translation of the comer points is consistently **(32, 32).** Experiments are conducted by changing the translation parameters from (8,8) to (80,80).

This algorithm is also tested on a pair of **AVHRR** (Advanced Very **High** Resolution Radiometer) data. The results from these experiments are not available and will be reported in future publications. Currently, the program is designed for images that differ only in translations along X- and Y- directions.

X	Y	X'	Y'	ΔX	ΔY
28	9	32	13	4	4
25	10	29	14	4	4
41	10	45	14	4	4
22	12	26	16	4	4
29	13	33	17	4	4
37	12	41	16	4	4
26	13	30	17	4	4
38	14	42	18	4	4
33	30	37	34	4	4

Table 1 X, Y are the coordinates of the comer points from the **k(=** 4)th decomposition of reference image; X', Y' are corresponding points of the test image. **AX,** AY are the X-translation and Y-translation respectively. The comer given in the table are typical points out of **30** points detected.

X	Y	X'	Y'	ΔX	ΔY
228	76	259	107	31	31
199	82	231	114	32	32
333	77	365	109	32	32
182	100	213	131	31	31
238	102	269	134	31	32
300	100	331	131	31	31
210	102	241	133	31	31
304	108	335	139	31	31
264	244	295	275	31	31

Table 2: Corner points and corresponding points at the original resolution.

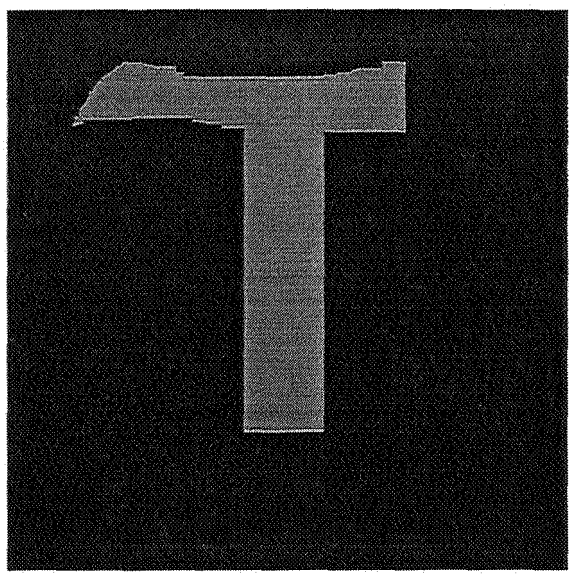


Figure 2: Hammer image

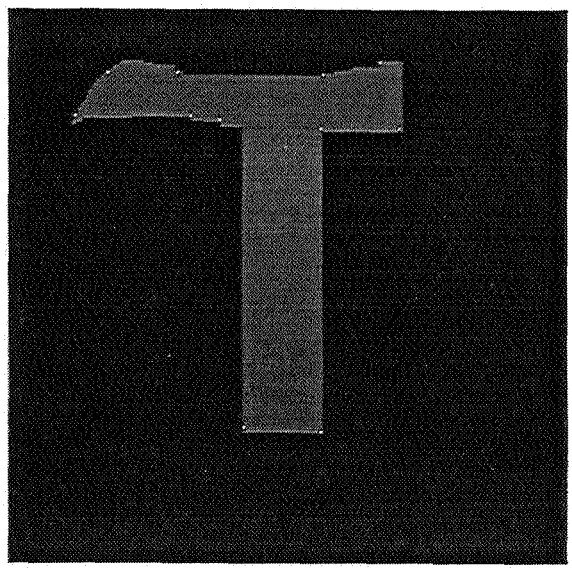
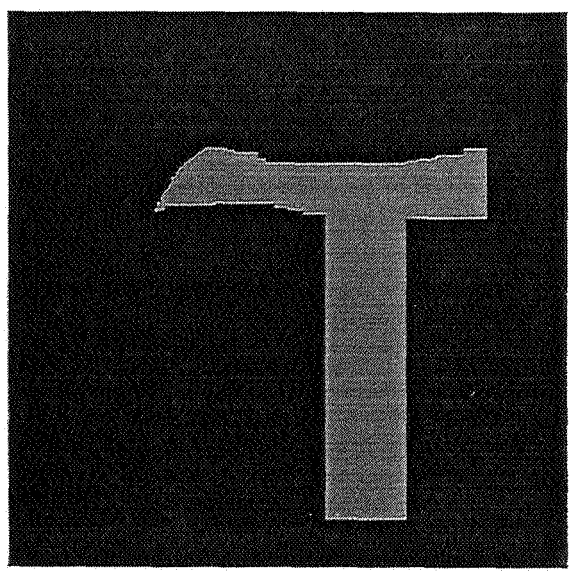


Figure 3: Corner detection results of the Hammer image shown in Figure 2. Corner points appear as bright points in the image. Notice in the upper right corner of the image only one of the two visible corners is retained by algorithm.

Figure 4: Translated image of Hammer



References

[1] Brown, L. G., "A Survey of Image Registration Techniques," ACM Computing Surveys, Vol. 24, No. 4, Dec. 1992

[2] J. S. Lee, Y N. Sun, and C. H. Chen, "Wavelet Transform for Corner Detection", international Conference on ASSP, 1992.

[3] Bergeaud, Francois and Stephane Mallat, "Matching Pursuits of Images," Proceedings of the IEE-SP International Symposium on Time-Frequency and Time Scale Analysis, 1994.

[4] Mallat, S. G., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 7, July 1989.

Efficient Algorithms for Robust Feature Matching

David M. Mount* Nathan S. Netanyahu† Jacqueline Le Moigne‡

Abstract

One of the basic building blocks in any point-based registration scheme involves matching feature points that are extracted from the sensed image to their counterparts in the reference image. This leads to the fundamental problem of point matching: given two sets of points, find the affine transformation that transforms one point set so that its distance from the other point set is minimized Because of measurement errors and the presence of outlying data points, it is important that the distance measure between two point sets be robust to these effects. We measure distances using the generalized Hausdorff distance.

Point matching can be a computationally intensive task, and there have been a number of algorithms and approaches proposed for solving this problem both theoretical and applied. We present two approaches to the point matching problem, in an attempt to reduce the computational complexity of the problem, while still providing guarantees on the quality of the final match. Our first method is an approximation algorithm, which is loosely based on a branch-and-bound approach due to Huttenlocher and Rucklidge [HR92, HR93]. We show that by varying the approximation error bounds, it is possible to achieve a tradeoff between then quality of the match and the running time of the algorithm. Our second method involves a Monte Carlo method for accelerating the search process used in the first algorithm With high probability this method succeeds in finding an approximately optimal match We establish the efficiency of our approachs empirically.

*Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland. Email: mountQcs.umd.edu. The support of the National Science Foundation under grant CCR–9712379 is gratefully acknowledged.

†Center for Automation Research, University of Maryland, College Park, and Center of Excellence in Space Data and Information Sciences, (CESDIS) Code 930.5 Space Data and Computing Division, NASA Goddard Space Flight Center, Greenbelt, Maryland. Email: nathanQcfar.umd.edu The support of the Applied Information Sciences Branch (AISB), Code 935, NASA/GSFC, under contract NAS 5555-37 is greatfully acknowledged

*Universities Space Research Association/CESDIS, Code 930.5, Space Data and Computing Division, NASA/GSFC, Greenbelt, Maryland. Email: lemoigneQcesdis.gsfc.nasa.gov The support of the AISB, Code 935, NASA/GSFC is greatfully acknowledged.

1 Introduction

There are two main approaches for image registration. One approach makes direct use of the original data (or edge gradient data) and the other is based on feature matching. Features could be control points, corners, line segments, etc. They are assumed to be available through any number of standard feature extraction algorithms. The former approach, which is based on a correlation measure between the two images, could prove to be computationally expensive and sensitive to noise unless substantial preprocessing is employed (See, e.g., Brown's survey paper [Bro92] for an elaborate discussion.) The latter approach, on the other hand, tends to yield more accurate results, as features are usually more reliable than intensity or radiometric values. We note, though, that the matching process may run into difficulties if feature extraction yields a significant number of missing or spurious points.

Feature-based methods can also be computationally expensive. The reader is referred to [TJ89], for a partial review on the computational complexity of point matching methods. In an attempt to arrive at a sound registration scheme that is both accurate and fast, we consider the problem of point matching, as part of a robust feature-based matching module. We believe that the availability of such a generic module will prove useful for a versatile image registration toolbox (currently under development by a team of researchers at NASA/GSFC [XMT+97]), and will enhance its overall functional capabiity The goal of this paper is to explore algorithmic methods for solving the robust point matching problem.

The algorithms presented in this paper can be broadly classified according to the following four components, proposed by Brown [Bro92] to classify any image registration method

Feature space: The domain in which information for matching is extracted are control points that were extracted in the image domain

Search space: The class of transformations that establish the correspondence between the sensed image and the reference data, are 2-dimensional affine transformations, allowing translation, rotation, scaling along each axis, and shearing.

Search strategy: The search for the optimal transformation is based on a branch-and-bound search of trans-

formation space.

Similarity metric: The figure of merit assigned to a match that is determined by a specific transformation is based on the generalized (directional) Hausdorff distance (described below in detail). This is a robust measure, in the sense that a fixed fraction of data may be spurious or missing without biasing the results.

The point matching problem can be defined abstractly as follows. We are given two points sets A and B lying in (conceptually at least) two different spaces. We are given a space $\mathcal{T}$ of dowable transformations mapping one space to the other, We are also given some measure of the distance between any two point sets, The problem is to find the transformation $t \in \mathcal{T}$ that *minimizes* the distance between $t(A)$ and B.

Because of measurement errors, obscuration, and missing or spurious data it is important that the distance between two point sets be measured in a manner that is robust to these effects. For example, minimizing the sum of squared distances between each point of $t(A)$ and its nearest neighbor in B may produce erroneous results if there are points of A which do not match any point of B For this reason we consider a well-known robust distance function, the ***generalized Hausdorff* distance** [HR92, HR93, HKR93]. Given a parameter k, $1 \leq k \leq |A|$, define the generalized (directed) Hausdorff distance from A to B to be

$$H_k(A, B) = K^{th}_{a \in A} \min_{b \in B} \text{dist}(a, b),$$

where K^{th} returns the kth smallest element of the set, and where $\text{dist}(a, b)$ is the Euclidean distance from point a to b. (Readers familiar with the work of Huttenlocher et al. should take note that we reverse the roles of A and B.) The parameter k is based upon ***a priori*** knowledge of the number of points of A that are expected to find a close match in B under the optimum transformation These points are called **inliers.** Observe that the definition is asymetric, When matching, it is possible to consider the directed distances both from A to B and from B to A Although we have not considered this, it would be a straightforward extension to our algorithms.

It is often more natural to express k in terms of a quantile. For $0 < q \leq 1$, define $H_q(A, B)$ to be $H_k(A, \text{B})$ where $k = \lfloor q|A| \rfloor$. The Hausdorff distance is robust to outliers, since as long as the fraction of inliers in A is at least $q \geq 0.5$, then the distance cannot be affected adversely by the outliers, no matter where they are placed. To use the terminology of robust statistics, this estimator has a 50% breakdown point [DH83, RL87].

There have been a large number of approaches for solving the point matching problem, arising from the fields of computer vision and pattern recognition, as well as from computational geometry. Most of the formulations of the point matching problem from computational geometry are not suitable for use in noisy, cluttered images either because they require exact matches [dRL95], or because they assume that there are no outliers. The latter may occur because the similarity measure requires either that the point sets have equal cardinality [AMWW88] or by using the (standard) Hausdorff measure, so that every point of A must have a nearby matching point in B [AAR94, CGH+97]. Even under these relatively restrictive assumptions, the computational complexity can be quite high. For example, the best known algorithm for determining the translation and rotation that *minimizes* the Hausdorff distance between sets of sizes m and n is $O(m^3 n^2 \log^2 mn)$ [AAR94, CGH+97]. This complexity may be unacceptably high for applications involving hundreds to thousands of points. In an attempt to circumvent this complexity, some researchers have considered Monte Carlo algorithms, which may generally fail to find the optimum match, but the probability of failure can be made arbitrarily small An example is the method based on geometric hashing proposed by Irani and Raghavan [IR96]. (For a more complete discussion, *see* [MNM97].)

Numerous point matching algorithms have been also proposed and applied by researchers in the fields of computer vision and pattern recognition. These range from relaxation-based methods (e.g., [RR80, TJ89]) to cluster detection in transformation space (by computing point-to-point correspondences) [SKB82, GS85, GSP86], to hierarchical decomposition of transformation space coupled with the application of a robust similarity measure [HR93, HKR93, OH97, HV97]. Most of the techniques presented in these papers are computationally intensive (in a worst case theoretical sense), or could take a long time to run, in practice.

In this paper, we propose two methods to ameliorate the high complexity of solving the point matching problem, while still guaranteeing the accuracy of the resulting match. One way to reduce the complexity is to provide the user of the algorithm a tradeoff between the algorithm's running time and performance guarantees on the final match. To this end we present an approximation algorithm for the point matching problem. The algorithm is quite flexible in that it can be applied to many natural subsets of affine transformations (including translations, rigid motions, homothetic (angle preserving) transformations, and general affine transformations), When a *priori* information is present on the bounds of the possible range of transformations, these can be incorporated into the model We also allow the user to adjust a set of parameters controlling the allowed relative, absolute, and quantile errors made by the algorithm. This flexibility makes it possible for the user to dramatically decrease the running time of the algorithm, often with a relatively small effective error in the final match.

The approximation algorithm is loosely based on a branch-and-bound search of transformation space described in various publications by Huttenlocher, Klanderman, and Rucklidge [HKR93, HR92, HR93, Ruc95, Ruc96]. In particular, the transformation space is subdivided hierarchically into rectangular regions, called cells. For each cell we either determine that it cannot contain the optimum transformation and remove it from further consideration, or we subdivide it into disjoint subcells. In this way the search focuses on the most promising regions of the transformation space. We sample a transformation from each cell, and record the best transformation encountered. For our approximation algorithm, this is repeated until it can be infered that all remaining cells cannot supply a significantly better transformation than the best seen so far

The running time of branch-and-bound search algorithm can be quite high, especially when the search space has many degrees of freedom and high accuracy is required. For this reason, we have augmented the search algorithm with a method that can radically decrease the execution time, but the price that the user pays is that the resulting algorithm may fail to find the optimum transformation with some small probability, which is under the user's control. (An algorithm having this property is called a **Monte Carlo algorithm.)** The method is based on a common idea, which we call **pinnrng.** Once the image of any constant number of points of A have been fixed, they determine a unique affine

transformation. (The actual number depends on the space of dowable transformations. For example, three points from A and three corresponding points from B, determine a unique affine transformation.) Rather than try triples blindly, we use the information acquired from the branch-and-bound search to advise us on the subsets of triples that are likely to lead to promising transformations. We show that this approach can lead to significant improvements in running time, but with a small probability of failing to find the optimum transformation.

The rest of the paper is organized as follows. In Section 2 we describe the branch-and-bound approximation algorithm. In Section 3 we present a variant of the method due to pinning. Section 4 provides preliminary empirical results, and Section 5 contains a brief discussion.

2 The Approximation Algorithm

Recall that we are given two points sets A and B and a space of allowable transformations $\mathcal{T}$. We are also given a distance quantile q, $0 < q \leq 1$. The problem is to find $t \in \mathcal{T}$ that minimizes the generalized Hausdorff distance $H_q(t(A), B)$. Since A and B will be fixed for the remainder of the discussion, let us call this $sim_q(t)$, for the similarity measure oft. (Note that the better the match, the smaller the similarity measure.) Let t_{opt} denote this transformation, and let sim_{opt} be this optimum similarity.

To provide a tradeoff between optimality and speed, we allow the algorithm to make an error in its answer. There are three natural ways to define this error. First, it may be suboptimal by some relative error factor, second, it may be suboptimal by some absolute additive error, and third, it may be less robust than the optimum, by using a somewhat smaller quantile in the Hausdorff distance. Rather than define three separate approximation problems, we provide three parameters that can control all of these effects. In particular, we are given three nonnegative values:

- ϵ_r, the relative error bound,
- ϵ_a, the absolute error bound,
- ϵ_q, the quantile error bound

Define $q' = (1 - \epsilon_q)q$, to be the weak *quantize.* Note that since $q' \leq q$, we have $sim_{q'}(t) \leq sim_q(t)$, for any $t \in \mathcal{T}$. We say that a transformation t is *approximately optimal* relative to these parameters if either of the following *approximate correctness conditaons:*

$$sim_{q'}(t) \leq (1 + \epsilon_r) sim_{opt} \quad \text{or}$$
$$sim_{q'}(t) \leq sim_{opt} + \epsilon_a$$

Thus, the approximate solution is slightly less robust, in that it considers only the weak quantile rather than the true quantile, and may exceed the optimum similarity by a relative error of ϵ_r or an absolute error of ϵ_a. In contrast, the approach used by Huttenlocher and Rucklidge [HR93, Ruc95] assumes a finite search space resulting from a fixed rasterization of transformation space. Their algorithm is guaranteed to return a result whose absolute error is within the limits of precision of the rasterization. Our search space is infinite, and hence there is a possibility that the algorithm may loop infinitely, Later we show that as long as $\epsilon_a > 0$ or $\epsilon_r sim_{opt} > 0$, then termination is guaranteed.

There are a number of different ways to describe the space $\mathcal{T}$ of allowable transformations. In our implementation we have assumed that $\mathcal{T}$ is a bounded subset of affine transformations. Assuming that the point sets A and B are both in 2-dimensional space, any affine transformation t applied to a point $a \in A$ can be expressed as a linear transformation M followed by a translation c:

$$t(a) \quad Ma + c,$$

where a and c are 2-element column vectors, and M is a 2×2 matrix. Since there are six parameters needed to define the transformation, this naturally leads to a 6-dimensional *transformation space.* There are somewhat more natural ways to describe an affine transformations (in terms of translation, rotation, scaling, and shearing), and these could have also been used as a basis for the search. We have chosen this particular representation because it is linear and hence easy to work with and analyze.

To prevent highly degenerate transformations we also allow the possibility of imposing additional constraints on the matrix M. These might include bounds on the determinant and norm of the matrix M. When transformations are sampled by the algorithm, these constraints are tested. (This might also be handled by computing reverse directional Hausdorff distances.)

From this perspective, the point matching problem is to find the transformation $t \in \mathcal{T}$ with the lowest similarity measure. Let us first describe the algorithm at a high level for the exact optimization problem, and then we will consider the modifications for an approximation algorithm. Following Huttenlocher and Rucklidge [HR93, Ruc95], we solve this by a branch-and-bound search of transformation space. We construct a search tree, where each node of this tree is identified with the set of transformations contained in some axis-aligned hyperrectangle in the 6-dimensional transformation space. These hyperrectangles are called *cells.* Each cell T is represented by a pair of transformations, (t_{lo}, t_{hi}), whose coordinates are the upper and lower bounds on the transformations of the cell. Any transformation whose coordinates lie between the corresponding coordinates of t_{lo} and t_{hi} lies in this cell.

The algorithm begins with an initial cell, which is known to contain the optimum solution. This is supplied by the user, based on *a priori* knowledge of the nature of the transformation (e.g., bounds on the possible translation and rotation). A cell is either *active,* meaning that it is a candidate to provide the optimum transformation, or *killed,* meaning that it cannot provide the optimum solution. At any time the union of the active and killed cells forms a cover of the initial cell into subcells with disjoint interiors. The algorithm proceeds by selecting one of the active cells to be *processed.* After processing, a cell is either killed, or it is split into two disjoint subcells, each of which is made active. We will discuss termination conditions later.

Let us consider the elements of the algorithm in somewhat greater detail. For each cell T that we process, we are interested in the transformation of this cell that has the smallest similarity measure. We compute an upper bound $sim_{hi}(T)$ and a lower bound $sim_{lo}(T)$ on this smallest similarity (discussed below). For each upper bound, there will be a transformation that serves as a witness to this upper bound. The algorithm maintains the best similarity sim_{best} of that it has encountered thus far in the search, and the associated transformation t_{best}. To process a cell, we first compute its lower bound. If the lower bound exceeds sim_{best} then we kill this cell. Otherwise, we compute the upper bound, and if it is less than sim_{best} then we update the the current best.

How are the upper and lower bounds computed? To compute an upper bound, we may sample any transformation from within the cell. In our implementation this is done by simply taking the midpoint t of the cell, and then computing $sim_q(t)$. In particular, this involves computing the image of each point of A under t, and then computing the nearest neighbor in B for each image point. Since the cell's optimum similarity is smaller than any sampled point, this is an upper bound on the minimum similarity of the cell. Nearest neighbors are computed by storing the points of B in a kd-tree data structure, and applying known efficient search techniques [FBF77, AM93, AMN+94].

To compute the lower bound, we apply the same general method used by Huttenlocher and Rucklidge. Given any cell $T \subset \mathcal{T}$, and given any point $a \in A$, consider the image of a under every $t \in T$. It is an easy matter to compute a bounding rectangle on this set. For example, if the coordinates of a are nonnegative, then this is the rectangle defined by the corner points $t_{lo}(a)$ and $t_{hi}(a)$. (The general formula involves a case analysis on the signs of the coordinates of a.) We call this bounding rectangle the ***uncertainty region*** of a relative to cell T. In this way, each cell is associated with a collection of uncertainty regions, one for each point of A. (See Fig. 1.)

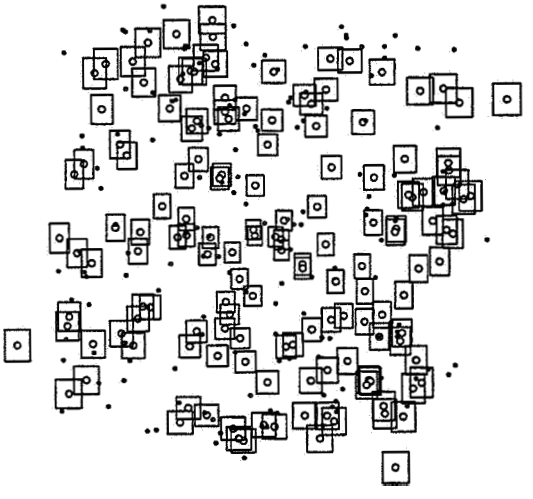


Figure 1: Uncertainty regions. The points of A are shown in white, each transformed by the midpoint of the cell, the points of B are shown in black, and the uncertainty regions are shown as rectangles.

Define the distance between an uncertainty region and a point $b \in B$ to be the minimum distance between b and any part of the uncertainty region. If b lies inside the uncertainty region, then the distance is zero. To derive our lower bound for T, we compute the distances from each uncertainty region to its closest point in B, and take the qth smallest such distance. Call this $sim_{lo}(T)$. This can be done by a straightforward generalization of the kd-tree-based nearest neighbor method described above [AM95]. Unlike Huttenlocher and Rucklidge [HR93, Ruc95], who perform nearest neighbor searching at image precision through a distance transform, our method can compute distances to machine precision, and require auxiliary space that is proportional to the number of points in B, rather than the image size.

Observe that the image of a point $a \in A$ under any transformation $t \in T$ lies within the corresponding uncertainty region, and hence this distance is a lower bound on the distance from $t(a)$ to its nearest neighbor in B. This implies that $sim_{lo}(T) \leq sim_q(t)$, for any $t \in T$. In other words, this is indeed a lower bound on the optimum similarity in this cell.

How are cells selected for processing and how is splitting performed? Here we differ from Huttenlocher and Rucklidge's simple rasterization method. In particular, we split each cell into two subcells (rather than $2^6 = 64$ subcells), and choose the cell to be processed and the splitting dimension so as to reduce, as much as possible, the uncertainty. Define the ***size*** of an uncertainty region to be its longest side. Define the ***size*** of a cell to be largest size of any of its uncertainty regions. (Note that cell size is not defined in terms of the size of the hyperrectangle in transformation space.) Our implementation actually computes a fast upper bound on this size, by first computing the bounding box for the points of A, and then computing the maximum size of the uncertainty region generated by any point in this bounding box.) The active cells are stored in a priority queue, ordered by cell size. The largest cell, that is, the cell with the largest uncertainty region, is chosen for processing at each stage.

If a cell survives processing, then it is split. Consider a cell T defined by a lower and upper bound range (t_{lo}, t_{hi}). We determine which of the six components of the transformation range contributes most to the size of the uncertainty region, and we split the cell through its midpoint along this dimension, To be precise, first consider the transformation $A = t_{hi} - t_{lo}$, formed by taking the component-by-component differences of these transformations. By the linearity of our representation, the size of the uncertainty region of a point a is the longer side of the vector $\Delta(a)$. The x- .and y-coordinates of this vector are

$$x = \Delta_{xx} a_x + \Delta_{xy} a_y + t_x,$$
$$y = \Delta_{yx} a_x + \Delta_{yy} a_y + t_y.$$

We define the component that contributes most to the size of a's uncertainty region to be the largest in absolute value among the six terms consisting of the $\Delta_{ij} a_j$'s and the t_i's. This is for a single point a. To define the component that contributes most to the size of an entire cell T, we take the maximum component over all points a. In our implementation, this is bounded by taking the the maximum over all points lying in the bounding box of A. Because there are six dimensions that affect the size of each uncertainty region, and given the linearity of the transformation space, we have the following.

Lemma 1

(i) *After six consecutive splits, the size of a cell (that is, the maximum side length of its largest uncertainty region) decreases by at least half.*

(ii) *If there are currently m active cells, then after at most 64m rnstances of cell processing, the size of the largest cell will decrease by at least half.*

This completes the description of the exact algorithm. We modify this to an approximation algorithm, as follows. First, we compute the upper bound $sim_{hi}(T)$ on the similarity of a cell using the weak quantile q', rather than the true quantile q. Thus we have $sim_{hi}(T) \geq sim_{q'}(t)$, for any $t \in T$. Second, the condition for killing a cell is not that the cell's lower bound be greater than the current best, but that it not be significantly lower than the current best. Given a cell T with lower bound $sim_{lo}(T)$, it is killed if either:

$$sim_{lo}(T) > \frac{sim_{best}}{1+\epsilon_r} \quad \text{or} \quad sim_{lo}(T) > sim_{best} - \epsilon_a$$

The algorithm terminates when all cells have been killed or when $sim_{best} < \epsilon_a$.

Here is an overview of the approximation algorithm. The input consists of the point sets A and B, the Hausdorff quantile q, the approximation parameters ϵ_r, ϵ_a, and ϵ_q, and the initial cell T_0.

(0) Initialize the priority queue to contain T_0. Set $sim_{best} = \infty$. Define the weak quantile q' to be $q(1-\epsilon_q)$. Repeat the following steps until the priority queue is empty or until $sim_{best} \leq \epsilon_a$. Build a kd-tree for the points of B

(1) Remove the largest cell T from the queue. Compute its lower and upper bounds. This involves the following steps:

  (a) Compute the uncertainty regions, for every point $a \in A$ with respect to T

  (b) For each uncertainty region, compute its nearest neighbor in B (where the distance from a point to a region is defined to be the shortest distance between the point and any point in the region).

  (c) Compute (using any fast selection algorithm) the qth quantile among these distances. Call this $sim_{lo}(T)$. If $sim_{lo}(T) > sim_{best}/(1+\epsilon_r)$, then kill this cell, and return to step (1).

  (d) Otherwise, sample an allowable transformation t from this cell. (At this point we may also apply some additional allowability conditions on the transformation.)

  (e) Compute the image of each point of A under t, and compute the nearest neighbors of these points with respect to B. Find the q'th smallest such distance, Call this $sim_{hi}(T)$.

(2) If $sim_{hi}(T) < sim_{best}$, then update sim_{best} and let t_{best} be the associated transformation.

(3) Split T into two smaller subcells T_1 and T_2, by splitting it along the dimension that contributes most to its uncertainty region size. Compute size bounds for T_1 and T_2

(4) Enqueue T_1 and T_2 in the queue of active cells.

The correctness of the algorithm is established by the following theorem

Theorem 1 *Assuming that $\epsilon_a > 0$ or $\epsilon_r sim_{opt} > 0$, then the this approximation will eventually terminate, and upon termination t_{best} satisfies the approximate correctness conditions given above.*

Proof First we show that the algorithm eventually terminates. Recall that the initial cell is bounded, and from Lemma 1 it follows the maximum cell size (the longest side of any uncertainty region) will fall below any given threshold after a sufficiently large number of cells have been processed,

We claim that when the size (the longest side length) of the largest active cell falls below

$$\frac{1}{\sqrt{2}} \max\left(\epsilon_a, \frac{\epsilon_r}{1+\epsilon_r} sim_{opt})/2\right),$$

then the algorithm terminates. Observe that under the hypotheses of the theorem this value is positive, and so the above scenario will eventually happen, Since the diameter of each uncertainty region is at most $\sqrt{2}$ times its size, it follows that when the above event takes place, the uncertainty falls below the max() term above, implying that the difference $sim_{hi}(T) - sim_{lo}(T)$ becomes smaller than this threshold. Also, since we sample a transformation from each cell, it follows that $sim_{best} \leq sim_{hi}(T)$ Thus we have

$$\begin{aligned} sim_{best} - sim_{lo}(T) &\leq sim_{hi}(T) - sim_{lo}(T) \\ &< \max\left(\epsilon_a, \frac{\epsilon_r}{1+\epsilon_r} sim_{opt}\right) \\ &\leq \max\left(\epsilon_a, \frac{\epsilon_r}{1+\epsilon_r} sim_{best}\right) \end{aligned}$$

There are two cases. If the maximum is achieved by ϵ_a, then we have

$$sim_{lo}(T) > sim_{best} - \epsilon_a$$

On the other hand, if the maximum is achieved by the other term, then we have

$$sim_{lo}(T) > \frac{sim_{best}}{1+\epsilon_r}$$

In either case, T will satisfy one of the two conditions needed to kill it. Thus, no cell can survive beyond this point, and the algorithm will terminate.

To show the correctness of the algorithm upon termination, recall that the algorithm has two termination conditions. First, if $sim_{best} \leq \epsilon_a$, then since the optimum similarity is nonnegative, t_{best} must be within the absolute error bound. On the other hand, suppose that the algorithm terminates because all the cells are killed. Consider the moment at which the cell T containing t_{opt} is killed Then either

$$\begin{aligned} (1+\epsilon_r) sim_{opt} &= (1+\epsilon_r) sim_q(t_{opt}) \\ &\geq (1+\epsilon_r) sim_{lo}(T) > sim_{best}, \end{aligned}$$

implying that t_{best} is (relatively) approximately optimal, or

$$sim_{opt} + \epsilon_a = sim_q(t_{opt}) + \epsilon_a \geq sim_{lo}(T) + \epsilon_a > sim_{best},$$

implying that t_{best} is (absolutely) approximately optimal. Hence when the algorithm terminates, all the cells — including the optimal cell — have been killed, thereby implying that the reported transformation is approximately correct. □

In Section 4 we will give empirical evidence that approximation can significantly speed up the running time of the branch-and-bound algorithm at a relative modest cost to the quality of the resulting match.

3 Accelerating using Pinning

Even though approximation can speed up search times in many instances, there are many other cases where the resulting running time is still very large. In particular, this happens when the initial search range is large and spans all six dimensions, and the optimum match has a very small similarity measure. The problem seems to be that the algorithm needs to apply a large number of splits to decrease the cell size small enough so that the killing criterion can be satisfied. For this reason, we introduce a process called *pinning* to help accelerate the search. Throughout this section we assume that the transformation space is the 6-dimensional space of affine transformations. Generalizations to other transformation spaces are not difficult.

Before discussing what pinning is, let us consider why the approximation algorithm may run slowly, and how it could

be sped up. It was argued in the proof of Theorem 1 that the algorithm terminates when the cell sizes decrease to around $\max(\epsilon_a, \epsilon_r sim_{opt})$. If this quantity is small compared to the initial uncertainty bounds, then the algorithm may require a great deal of splitting before achieving this size. Of course, not all cells need to be split to this small size, but at the very least, observe that if the uncertainty region of a cell contains even a single point of B, then this region contributes a distance of zero to the set of distances used in computing the lower bound for this cell. Thus to even get a nonzero lower bound, a significant fraction (meaning at least q) of uncertainty regions must be empty. As cell sizes shrink, before we get to a point that many cells have nonempty uncertainty regions, we will reach a stage where many cells have uncertainty regions with a single point. We claim that at this stage, it is already possible to derive nontrivial lower bounds, and hence to kill cells. This is the main idea behind pinning.

Let us think of the points of A as being partitioned into two classes, ***inliers,*** i.e., points that are mapped by the optimum transformation to lie within distance sim_{opt} of their nearest neighbor in B, and ***outliers,*** i.e., all of the remaining points of A. By the definition of sim_{opt} as a q-quantile Hausdorff distance, at least a fraction q of points of A are inliers. (Normally the concept of an inlier is based on the underlying point distribution, but this functional definition is sufficient for our purposes.) To simplify the explanation, suppose for the moment that the inliers are completely free of any error, that is, the q-quantile Hausdorff distance is zero for the optimum transformation, t_{opt}. If we were able to find three inliers of A and match them against the corresponding elements of B, and further — if these three points of A were affinely independent (i.e., not collinear) — then these three point correspondences would uniquely determine the affine transformation that achieves a Hausdorff distance of zero. The process whereby triples from A are matched against prospective corresponding triples from B in order to determine a candidate transformation is called ***pinning.***

In principle, the pinning operation is similar to point matching methods that are based on point to point correspondence. This notion is fairly common, as was briefly mentioned in the 1 Of course, in noisy environments it is unreasonable to assume that points could be matched free from any error or that the numerical extraction of the affine transformation would be error free. To generalize this to noisy environments, suppose that for each inlier $a \in$ A there is a point of B that lies within some small distance σ of its optimum image $t_{opt}(a)$. We assume that an upper bound on σ is provided to the search algorithm. For registration applications, such an estimate would be based on knowledge of the accuracy of the feature extraction process, under the assumption that a feature point does indeed exist. The final Hausdorff distance will be no larger than σ If σ is small relatively to the size of the image, and if the three points are 'geometrically well distributed", then we would expect that the match that results from pinning should be fairly close to the optimum. Intuitively, a geometrically well distributed triple should define a large, "fat" triangle among the points of A. To make this more formal, we present the following definition.

Definition: Given a set of points A in the plane and $a \geq 1$, a triple $\{a_1, a_2, a_3\} \subseteq A$ is ***well-distributed*** relative to a if every point $a \in$ A can be expressed as an affine combination $a = \alpha_1 a_1 + \alpha_2 a_2 + \alpha_3 a_3$, such that, $|\alpha_i| \leq a$.

Given this definition, the following lemma follows easily from the triangle inequality.

Lemma *2 Given a planar point set A, a well-distributed triple $\{a_1, a_2, a_3\}$ relative to some $a > 1$, and an affine transformation t that maps each of the points a_i to a point that lies within distance σ of a_i, then for any point $a \in A$, the distance from $t(a)$ and a is at most $6a\sigma$*

This lemma implies that if pinning is applied to a well-distributed triple of points and if the images of the pinned points are perturbed by a small amount σ, then the perturbation induced by the resulting transformation will be proportional to σ

In general, the number of subsets that may have to be tested may be quite large. However, branch-and-bound search provides a particularly simple way of identifying the most promising correspondences to try. In particular, as the search proceeds and cells are split, the search arrives eventually at a stage at which a significant fraction of uncertainty regions will contain at most a single point of B An uncertainty region is said to be *"pinnable"* if there is at most one point of B in the region, or if the region is empty and there is at least one point of B within distance σ of the region. If the current cell has a significant fraction of pinnable uncertainty regions, then we say that this cell is ***eligible*** for ***pinning.*** These constant factors are left unspecified for now (See Fig. 2.)

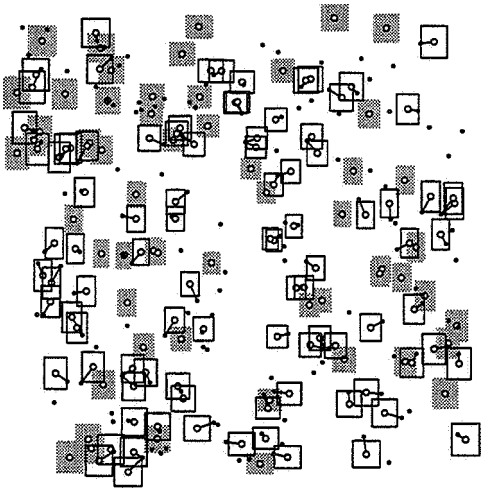


Figure 2: Pinnable uncertainty regions. The transformed points of A are shown in white, points of B are shown in black. Pinnable uncertainty regions are shown with solid lines, and unpinnable uncertainty regions are shaded. For each pinnable uncertainty region, a line segment joins the associated point of A with its nearest neighbor in B

Consider the cell containing the optimum transformation. If it is eligible for pinning, then for each of its pinnable uncertainty regions, there is a single obvious candidate for the corresponding point of A to be mapped to. That is, if the uncertainty region is nonempty, then there is a point of B lying within the uncertainty region, and otherwise it is the nearest neighbor of the uncertainty region. Further, if $a \in A$ is an inlier, then the nearest neighbor in B should lie within distance σ of a, and hence within σ of the uncertainty region.

To apply pinning we need to make the assumption that the uncertainty regions of outliers are not more likely to be

pinnable than inliers. (Or at least, if this assumption is violated, then we should have a bound on the difference in the probabilities.) Under this assumption, if we randomly sample three pinnable uncertainty regions of A, then with probability q^3, all three points will be inliers. Thus, the three respective nearest neighbors of B will be within distance σ from each of these points. Further, it follows from Lemma 2 that if these points of A are well-distributed, then the pinning transformation that results from considering this correspondence will be a reasonable approximation to the optimum, in the sense that the Hausdorff distance should not exceed the optimum Hausdorff distance by an amount greater than a constant c times σ (as determined by Lemma 2).

Thus, under the above made assumptions, and depending on q, if we take a sufficiently large number N of random samples, then the probability of generating a pinning transformation whose distance from optimum is more than $c\sigma$ is at most $(1-q^3)^N$ (To see this, observe that $(1-q^3)$ is the probability that all three sampled points of A are outliers, and for failure this event must be repeated N times.) To reduce the chances of failing to find such a good triple to some small failure probability, p_f, the number of well-distributed samples should satisfy

$$N(q, p_f) \geq \frac{\ln p_f}{\ln(1-q^3)}$$

For example, to reduce the probability failure below 0.01, and for $q = 0.5$, roughly 35 sampled triples are sufficient. Note that this constant — although large — is independent of the number of feature points in either of the images.

If on the other hand, after $N(q, p_f)$ well-distributed sampled triples we find that all exceed the current best similarity by at least $c\sigma$, then we conclude that this cell does not contain the optimum transformation, and kill it. This is the basis of pinning. Observe that this is a Monte Carlo process, in the sense that with probability p_f we may erroneously kill the cell containing the optimum transformation

Now let us describe the pinning process in greater detail. First, we need make the following assumption upon which pinning relies.

Pinning Assumption: For a cell containing the optimum transformation, the probability that an uncertainty region associated with this cell is pinnable, and the probability that a pinnable triple is well-distributed is not greater for outliers than for inliers.

As we mentioned above, it is not necessarily realistic to assume that this assumption holds. But if it is possible to estimate the difference in the conditional probabilities for inliers and outliers, and these probabilities are not significantly different, then the number of samples derived below can be modified accordingly.

We can incorporate pinning into the existing branch-and-bound approach as follows. We supply the algorithm with three additional pieces of information, i.e., the expected perturbation σ between inliers and their corresponding points in the final match, and two quantities called the sampling quantile q_s and a minimum sample size N_s. First, we augment the procedure that computes the distance from each uncertainty region to its nearest neighbor to compute also the number of points of B that lie within the uncertainty region. This can be done by modifying the kd-tree nearest neighbor algorithms [AM95]. If this number is at most one, and the nearest neighbor is within σ of the uncertainty region, then we flag this region as being pinnable.

After processing all the uncertainty regions, if the fraction of pinnable uncertainty regions exceeds q_s, then we declare that the cell itself is eligible for pinning. If so, we sample N_s triples of points from A, such that each triple is well-distributed (relative to some given constant). We generate the pinning transformation mapping each sampled triple to the nearest neighbor of the corresponding uncertainty region, and compute the similarity of this transformation. If the similarity of all the N_s transformations exceeds the current best sim_{best} by an additive amount of $c\sigma$, then we kill this cell. Otherwise, the cell remains active, and the normal processing resumes.

Since we test the similarity of each sampled transformation, if we find that any sampled transformation is better than the current best (and the transformation satisfies the additional allowability conditions), then we take this transformation to be the current best. Thus pinning not only helps in improving lower bounds, but also helps in generating better upper bounds.

Our implementation differs slightly from this description We do not check that the points are well-distributed. Instead we apply a somewhat more precise test. We generate the pinning transformation and then test whether this transformation lies within or is sufficiently close to the current cell. The definition of "sufficiently close" is that the translational component should be within some fixed constant factor of σ, and the entries of the matrix M should be within a constant factor of the ratio of σ to the diameter of the point set A. Our reasoning is that points do not need to be well-distributed, as long as they generate a transformation that is reasonably "true" to the cell. Because of its Monte Carlo nature, the fear in pinning is that the cell containing the optimum may be killed accidentally. If the current cell contains the optimum transformation, then we expect that the transformation generated by a triple of inliers should either lie inside or close to the cell. If not, then we conclude that the triple was unacceptable. Because it is easy to test this condition (once the transformation has been generated), we generate triples until this condition is satisfied. If we fail after a repeated number of tries (10 in the current implementation), then we treat this sample as a failure.

Under what circumstances does pinning offer an advantage over pure branch-and-bound? At this point we do not have a theoretical analysis or even an exhaustive empirical analysis explaining when this happens. However, based on our empirical experiences with the algorithm on uniformly distributed point sets (see Section 4), in many instances pinning can offer significant improvements in running time over the pure branch-and-bound search. In particular, we have observed that pinning seems to offer the greatest advantage when the deviation in the location of feature points σ is relativey small with respect to the expected distance 6 from a random point and its nearest neighbor in B.

We attempt to provide here some justification for this empirical observation. Assume that the points of B are uniformly distributed, and for simplicity suppose that the center of each uncertainty region is sufficiently random that the expected distance to the nearest point of B is 6. When the sizes of the uncertainty regions decrease to roughly δ, then pinning becomes possible, because many uncertainty regions of the cell containing the optimum transformation should contain at most one point of B Also note that at this point, the current best similarity is expected to be pro-

portional to δ. This is because we cannot expect to locate the optimum much more precisely than the cell size (unless we just happen to be very lucky in sampling a transformation). However, by the reasoning made earlier in this section, once we reach this stage, and pinning can be applied to the optimum cell, we expect to discover a transformation whose similarity is not greater than $sim_{opt} + c\sigma$ Recall that $sim_{opt} \leq \sigma$, and so sim_{best} is now proportional to the optimum. Thus, if the ratio δ/σ is large, then the pinning algorithm gains a big advantage as the tighter upper bound makes it possible to kill more unpromising cells.

Also consider the conditions needed to kill a nonoptimal cell. Assuming, as before, that sim_{best} is proportional to S at this stage of the search, then (considering just the relative error) a cell T cannot be killed until

$$sim_{lo}(T) > \frac{sim_{best}}{1+\epsilon_r} \gtrsim \frac{\delta}{1+\epsilon_r}.$$

The distance of an uncertainty region to its nearest neighbor in B is expected to be at most 6 minus the radius of the uncertainty region It follows from simple algebra that if the cell's size is larger than $2\epsilon_r\delta$, then this condition is not likely to be satisfied by most of the cell's uncertainty regions. But pinning already applies when the cell's size is roughly 6. Thus the pure branch-and-bound approach must continue to split the cell until its size is reduced by an amount proportional to $2\epsilon_r$ before it can be killed. By Lemma 1, this may require $6\log_2(1/2\epsilon_r)$ additional splits, and could result in the generation of $(1/2\epsilon_r)^6$ additional cells. For even moderately small values of ϵ_r, this value may be significantly larger than the N_s samples generated by pinning, and hence pinning may offer a significant improvement over pure branch-and-bound.

4 Empirical Study

We implemented both the approximation algorithm and the modification using pinning in the C++ programming language. We also implemented a driving program to run a series of tests, and to gather statistics about the progress and performance of the two algorithms. In order to perform nearest neighbor queries we used the ANN approximate nearest neighbor library [MA97] to generate a kd-tree for the points of B (where nearest neighbors are computed exactly), and to compute rectangular range queries we implemented an extension to this library, The program can either read input sets from a file or generate inputs randomly from a number of distributions.

For our experiments, the point sets were generated as follows. First the user specifies the number of points in the sets A and B, bounding rectangles for A and B, the desired fraction of inliers of A, and the standard deviation σ of a random perturbation to be applied to each image point. (Here σ is not the same as the upper bound σ of the previous section, although the two are related.) Then the program generates an affine transformation t at random from a collection of bounds on the x and y translation, rotation, x and y scale, and shearing. The program then generates the desired number of uniformly distributed inliers as follows. A point a is generated uniformly at random from A's bounding rectangle. For each such point a, it computes $t(a)$, adds a Gaussian error with mean zero and standard deviation σ, and then tests whether the resulting point lies within the bounding rectangle of B. If so, the point is accepted This is repeated until the desired number of inliers are generated

The remainder of point sets A and B are filled out with uniformly distributed points in their bounding rectangles. The given transformation t is called the ***target transformation.*** It will generally not be the optimum transformation (unless $\sigma = 0$), but should be close enough to the optimum that we may use it for validating the results of the experiments.

The user supplies the desired quantile q for the Hausdorff distance, the error factors ϵ_r, ϵ_a, and ϵ_q, and the bounds on the initial cell. If pinning is being applied we also supply two additional parameters, q_s, the fraction of pinnable uncertainty regions needed invoke pinning for a cell, and N_s, the number of well-distributed triples to sample from the pinnable regions. The program can be run using either the "pure" branch-and-bound approximation algorithm described in Section 2, or with "pinning" as described in Section 3.

We also add a feature that was hinted at in the previous sections. Even though we consider the entire space of affine transformations, it is often desirable to restrict the search to a particular subspace. We wanted to restrict the search to transformations that are relatively rigid, in that they should consist primarily of translation and rotation, and only allow a small amount of scaling and shearing. This was done by implementing additional ***allowable conditions*** on the transformations. In our case, this was done by requiring that the determinant and norm of the transformation matrix M be close to 1 In particular, in our experiments we required that they both lie within the interval specified as part of the input, e.g., $[0.95, 1.05]$. Any sampled transformation that did not satisfy these requirements was immediately rejected, without computing its similarity. Furthermore, if it could be established that every transformation in a cell does not satisfy this requirement, then the entire cell is eliminated.

The program measures a number of statistics on its own performance, including CPU seconds, the number of cells expanded, and the number of times the Hausdorff-based similarity was computed. In addition, the program computes a quantity called the *effective error* Normally, this should be done by comparing the algorithms performance against the optimum, but since we do not know the optimum transformation, we measured the error against the target transformation, which was used to generate the data sets. The algorithm computes the similarity of the target transformation sim_{targ} and compares it against the similarity returned by the program sim_{best}. Both are computed relatively to the weak quantile $q(1-\epsilon_q)$. The effective error is defined to be the minimum of the relative error of sim_{best}

$$e_r = \frac{sim_{best} - sim_{targ}}{sim_{targ}},$$

and the absolute error

$$e_a = sim_{best} - sim_{targ}.$$

Note that because sim_{targ} is not the optimum, it is actually possible for the algorithm to return a (typically small) negative effective error For the approximation algorithm, the effective error cannot exceed $\max(\epsilon_r, \epsilon_a)$. For the Monte Carlo pinning algorithm this might happen only in the unlucky (and unlikely) event that the optimum cell was killed.

We have concentrated thus far on two main experiments. The purpose of the first experiment was to study the performance of the algorithm(s) as a function of the relative error ϵ_r, whereas the purpose of the other experiment was to study the performance(s) as a function of the perturbation σ In both experiments, all of the datasets consisted of 100 points,

50 of which were inliers. The model points were generated (at random) inside a bounding box of size 200×200, and the object points were generated, as was explained previously, by applying a target transformation (picked at random) to the object points and adding perturbation The bounds for the target transformation were [-90, +90] for rotation, and [-25, +25] for translation in both x and y directions. Also, the initial cell bounds provided in all cases were [-3, +2] for rotation and [-5, +4] for both translations. Finally, we fixed $\epsilon_a = 0.1$, $\epsilon_q = 0$, $q_s = 0.2$, and $N_s = 15$ throughout the experiments. (The latter corresponds to a failure probability of approximately 10% in the case of pinning.)

For the first experiment, we fixed the perturbation at $\sigma = 1$, and for each value of ϵ_r experimented with 10 different data sets. Similarly, for the second experiment, we fixed $\epsilon_r = 0.1$, and for each value of σ experimented with 20 different data sets. Both pure branch-and-bound search and pinning were invoked for each data instance. Running time summaries of the two experiments are provided in Figs. 3, 4 below (Each reading is the median value of the running times recorded for an entry in question The running times themselves vary greatly by nature of the algorithm(s).)

From the graphs of Fig. 3 we observe that pinning can be advantageous over a wide range of ϵ_r values. This is especially true for small values of ϵ_r However, as ϵ_r increases, the running time of the pure algorithm drops significantly. Interestingly, though, the effective epsilon observed for the pure method is much smaller than the required relative error In fact, in many instances the effective epsilon is actually negative, even for ϵ_r as large as 10. (Complete graphs will be provided in [MNM97].) From the graphs of Fig. 4 we again observe that pinning can be significantly advantageous, provided that the perturbation is relatively small. For large σ, both methods seem to perform comparably (The artifact that occurs near $\sigma = 0.2$ is not clear at this point and will be studied further-) Finally, we note that 19 instances (out of 240) were recorded, where pinning failed to report a correct approximation, i.e., the failure probability was slightly better than 10%.

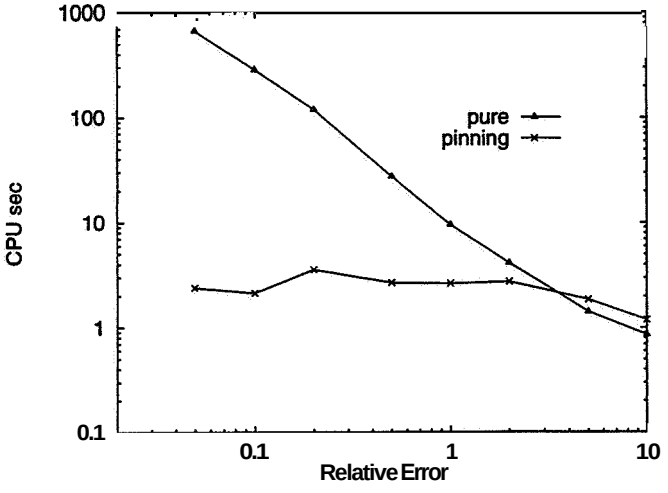


Figure 3: Execution time of pure branch-and-bound as a function of relative error, ϵ_r.

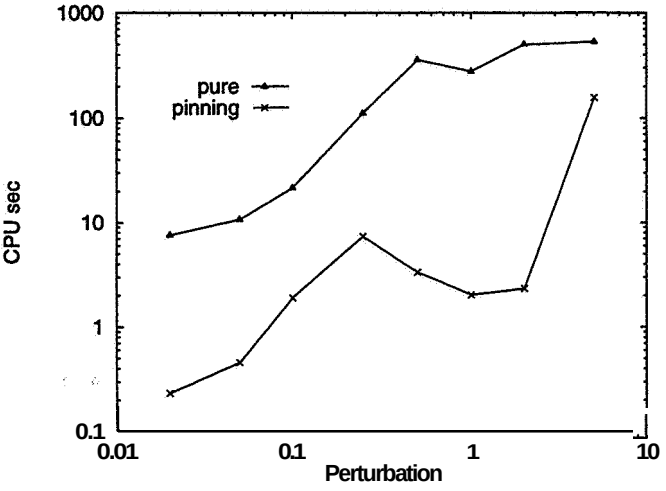


Figure 4: Execution time of pure branch-and-bound and pinning as a function of the perturbation σ

5 Discussion

In this paper we have presented two algorithms for robust feature matching. Following the framework of Huttenlocher et al., we derived a branch-and-bound approximation algorithm and a faster randomized variant, called pinning. In addition, we have presented promising preliminary results (on synthetic data).

Future research will include a more elaborate empirical study of the performance of the algorithms and of their suitability to remotely sensed data. In particular, we will look into the possibility of incorporating the algorithms in a recently developed multiresolution, wavelet-based registration scheme [LeM95]. Finally, in order to further reduce running times, we will derive variants for special cases of 3-dimensional and 4-dimensional transformations.

Acknowledgements

The second and third authors would like to thank William J Campbell and Robert F Cromp of Code 935, NASA/GSFC, for their ongoing support and for sustaining the activities of the image registration group at NASA/GSFC.

References

[AAR94] H. Alt, O. Aichholzer, and G. Rote. Matching shapes with a reference point. In Proc. **10th Annu. ACM Sympos. Comput. Geom.,** pages 85–92, 1994.

[AM93] S. Arya and D. M Mount. Algorithms for fast vector quantization. In J A. Storer and M. Cohn, editors, Proc. of ***DCC '93:*** **Data Compression Conference,** pages 381–390. IEEE Press, 1993.

[AM95] S. Arya and D. M Mount. Approximate range searching. In Proc. **11th Annu. ACM Sympos. Comput. Geom.,** pages 172–181, 1995.

[AMN+94] S. Arya, D M. Mount, N S. Netanyahu, R. Silverman, and A. Wu. An optimal algorithm for approximate nearest neighbor searching. In

Proc, 5th ACM-SIAM Sympos. Discrete Algorithms, pages 573–582, 1994.

[AMWW88] H. Alt, K. Mehlhorn, H Wagener, and E. Welzl. Congruence, similarity and symmetries of geometric objects. ***Discrete Comput. Geom.,*** 3:237–256, 1988.

[Bro92] L. G. Brown. A survey of image registration techniques. ***ACM Computing Surveys,*** 24:325–376, 1992.

[CGH+97] L. P Chew, M. T Goodrich, D. P Huttenlocher, K. Kedem, J. M Kleinberg, and D. Kravets. Geometric pattern matching under Euclidean motion. ***Comput. Geom. Theory Appl.,*** 7:113–124, 1997

[DH83] D L. Donoho and P J Huber. The notion of breakdown point. In P J. Bickel, K. Doksun, and Jr J L. Hodges, editors, ***A Festschrift for Erich L. Lehman,*** pages 157–184. Wadsworth, Belmont, California, 1983.

[dRL95] P J de Rezende and D. T Lee. Point set pattern matching in d-dimensions. ***Algorithmrca,*** 13:387–404, 1995.

[FBF77] J H. Friedman, J L. Bentley, and R. A. Finkel. An algorithm for finding best matches in logarithmic expected time. ***ACM Transactions on Mathematical Software,*** 3:209–226, 1977

[GS85] A. Goshtasby and G. C Stockman. Point pattern matching using convex hull edges. ***IEEE Transacttons on Systems, Man, and Cybernetic c*** ~1,5:631–637, 1985.

[GSP86] A. Goshtasby, G. C. Stockman, and C. V Page. A region-based approach to digital image registration with subpixel accuracy. ***IEEE Transactions on Geoscience and Remote Sensing,*** 24:390–399, 1986.

[HKR93] D. P Huttenlocher, G A. Klanderman, and W J Rucklidge. Comparing images using the Hausdorff distance. ***IEEE Transactions on Pattern Analysis and Machine Intelligence,*** 15:850–863, 1993.

[HR92] D. P Huttenlocher and W J. Rucklidge. A multi-resolution technique for comparing images using the Hausdorff distance. Technical Report 1321, Dept. of Computer Science, Cornell University, Ithaca, NY, 1992.

[HR93] D. P Huttenlocher and W J Rucklidge. A multi-resolution technique for comparing images using the Hausdorff distance. In ***Proc. IEEE Conf. Comput. Vision Pattern. Recogn.,*** pages 705–706. IEEE, 1993.

[HV97] M. Hagedoorn and R. C. Veltkamp. Reliable and efficient pattern matching using an affine invariant metric. Technical Report RUU-CS-97-33, Dept. of Computing Science, Utrecht University, The Netherlands, 1997

[IR96] S. Irani and P Raghavan. Combinatorial and experimental results for randomized point matching algorithms. In ***Proc. 12th Annu. ACM Sympos. Comput. Geom.,*** pages 68–77, 1996.

[LeM95] J LeMoigne. Towards a parallel registration of multiple resolution remote sensing data. In ***Proc. Int'l Geo. and Remote Sensing Sympos.*** IEEE, 1995.

[MA97] D. M. Mount and S. Arya. Ann: A library for approximate nearest neighbor searching. Unpublished manuscript, to be presented at the CGC 2nd Annual Fall Workship on Computational Geometry, 1997

[MNM97] D. M. Mount, N S. Netanyahu, and J Le Moigne. Efficient algorithms for robust feature matching. Full paper version, in preparation, 1997

[OH97] C. F Olson and D. P Huttenlocher. Automatic target recognition by matching oriented edge pixels. ***IEEE Transactions on Image Processing,*** 6:103–113, 1997

[RL87] P J Rousseeuw and A M. Leroy. ***Robust Regression and Outlier Detection.*** Wiley, New York, 1987

[RR80] S. Ranade and A. Rosenfeld. Point pattern matching by relaxation. ***Pattern Recognition,*** 12:269–275, 1980.

[Ruc95] W J Rucklidge. Locating objects using the Hausdorff distance. In ***Proc. 5th Int'l Conf on Computer Visron,*** pages 457–464. IEEE, 1995.

[Ruc96] W J Rucklidge. ***Efficient visual recognition using the Hausdorff distance.*** Number 1173 in Lecture Notes in Computer Science. Springer-Verlag, Berlin, 1996.

[SKB82] G. C. Stockman, S. Kopstein, and S. Benett. Matching images to models for registration and object detection via clustering. ***IEEE Transactions on pattern Analysis and Machine Intelligence,*** 4:229–241, 1982.

[TJ89] J. Ton and A. K. Jain. Registering landsat images by point matching. ***IEEE Transactions on Geoscience and Remote Sensing,*** 27:642–651, 1989.

[XMT+97] W Xia, J. Le Moigne, J. C Tilton, B-T Lerner, E. Kaymaz, J Pierce, S. Raghavan, S. Chettri, T El-Ghazawi, M. Manohar, N S. Netanyahu, W J Campbell, and R. F Cromp. A registration toolbox for multi-source remote sensing applications. In ***Proc. Int'l Conf. on Earth Observation and Environmental Information,*** 1997

Effects of Lossy Compression on Digital Image Registration

Anthony J. Maeder
School of Engineering, University of Ballarat,
P.O. Box 663, Ballarat 3353, Victoria, Australia

Keywords: image registration, image matching, stereo disparity, lossy compression, **JPEG.**

ABSTRACT

Area-based, feature-based and various hybrid approaches for image matching to obtain spatial correspondence between images, form the basis of many image registration techniques. If image data is degraded prior to the matching calculations, the accuracy of the registration will be affected. Some experiments using ***JPEG*** compressed data under different matching conditions are reported. ***An*** indication is obtained that lossy compression rates of up to 10:1 may not cause significant reduction in registration accuracy

1. INTRODUCTION

Digital image registration is important in many different applications areas (e.g. bio-medical, geographical, surveillance) where the physical correspondence of information contained in a set of related images must be established in order to solve further more complex image analysis problems (e.g. measurement, recognition). Some common generic tasks are the fusion of images acquired from multiple imaging modalities, the spatial alignment of moving objects in a temporal image sequence, construction of image mosaics and panoramas, and the extraction of **3D** digital surface models from overlapped images via stereo disparity calculations.

The fundamental basis of registration can be expressed very simply in the discretized digital image domain. Given two images I_1 and I_2, each consisting of a 2D discretized rectilinear grid of points (referred to as (u_1,v_1) and (u_2,v_2) respectively) representing pixel intensity values, and given a correspondence function $C((u_1,v_1),(u_2,v_2))$, which indicates how closely the I_1 image information at position (u_1,v_1) compares with that in I_2 at position (u_2,v_2), we seek in general a transformation R that maps each pixel position in image I_1 to an ideal pixel position in I_2 i.e.

R $(u_1,v_1) \rightarrow (u_2,v_2)$
iff $C((u_1,v_1),(u_2,v_2))$ exists and maximized over (u_2,v_2);
undefined otherwise

The set of all R values (possibly interpolated where they do not exist) constitutes a disparity map, from which can be derived a digital elevation map (DEM) representing the 3D surface height.

Due to physical constraints of image capture and display, pixel intensities represent information for the immediate area around each grid position, rather than providing an accurate infinitesimal point value. Thus an alternative formulation for the registration transformation might define the position (p,q) of the centre of such an area, not coincident with any point on the discretized grid of I_2 pixel positions. Equivalently a neighbourhood of adjacent pixel positions $N(u_2,v_2)$ might be defined, to which the pixel at (or indeed $N'(u_1,v_1)$) *is* mapped with a weighting described by some appropriate point-spread function or similar formula.

Registration transformations may be obtained in various ways and may take widely differing forms. If the nature of the optical and physical distortions which occur during acquisition of the images is known, systematic analytical or numerical inversion of these distortions ***can*** be undertaken. This approach can be successful even if the distortions are only approximated (e.g at low orders), or if the family of possible distortions to ***be*** considered is limited (e.g. to affine changes). However, often the distortions are completely unknown and in such cases the paradigm of an interpolated surface deformation or warping **is** often adopted to provide the registration transformation. Such warpings are typically based on polynomial or spline equations [9] and consequently can be defined to ensure certain desirable mathematical properties are met by the resulting registration (e.g. continuity, smoothness). Systematic inversion approaches provide a direct form of the correspondence function, whereas warping-based techniques require a search based on comparison of some details of the intensity values in the two images to be registered, in order to calculate optimal correspondence.

2. IMAGE MATCHING

The comparison and search approach described above **is** the basis of image 'matching', requiring selected image processing operations to be performed on part or all of the two images prior to searching for the optimal correspondence function value. The results of the matching process are thus critical values on which the successful optimizing of the correspondence function depends. Typical matching processes rely on either 'area-based' or 'feature-based' approaches and have been extensively reviewed in the contexts of stereo photogrammetry and computer vision [3],

[2]. Area-based matching undertakes comparisons of intensity values and distributions in neighbourhoods of spatially adjacent but independently considered pixels. Feature-based matching compares information derived from analysing the image intensity values at a higher structural level related to visual understanding of the scene. For example, cross-correlation is a popular area-based method [5] and edge-association is a common feature-based method [8].

In general, area-based matching provides a denser collection of points at which correspondence function values exist than does feature-based matching. However, area-based matching can be sporadically inaccurate and confused because local region context information is not taken into account as part of the match calculations. For this reason, some efforts have been invested in trying to combine area-based and feature-based matching into a hybrid approach, which would overcome many such ambiguities. Techniques have been proposed which first apply feature-based matching and then apply area-based matching in the unmatched or poorly matched vicinity of the feature matches [4].

The above hybrid approach offers promise but lacks the facility to characterise and control the accuracy of the match. Incorporating more visual context information would allow more closely directed matching calculations to be undertaken. We have experimented with some 'region-based' approaches which perform elementary image segmentations to provide contexts which are more visually appropriate. For example, coherent spatial regions defined coarsely using quadtrees were applied to limit the extent of potential match positions in matching for motion estimation [1]. Another technique has applied multiresolution shape alignment based on polygonal region outline representation as a basis for localised control of matching [7].

Recent work has considered using error-estimate-weighted combinations of match values arising independently from different matching techniques to increase overall accuracy of the resulting correspondence function values [6]. This work has indicated that when the raw image data quality is significantly degraded, better registration can be obtained by applying several different matching techniques and allowing the weighting process to select those of greatest estimated accuracy relative to the particular technique generating them, than by choosing only one matching technique to apply to the whole image. This finding raises the question of how much image information is necessary to achieve acceptable matching, or conversely how strongly matching is affected by degrading of image quality The experiments reported in the remainder of this paper offer some preliminary results in response.

3. COMPRESSION EXPERIMENT§

A major source of irreversible degradation in digital image data arises from lossy compression, which results in independent pixel intensity shifts in the reconstructed image. For many commonly used lossy compression techniques (e.g. JPEG, wavelets, fractals), these intensity shifts are highly non-uniform spatially amongst adjacent pixels, and so systematic analysis of the effects of the shifts is difficult to perform. By instead assessing the degree of variation introduced into the match positions computed by different matching processes using a range of compressed images, in this case targeted for 3D surface reconstruction for remote sensing applications, typical effects on the subsequent correspondence function and so on image registration can be estimated.

This procedure has been used in the work described here to answer a number of questions arising from different possibilities for compression and matching of the data. How does the amount of compression affect the matching accuracy? Is one type of matching approach more severely affected by compression than another? How does the loss of accuracy from compressing of DEMs compare with that from compression of the original images? Three experiments are described here, and the results for these experiments are presented for two image sets, both of them stereo pairs which have been intensity stretched to cover a full 8-bit range, and digitally reoriented to achieve approximate epipolarity Images A_1 and A_2 are **SPOT** satellite scenes containing a mixture of rough mountain topography, farmland and urban settlement. Images B_1 and B_2 are aerial survey views of a freeway intersection with surrounding grassland and buildings. The distribution of intensity values is characterized by $\sigma = 7.4$ for A and $\sigma = 6.3$ for B.

A single image from each pair, and a typical smoothed area-matched derived DEM, are shown in Figure 1 Results for the three experiments are shown in separate tables and discussed individually below Images were compressed using the Independent JPEG Group CJPEG coder and the Q values shown in the results are the quality factor settings used by the coder Compression is reported as bits per pixel (bpp) in all cases, since both pixel and DEM values are represented over 8 bit ranges, and the overall matching rate is shown as the percentage of pixels matched (ppm). Errors between reference and reconstructed images or DEMs are computed as the mean absolute difference per pixel (mad), again expressed as a percentage of the full range of intensity or height values available (i.e. 8 bits in both cases).

The first experiment (see Table 1) considered the effects on the accuracy of match values when both images in the pair were compressed with similar quality to each other For each image of both pairs, the compression achieved and the accuracy with which the intensity values were reconstructed on decoding are shown first. In both cases, the difference in compression rates and reconstruction accuracies between the two images of a pair was insignificant. However, owing to the substantially different nature of the two scenes depicted (highly variable local intensity values in A versus fairly consistent local intensity values over much of B), the compression rates achieved for the two image pairs at similar quality factors varied by a factor of 2. At high quality factors, the accuracy of intensities in the reconstructed images was fairly close for both image pairs, but as the quality factor was

decreased a substantial divergence in accuracy trends occurred. The worse accuracy for Image A at high compression rates was caused by the inability of JPEG to model high spatial frequencies well via the DCT, when high order coefficients were eliminated.

Next, DEMs were derived from the reconstructed image pairs, using either an area-based or a feature-based matching process. The area-based method applied weighted cross-correlation on a 5x5 pixel neighbourhood, while the feature-based method used Laplacian-of-Gaussian thresholded edge association. In each case, the parameters of the matching processes were varied until the matches yielded a similar total matched coverage of the particular image pair The overall coverage achieved for image pair A was worse than for image pair B due to the highly textured image content, which was harder to match under the subtle variations in intensity between the images of the A pair than the strongly contrasting regions present in B. Linear interpolation was used to provide DEM values at unmatched locations.

At high quality factors, the DEMs generated for image pair A were superior in accuracy to those obtained for image pair B, but they rapidly converged to similar values as the quality factor is decreased. This initial difference was partly due to the better representation of textured high spatial frequencies by JPEG in image pair A, and partly due to the inadequacy of linear interpolation for modelling the sharp edges of image pair B. Note however that the accuracy obtained compared with the amount of compression achieved was consistently better for image pair B, due to the better overall compression performance noted above. These results suggest that the matching technique adopted need not affect the matching accuracy substantially, while the amount of compression to which the image is subjected has a very profound influence.

The second experiment (see Table 2) considered the effect on the calculation of match values when the images in the pair were stored at significantly different qualities from each other The purpose of this experiment was to establish whether significantly better results could be obtained for matching accuracy by preserving one of the images at higher quality than the other Both possible choices of which image was preserved at higher quality were considered (I_1 was always matched onto I_2). For convenience, the higher quality image was compressed at a fixed quality factor of 95 in all cases. This experiment revealed for image pair A that both match methods performed better when I_2 was stored at the fixed high quality and I_1 was at a low quality factor, than vice versa For image pair B, this was only so for one matching method (feature-based), perhaps due to the effects of interpolation as mentioned before. For image pair A, a consistently better match accuracy was obtained using the feature-based matching method at low quality factors, suggesting that the 'blurring' effect assocjated with the compression affected the area-based matching method more adversely Although the variations in results were relatively minor, there appeared to be a consistent benefit in applying feature-based matching when I_2 at a fixed high quality factor

The third experiment (see Table 3) explored the matter of whether there might be an advantage in compressing the DEM rather than the second image in the pair Just as a DEM may be generated from a pair of compressed images, so also might a second image be generated from a compressed image and DEM. Under these circumstances, an orthoimage might be more useful to store than either of the initial image pair. Depending on the application, once a DEM has been derived it may not be necessary to repeat that calculation; however, in some circumstances it would be necessary (e.g. if more data is added or a different method of DEM construction is to be applied). The distribution of (normalised) height values in the two DEMs was σ = 8.3 for A and σ = 6.0 for B. The compression rates achieved for the two DEMs varied substantially, as the variances were somewhat different. Nevertheless, the accuracy of the reconstructed DEM values was rather similar for both A & B, suggesting that these might depend more strongly on the quality factor used than on the nature of the DEM. Comparing these results with those of Tables 1 and 2, it can be seen that for A, it is more efficient to store the two images of the pair, whereas for B it is more efficient to store one image and the DEM, at low quality factors.

4. CONCLUSION

JPEG image compression was applied to some typical remotely sensed images to provide sample degraded data for matching experiments. Both pixel and feature based matching techniques were considered, along with the use of different quality factors affecting the compression rates and reconstruction accuracies. A preference exists for using feature-based matching at low compression quality factors. Also, where one image in a pair is to stored at higher quality than the other, it is preferable that this be the image being matched onto (i.e. 12). The results presented indicate that at compression rates of less than 1 bpp, matching accuracies in the range 5% to 10% mad per pixel can be achieved on substantially different image data.

The assistance of Stewart Haines in undertaking the experiments reported here is gratefully acknowledged. This work was performed with support from grants under the Australian Research Council Industry Collaborative program and the Research Equipment & Infrastructure Facilities program.

REFERENCES

[1] Baker M. & Maeder A "Segmentation-based coding of motion fields for video compression," ***Proceedings of SPIE,*** 2668,1996.

[2] Barnard S.T & Fischler M.A. "Computational stereo", ***Computer Surveys*** 14(4), 553-572, 1982.

[3] Dhond U.R. & Aggarwal J.K. "Structure from stereo - a review", ***IEEE Transactions on Systems, Man, & Cybernetics,*** 19(6), 1489-1510, 1989.

[4] Greenfeld J.S. ***"An*** operator-based matching system", ***Photogrammetric Engineering & Remote Sensing,*** 57(8), 1049-1055, 1991

[5] Hannah M.J "A system for digital stereo image matching", ***Photogrammetric Engineering & Remote Sensing*** , 55(12), 1765-1770, 1989

[6] Jones M. & Maeder A. "Compound stereo image disparity technique", ***Proceedings of 1995 Conference on Digital Image Computing. Techniques & Applications,*** Australian Pattern Recognition Society, 20-25, Brisbane, December 1995.

[7] Maeder A & Jones M. "Multiresolution shape matching for image fusion", ***Proceedings of 1st IEEE Conference on Image Processing,*** 701-704, Austin TX, November 1994.

[8] Marr D. & Poggio T "A computational theory of human stereo vision", ***Proceedings of the Royal Society of London,*** B204, 301-328, 1979.

[9] Wolberg G. "Digital Image Warping", ***IEEE Computer Society Press,*** 1990.

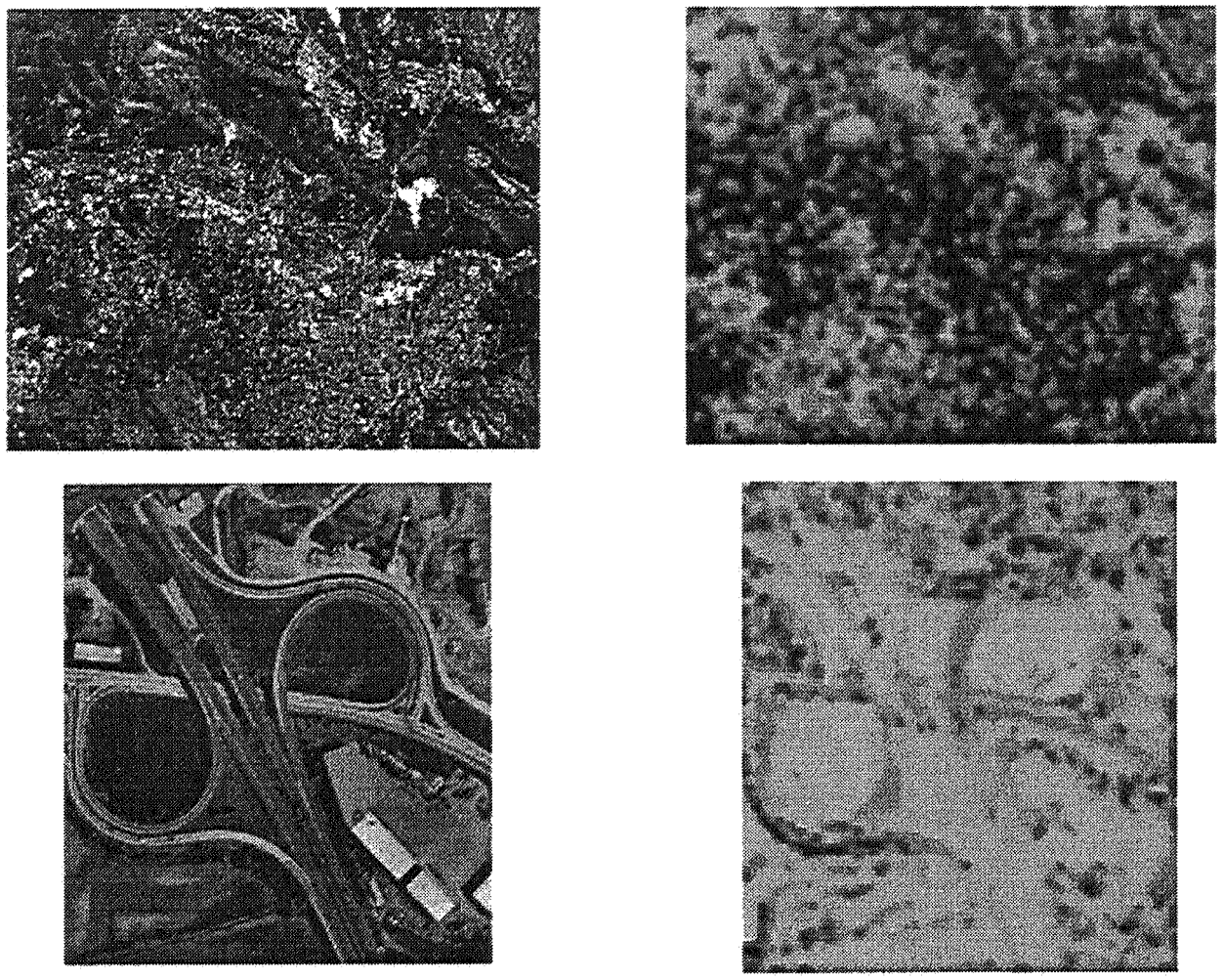


Figure 1: One of the images of the pair A (above) and B (below) and their associated DEMs

	Q	I_1 (bpp)	I_1 (mad) %	I_2 (bpp)	I_2 (mad) %	Area (ppm) %	DEM (mad) %	Feature (ppm) %	DEM (mad) %
Image A	95	6.51	0.6	6.53	0.6	61	1.4	61	1.3
	90	5.05	1.2	5.06	1.2	60	2.9	61	2.3
	80	3.82	2.3	3.84	2.4	58	5.1	60	4.2
	60	2.71	4.2	2.73	4.3	56	7 1	59	6.6
	40	2.05	5.6	2.06	5.7	57	8.5	59	8.6
Image B	95	3.69	0.5	3.58	0.5	83	2.5	84	2.5
	90	2.67	0.7	2.58	0.7	83	3.5	84	3.9
	80	1.87	1.1	1.80	1.0	82	5.4	83	5.7
	60	1.29	1.4	1.25	1.4	81	7.0	83	7.3
	40	1.01	1 7	0.98	1 7	79	8.6	82	8.7

Table 1: Results for images and DEMs of reconstructed images when $Q(I_1) = Q(I_2)$

	Q	Area (ppm) %	DEM (mad) %	Feature (ppm) %	DEM (mad) %
Image Pair A	95	61	1.4	61	1.3
$Q(I_1) = 95$	90	60	2.3	61	1.9
	80	59	3.9	60	3.2
	60	59	6.1	61	4.6
	40	57	7.3	61	6.5
$Q(I_2) = 95$	95	61	1.4	61	1.3
	90	60	2.2	61	1 7
	80	59	3.8	61	2.5
	60	57	5.4	59	4.0
	40	59	6.7	61	4.8
Image Pair B	95	83	2.5	84	2.5
$Q(I_1) = 95$	90	83	3.1	84	3.4
	80	82	4.3	83	4.6
	60	81	5.8	83	5.9
	40	79	6.6	82	6.9
$Q(I_2) = 95$	95	83	2.5	84	2.5
	90	82	3.2	84	3 1
	80	82	4.3	83	4 1
	60	80	5.6	83	5.2
	40	78	6 7	83	5 9

Table 2: Results for DEMs of reconstructed images when $Q(I_1) \neq (I_2)$

	Q	DEM (bpp)	DEM (mad) %
Image Pair A	95	6.51	0.6
	90	5.05	1.2
	80	3.82	2.3
	60	2.71	4.9
	40	2.05	7.2
Image Pair B	95	2.32	0.6
	90	1.64	1.2
	80	1 16	2.4
	60	0.82	4.7
	40	0.65	6.9

Table 3: Results for reconstructed DEMs

Session VIII

Computer Vision

Registration of Uncertain Geometric Features: Estimating the Pose and its Accuracy

Xavier Pennec
INRIA - Projet Epidaure
MIT - Artificial intelligence Lab.
NE43-741, 545 Technology square,
Cambridge MA 02-139, **USA**
Email: xpennec@ai.mit.edu

Abstract

We provide in this article a generic framework for the pose estimation from geometric features. We propose more particularly three algorithms: a gradient descent on the &e-mannian least square distance and on the Mahalanobis distance, and an Extended Kalman Filter For all these methods, we provide a way to compute the uncertainty of the resulting transformation. The analysis and comparison of the algorithms show how to combine them in a powerful and high level statistical registration algorithm. An application in medical image analysis demonstrate the usefulness of the method.

1 Introduction

In computer vision, most registration algorithms rely on feature extracted from the images. In this framework, the problem can generally be separated into two steps: (1) finding the correspondences between features (matches) and (2) computing the geometric transformation that maps one set of features to the other **In** this article, we do not discuss matching methods per se, but the estimation of the geometric transformation and its accuracy.

We first review the existing methods for computing the **3D** rigid motion from two sets of matched points and how one can estimate its uncertainty. However, models of the real world often lead to consider more complex features like lines [Gri90], planes, oriented points [FA96] or frames [PA96] and no traditional method easily extends to such features. Moreover, we have shown **in** [PA97] that geometric features generally do not belong to a vector space but rather to a manifold and that it induces paradoxes if we try to use the standard techniques on points with them. Thus, we have developed in [Pen96] a rigorous theory of uncertainty on geometric features. The basic idea is that our measures on features should be invariant by the action of a given transformation group. In the geometric framework, this can be ensured by using an invariant Riemannian distance on the manifold (which means that the transformation group is a kind of "rigid" one for the features) We have developed the existence conditions for such an invariant metric, a coherent definition of the expectation and of the covariance matrix of random features. Then, other statistical operations can be defined, such **as** the Mahalanobis distance and the χ^2 test

Using these basic operations, we develop in the second section three generic pose estimation algorithms. The first is the generalization of the Least Squares and corresponds to the minimization of the squared Riemannian distance between matched features. The minimum is obtained by gradient descent The two others minimize the Mahalanobis distance (in the rigorous sense of above) either by gradient descent or Kalman filtering. For the three methods, we provide a way to compute an estimation of the result accuracy

In the thud section, we analyze and compare the three algorithms in terms of accuracy of the estimated transformation, accuracy of its uncertainty estimation and of computation time. This leads to a useful classification for the use of these algorithms and a way to combine them in order to build a general but practical pose estimation algorithm including the estimation of the noise on features and an outliers rejection stage for robustness.

In the last section, we apply this theory to a practical case in Medical Image Analysis: the estimation of the relative motion of bones in the pelvis from MRI images. In this study, the estimation of the pose uncertainty turns out to be of the highest importance since it prevents a possibly wrong conclusion.

1 1 Pose estimation from matched points

In order to introduce the pose estimation problem, we **as**sume here that our objects of interest are represented by two sets of n matched points $\{x_i\}$ and $\{y_i\}$ in $\mathbb{R}^3$ We call them respectively the model and scene points. Thus, if the transformation between these two objects is rigid and if we have exact measurements, we have $y_i = R.x_i + t$ We could then determine the movement parameters with only **3** couples of matched points: this gives 9 scalar equations for 6 unknown **(3** for the translation and **3** for the rotation) and **3** invariants (the distance between points in each triplet)

Unfortunately, we never have exact data and we have to cope with measurement errors. Because of the **3** invariants the system is over-constrained. The usual solution is to look for the parameters of the rotation R and the translation t that minimize the sum of square errors:

$$C(R,t) = \frac{1}{2}\sum_i \|y_i - R.x_i - t\|^2 \qquad (1)$$

This problem is called the *orthogonal Procrustes problem* in statistics, the *absolute orientation problem* in photogrammetry and the *pose estimation problem* in computer vision Several closed form solutions have been developed, using unit quaternions [Hor87, Aya91, Fau93], the singular value decomposition (SVD) [Sch66, AHB87, Ume91], the Polar decomposition [HHN88] or dual quaternions [WS91] In our implementation, we use the unit quaternion method

Since all these methods compute the global minimum of the same criterion, the solutions are obviously mathematically identical. A study of the numerical point of view [LEF95] concludes that the difference in accuracy between the four methods is not significant (compared to the numerical precision of the machine) and that the computation times are comparable (a maximum difference of 15% which can be due to variations in the optimization of the implementation). The only (very slight) difference is the sensibility of the methods when the set of points becomes close to a degenerated configuration (aligned points).

The second problem is that the rigid transformation we have computed is of course not the exact transformation between the exact points. It is important and even vital in some medical applications to estimate the accuracy of our estimation. If uncertainty handling is a central topic in several works, like [DW88a, DW88b] or [Aya91, ZF92], there are much fewer studies dealing with the accuracy of the estimated motion Early experimental works can be found in [FH84], [Sny89], or [HJL+89]. A theoretical evaluation of the errors on rotation estimation is introduced in [Kan93, Kan94]. We have described in [PT97] how to estimate the uncertainty on this least squares pose estimation. We just recall here the results. Let r be the rotation vector and $R(\mathrm{r})$ the correspondingrotation matrix the action of the rotation vector on a point x is $\mathrm{r} \star \mathrm{x} = R(\mathrm{r})$ and the action of a rigid transformation $\mathrm{f} = (\mathrm{r}, t)$ is $\mathrm{f} \star \boldsymbol{x} = R(\mathrm{r}).x + t$ In the case of an isotropic, identical and independent noise with standard deviation σ_x on model points and σ_y on scene points (the least squares estimator is optimal only at this condition), the covariance on the estimated rigid transformation $\hat{\mathrm{f}} = (\hat{\mathrm{r}}, \hat{t})$ is the inverse of the criterion Hessian matrix scaled by the variance of the residuals $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$

$$\Sigma_{\hat{\mathrm{f}}\hat{\mathrm{f}}} = \sigma_z^2 . H_{\hat{\mathrm{f}}}^{(-1)} \qquad (2)$$

with

$$H_{\mathrm{f}} = \sum_i \left(\frac{\partial (\mathrm{f} \star x_i)}{\partial \mathrm{f}} \right)^{\mathrm{T}} \left(\frac{\partial (\mathrm{f} \star x_i)}{\partial \mathrm{f}} \right)$$

Moreover, σ_z^2 can be directly estimated from the residuals with $\hat{\sigma}_z^2 = 2.C(\hat{\mathrm{f}})/(3.(n-2))$

Now, if we have an uncertainty information on our points (i.e. covariance matrices $\Sigma_{\boldsymbol{x_i x_i}}$ and $\Sigma_{\boldsymbol{y_i y_i}}$), the least squares estimator is no longer optimal and we have to minimize the Mahalanobis distance. We have proposed in [PT97] a approximated solution based on the Extended Kalman Filter We do not detail this method here as it will appear as a special case of section 2.3.

1.2 Geometric features

Models of the real world often lead to consider geometric features that are more complex than points, such as lines [Gri90], planes, oriented points [FA96] or frames [PA96]. No traditional registration method easily extends to such features. Moreover, we have shown in [PA97] that geometric features generally do not belong to a vector space but rather to a manifold and that it can induce paradoxes if we used-rectly the standard techniques, Thus, we have developed in [Pen96] a rigorous theory of uncertainty on geometric features- The basic idea is that our measures on features should be invariant by the action of a given transformation group We summarize here the basic concepts of this theory.

In the geometric framework, one specifies the structure of a manifold M by a *Riemannian metnc*. This is a continuous collection of dot products on the tangent space at each point x of the manifold. Thus, if we consider a curve on the manifold, we can compute at each point its instantaneous speed vector and its norm, the instantaneous speed. To compute the length of the curve, we can proceed as usual by integrating this value along the curve. The distance between two points of a connected Riemannian manifold is the minimum length among the curves joining these points. The curves realizing this minimum for any two points of the manifold are called geodesics. Let o be a point of the manifold that we call the origin. We can develop the manifold in the tangent space at o along the geodesics (think of rolling a sphere along its tangent plane at the north pole). The geodesics going through the origin are transformed into straight lines and the distance along these geodesics are conserved. This realizes a chart called *the principal chart* that covers all the manifold (except a set of null measure called the cut locus) Let $\vec{y}$ be the representation of y in this chart. Then its distance to the origin is $\mathrm{dist}(o, \mathrm{y}) = \|\vec{y}\|$. This chart is somehow the "most linear" chart of the manifold with respect to the origin.

Now, since we are working with features on which acts a transformation group $\mathcal{G}$ that models the possible image view points, it is natural to choose an invariant Riemannian metric (the existence condition is detailed in [Pen96]) This way, all the measurements based on distance axe independent of the image reference frame or the transformation of the image: $\mathrm{dist}(\mathrm{x}, \mathrm{y}) = \mathrm{dist}(g \star \mathrm{x}, g \star \mathrm{y})$ Let f_{x} be a "placement function", i.e. a transformation of G such that $\mathrm{f}_{\mathrm{x}} \star \mathrm{o} = \mathrm{x}$. The distance can be expressed as: $\mathrm{dist}(\mathrm{x}, \mathrm{y}) = \mathrm{dist}\left(\mathrm{f}_{\mathrm{x}}^{(-1)} \star \mathrm{y}, o\right) = \|\mathrm{f}_{\vec{x}}^{(-1)} \star \vec{y}\|$. One can show that this formula is independent of the choice of the placement function.

In fact, we have shown that all interesting operations on deterministic and uncertain geometric features and transformations can be expressed in the principal chart of the manifold and the group with only a few basic algorithmic blocs that we call atomic Operations.

For the group, we have to implement

- *the cornpositron* $\vec{\mathrm{f}} \circ \vec{\mathrm{g}}$ and its Jacobians $\frac{\partial (\vec{\mathrm{f}} \circ \vec{\mathrm{g}})}{\partial (\vec{\mathrm{f}} \circ \vec{\mathrm{g}})}$, $\frac{\partial (\vec{\mathrm{f}} \circ \vec{\mathrm{g}})}{\partial \vec{\mathrm{g}}}$,
- *the inversion* $\vec{\mathrm{f}}^{(-1)}$ and its Jacobian $\frac{\partial (\vec{\mathrm{f}}^{(-1)})}{\partial \vec{\mathrm{f}}}$

For the feature manifold, the basic operations are

- *the action* $\vec{\mathrm{f}} \star \vec{\mathrm{x}}$ and its Jacobians $\frac{\partial (\vec{\mathrm{f}} \star \vec{\mathrm{x}})}{\partial \vec{\mathrm{f}}}$ and $\frac{\partial (\vec{\mathrm{f}} \star \vec{\mathrm{f}})}{\partial \vec{\mathrm{x}}}$,
- and *the placement function* $\vec{\mathrm{f}}_{\vec{\mathrm{x}}}$ with its Jacobian $\frac{\partial (\vec{\mathrm{f}}_{\vec{\mathrm{x}}})}{\partial \vec{\mathrm{x}}}$

To simplify further computations, we can add to these atomic operations the Jacobian of the

- *left translation of the identity* $J_L(\vec{\mathrm{f}}) = \left. \frac{\partial (\vec{\mathrm{f}} \circ \vec{\mathrm{e}})}{\partial \vec{\mathrm{e}}} \right|_{\vec{\mathrm{e}} = Id}$

From this small set of atomic operations, we can implement the composition and inversion of a probabilistic trans-

formation, the action of such a transformation on a probabilistic (or deterministic) feature, the distance and Mahalanobis distance between transformations, etc. We have implemented this framework for rigid transformations acting on different kinds of features such as **frames**, semi- or non-onented **frames** and points. Frames are composed of a point and an orthonormal trihedron. In semi-oriented frames, the trihedron is composed of two directions $(t_1, t_2) \equiv (-t_1, -t_2)$ and an oriented vector $\boldsymbol{n}$ and models the differential properties of a point on a surface. For non oriented frames, the normal is also un-oriented $(\boldsymbol{n} \equiv -\boldsymbol{n})$

2 Pose estimation from matched features

Since we now have a distance between two geometric features and a Mahalanobis distance between two probabilistic features, the generalization of the two criteria for pose estimation from points is straightforward. However, it is rather unlikely to find a closed form solution like in the point case. Thus, we focus on iterative but generic algorithms.

2.1 Least squares on Riemannian distance

Let $\{x_i\}$ and $\{y_i\}$ be two sets of matched features. The Least squares criterion is easily written using the Riemannian distance:

$$\begin{aligned} C(f) &= \frac{1}{2} \sum_i \mathrm{dist}(y_i, f \star x_i)^2 \\ &= \frac{1}{2} \sum_i \mathrm{dist}\left(f_{y_i}^{(-1)} \circ (f \star x_i)\ ,\ o\right)^2 \end{aligned}$$

2.1.1 Error vector

Now, thanks to the good properties of the principal chart, it turns out that this criterion can be expressed as a classical sum of square of vector norms. Let $\vec{z}_i = \vec{f}_{y_i}^{(-1)} \circ (\vec{f} \star \vec{x}_i)$ be the "error vector" in the principal chart The criterion becomes

$$C(f) = \frac{1}{2} \sum_i \|\vec{z}_i\|^2$$

This expression is only true in the principal chart and is false for any other representation

To minimize the criterion and determine the uncertainty at the minimum, we will need to compute its first and second derivatives. From the atomic operations, and using the composition rule for differentials, we can compute the error vector $\vec{z}_i$ and its Jacobians $\frac{\partial \vec{z}_i}{\partial \vec{x}_i}$, $\frac{\partial \vec{z}_i}{\partial \vec{y}_i}$, and $\frac{\partial \vec{z}_i}{\partial \vec{f}}$ (see [Pen96] for equations) If the uncertainty of the measurements $\vec{x}_i$ and $\vec{y}_i$ is represented by the covariance matrices $\Sigma_{\vec{x}_i \vec{x}_i}$ and $\Sigma_{\vec{y}_i \vec{y}_i}$, the uncertainty of the error vector is:

$$\Sigma_{\vec{z}_i \vec{z}_i} = \frac{\partial \vec{z}_i}{\partial \vec{x}_i} \Sigma_{\vec{x}_i \vec{x}_i} \frac{\partial \vec{z}_i}{\partial \vec{x}_i}^{\mathrm{T}} + \frac{\partial \vec{z}_i}{\partial \vec{y}_i} \Sigma_{\vec{y}_i \vec{y}_i} \frac{\partial \vec{z}_i}{\partial \vec{y}_i}^{\mathrm{T}}$$

2.1.2 Derivatives of the criterion

Since the criterion is $C = \frac{1}{2} \sum_i \vec{z}_i^{\mathrm{T}} . \vec{z}_i$, the first derivative is:

$$\Phi = \left(\frac{\partial C}{\partial \vec{f}}\right)^{\mathrm{T}} = \sum_i \frac{\partial \vec{z}_i}{\partial \vec{f}}^{\mathrm{T}} . \vec{z}_i$$

Neglecting the term in $\ddot{z}.z$ with respect to the terms in $\dot{z}^2$ in the second derivatives (see [GMW81, section 4.7]), we obtain:

$$H = \frac{\partial \Phi}{\partial \vec{f}} \simeq \sum_i \frac{\partial \vec{z}_i}{\partial \vec{f}}^{\mathrm{T}} \frac{\partial \vec{z}_i}{\partial \vec{f}} \qquad \text{and} \qquad \frac{\partial \Phi}{\partial \vec{z}_i} \simeq \frac{\partial \vec{z}_i}{\partial \vec{f}}^{\mathrm{T}}$$

2.1.3 A gradient descent algorithm

Let $\vec{f}$ be an estimation of the sought transformation. We have in the principal chart the following Taylor expansion (see [Pen96])

$$\begin{aligned} C(\vec{f} \circ \vec{\delta f}) = \ & C(\vec{f}) + \Phi^{\mathrm{T}} . J_L(\vec{f}) . \vec{\delta f} \\ & + \frac{1}{2} \vec{\delta f}^{\mathrm{T}} . J_L(\vec{f})^{\mathrm{T}} . H . J_L(\vec{f}) . \vec{\delta f} \\ & + O\left(\|\vec{\delta f}\|^3\right) \end{aligned} \tag{3}$$

where $\Phi = \frac{\partial C}{\partial \vec{f}}^{\mathrm{T}}$ and $H = \frac{\partial^2 C}{\partial \vec{f}^2}$ The minimum of this approximated criterion is given for:

$$\vec{\delta f} = -J_L(\vec{f})^{(-1)} . H^{(-1)} \Phi$$

Now, we can obtain a much better estimation of the transformation by iterating the previous algorithm. As an initial estimate, we can choose the identity if nothing else is given. Let $\vec{f}_t$ be the estimation of the transformation at step t The estimation at step $t+1$ is

$$\vec{f}_{t+1} = \vec{f}_t \circ \vec{\delta f}_t \qquad \text{with} \qquad \vec{\delta f}_t = -J_L(\vec{f}_t)^{(-1)} . H_t^{(-1)} . \Phi_t \tag{4}$$

The process is stopped when the norm $\|\vec{\delta f}_t\|$ of the adjustment transformation becomes too small (we use $\varepsilon = 10^{-10}$, which is almost the numeric precision of the machine) or when the number of iterations becomes too high Practically, we observe that this algorithm always converges in about 10 iterations for more than 3 matches.

2.1.4 Estimation of the uncertainty at the minimum

We have obtained above the transformation minimizing the criterion for the observed data Now, since the data are corrupted by noise (represented by their covariance matrix), the estimated transformation could be slightly different from the exact one. We are now trying to characterize the transformation uncertainty as a function of the data uncertainty

We assume for the moment that both the data and the state are vectors. Let $\hat{\chi}^{\mathrm{T}} = (\hat{z}_1, \quad \hat{z}_n)$ be the vector of observed data and $\hat{f}$ the corresponding state (we have added here the hats to stress the difference between actual and potential measurements). We assume that all our measurements are independent so that $\mathrm{E}, \quad = \mathrm{DIAG}(\Sigma_{z_1 z_1}, \quad \Sigma_{z_n z_n})$. A Taylor expansion of the criterion around the actual values $(\hat{\chi}, \hat{f})$ gives:

$$\begin{aligned} C(\chi, \mathbf{f}) = & \hat{C} + \hat{\Phi}^{\mathrm{T}} . (f - \hat{f}) + \frac{\partial C}{\partial \hat{X}} . (\chi - \hat{\chi}) \\ & + \tfrac{1}{2} (f - \hat{f})^{\mathrm{T}} \hat{H} (f - \hat{f}) \\ & + O(\|f - \hat{f}\|^3) + O(\|\chi - \hat{\chi}\|^2) \end{aligned}$$

In this formula, $\hat{\Phi}$ and $\hat{H}$ are the Jacobian and the Hessian matrix of the criterion (with respect to f) estimated at $(\hat{\chi}, \hat{f})$ The minimum of the criterion is characterized by a null derivative with respect to the state. Using the above Taylor expansion, this derivative is:

$$\begin{aligned} \Phi(\chi, \mathbf{f}) = & \hat{\Phi} + \hat{J}_\Phi . (\chi - \hat{\chi}) + \hat{H}\ (f - \hat{f}) \\ & + O(\|f - \hat{f}\|^2) + O(\|\chi - \hat{\chi}\|^2) \end{aligned}$$

where $\hat{J}_\Phi = \frac{\partial^2 C}{\partial \chi \partial \hat{f}}(\hat{\chi}, \hat{f})$. Equating to zero, we obtain the new state:

$$f = \hat{f} - \hat{H}^{(-1)}.\hat{J}_\Phi.(\chi - \hat{\chi}) + O(\|\chi - \hat{\chi}\|^2)$$

Thus, the covariance of $\hat{f}$ is:

$$\Sigma_{\hat{f}\hat{f}} = \mathbf{E}\left((f - \hat{f}).(f - \hat{f})^T\right) = \hat{H}^{(-1)}.\hat{J}_\Phi.\Sigma_{\hat{\chi}\hat{\chi}}\hat{J}_\Phi^T\,\hat{H}^{(-1)}$$

Now, using the particular form of χ and $\Sigma_{\chi\chi}$, we can simplify $\hat{J}_\Phi.\Sigma_{\hat{\chi}\hat{\chi}}\hat{J}_\Phi^T$ to obtain

$$\Sigma_{\hat{f}\hat{f}} = \hat{H}^{(-1)}\left(\sum_i \frac{\partial \hat{\Phi}}{\partial \vec{z}_i}\,\Sigma_{z_i z_i}\,\frac{\partial \hat{\Phi}}{\partial \vec{z}_i}^T\right)\hat{H}^{(-1)} \qquad (5)$$

Last but not least, the data and the state are not vectors in our case, but points respectively in a manifold and in a transformation group. However, thanks to be properties of the principal chart, this derivation is still valid if we replace $(f - \hat{f})$ by $J_L(\hat{\vec{f}}).(\hat{\vec{f}}^{(-1)} \circ \vec{f})$ and $(\chi - \hat{\chi})$ by a somehow similar expression Now, it turns out that the definition of the covariance is changed accordingly and that finally nothing is changed in equation (5).

2 1.5 Uncertainty for the least-squares solution

As for points, the least-square estimation is optimal only if $\Sigma_{z_i z_i} = \sigma_z^2\, Id$ Then, taking into account that $\frac{\partial \Phi}{\partial \vec{z}_i} \simeq \frac{\partial \vec{z}_i}{\partial \vec{f}}^T$, the uncertainty is simplified into

$$\Sigma_{\hat{f}\hat{f}} = \sigma_z^2.H^{(-1)}$$

Moreover, the variance σ_z^2 can be estimated from the data with

$$\hat{\sigma}_z^2 = \frac{\sum_i \|\hat{\vec{z}}_i\|^2}{3(n-3)} \text{ for points} \qquad \hat{\sigma}_z^2 = \frac{\sum_i \|\hat{\vec{z}}_i\|^2}{6(n-1)} \text{ for frames}$$

However, this hypothesis is quite strong for frame-based features: it means that the variance σ_θ^2 on the rotation vector representing the trihedron is equal to the variance σ_x^2 on the position of the frame Thus the choice of the units (radian, degree, meter or millimeter) is crucial. Most of the time, this hypothesis does not hold and we have to use equation (5).

2.2 Gradient descent on the Mahalanobis distance

To allow different and non isotropic covariance matrices for different measure, we can minimize the sum of squared Mahalanobis distances after registration It turns out that this Mahalanobis distance can be expressed with exactly the same error vector as before:

$$\begin{aligned} C(\mathrm{f}) &= \frac{1}{2}\sum_i \mu^2(\mathrm{y}_i, \mathrm{f} * x_i) \\ &= \frac{1}{2}\sum_i \mu^2(\mathrm{z}_i, 0) = \frac{1}{2}\sum_i \vec{z}_i^T\,\Sigma_{z_i z_i}^{(-1)}.\vec{z}_i \end{aligned}$$

Thus, the algorithm is the same than for the least squares, but the derivatives of the criterion are different. The first derivative is

$$\Phi = \left(\frac{\partial C}{\partial \vec{\mathrm{f}}}\right)^T = \sum_i \frac{\partial \vec{z}_i}{\partial \vec{\mathrm{f}}}^T\,\Sigma_{z_i z_i}^{(-1)}.\vec{z}_i$$

Similarly, the second derivatives can be approximated by

$$H = \frac{\partial \Phi}{\partial \vec{\mathrm{f}}} \simeq \sum_i \frac{\partial \vec{z}_i}{\partial \vec{\mathrm{f}}}^T.\Sigma_{z_i z_i}^{(-1)}\,\frac{\partial \vec{z}_i}{\partial \vec{\mathrm{f}}}$$

and

$$\frac{\partial \Phi}{\partial \vec{z}_i} \simeq \frac{\partial \vec{z}_i}{\partial \vec{\mathrm{f}}}^T\,\Sigma_{z_i z_i}^{(-1)}$$

Now, the Taylor expansion for the criterion is the same, and the evolution for the gradient descent is still given by equation (4). We can use the same starting value and stopping criterion as before. Practically, we have observe a convergence in about 15 iterations when starting from identity and in 5 to 10 iterations when starting from the least-squares solution,

For the uncertainty of the solution, we replace the values of H and $\frac{\partial \Phi}{\partial \vec{z}_i}$ into equation (5) and obtain:

$$\Sigma_{\hat{f}\hat{f}} = \hat{H}^{(-1)}$$

2.3 Kalman filtering

Assume that we have an initial estimate $\vec{\mathrm{f}}_0$ with covariance $\Sigma_{\vec{f}_0\vec{f}_0}$ The natural criterion to minimize becomes

$$\begin{aligned} C(\mathrm{f}) &= \frac{1}{2}\mu^2(\mathrm{f}, \mathrm{f}_0) + \frac{1}{2}\sum_i \mu^2(\mathrm{y}_i, \mathrm{f} * \mathrm{x}_i) \\ &= \frac{1}{2}(\vec{\mathrm{f}}_0^{(-1)} \circ \vec{\mathrm{f}})^T.J_L(\vec{\mathrm{f}}_0)^T\,\Sigma_{\vec{f}_0\vec{f}_0}^{(-1)}.J_L(\vec{\mathrm{f}}_0).(\vec{\mathrm{f}}_0^{(-1)} \circ \vec{\mathrm{f}}) \\ &\quad + \frac{1}{2}\sum_i \vec{z}_i^T.\Sigma_{z_i z_i}^{(-1)}.\vec{z}_i \end{aligned}$$

In the case of linear transformations of Gaussian random vectors, the maximum likelihood estimation and minimum error variance estimation corresponds, and could be expressed as a recursive filter known as the Kalman filter [Jaz70]. In the general case of non-linear transformations (which is indeed our case here), the filter could be extended by using linear approximations around the current state and measurements [Aya91]. At each step the input is

- an estimate $(\vec{\mathrm{f}}_{i-1}, \Sigma_{\vec{f}_{i-1}\vec{f}_{i-1}})$ of the state,
- the measurement $(\vec{z}_i, \Sigma_{\vec{z}_i\vec{z}_i})$
- and the linearized dependency between the state and the measurement $M_i = \frac{\partial \vec{z}_i}{\partial \vec{\mathrm{f}}}\Big|_{\vec{\mathrm{f}}=\vec{\mathrm{f}}_{i-1}}$

The output is the updated estimation of the state using the following equations:

$$\begin{aligned} K_i &= \Sigma_{\vec{f}_{i-1}\vec{f}_{i-1}}.M_i^T.(\Sigma_{\vec{z}_i\vec{z}_i} + M_i.\Sigma_{\vec{f}_{i-1}\vec{f}_{i-1}}.M_i^T)^{-1} \\ \vec{\mathrm{f}}_i &= \vec{\mathrm{f}}_{i-1} - K_i.\vec{z}_i \\ \Sigma_{\vec{f}_i\vec{f}_i} &= (Id - K_i.M_i).\Sigma_{\vec{f}_{i-1}\vec{f}_{i-1}} \end{aligned}$$

However, we must be careful that the state $\vec{\mathrm{f}}_i$ is constrained to belong to the principal chart and the above Kalman evolution equation could make it go out of its definition domain. Thus, we have to verify after each step that the state is still in the definition domain and correct if it is out (this is done using an additional atomic operation called "Domain" in [Pen96]).

3 Algorithm comparison

In the following, we denote by LSQ, MAHA and KAL the three pose estimation methods developed above. Since we can forget the trihedron in all types of frames to keep only a point, we can also apply the standard pose estimation techniques on points, such as the unit quaternion method, which we denote by QUAT We analyze here the relative accuracy of the different methods, their computation time and the accuracy of the uncertainty on the transformation (which should predict the absolute accuracy).

Statistics are realized by choosing a random sample of N features in an image 512x256x128. This is the exact model. We choose a random rigid movement (the "exact" transformation), and transform the exact model into the exact scene. Then each exact feature is perturbed by a Gaussian noise with a given covariance Σ in order to constitute the observed model and scene.

3.1 Accuracy of the methods

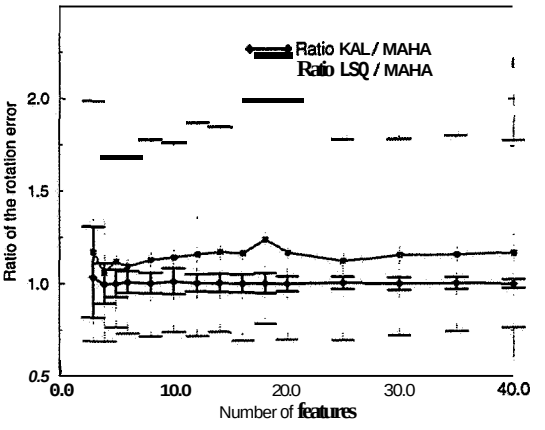
With the same couple of (synthetic) model and scene, we compute the transformation using QUAT, LSQ, KAL and MAHA For each of these estimations, we compute the residual transformation with respect to the exact transformation To simplify the analysis, we only compare the residual angle of rotation θ and the length d of the residual translation As these values are varying a lot with the considered set of matches and since we only want to study the relative accuracy of the three methods, we have chosen the error of MAHA as the reference. Thus, we compute the ratios $\varepsilon_{\theta_{\mathrm{LSQ}}} = \theta_{\mathrm{LSQ}}/\theta_{\mathrm{MAHA}}$ and $\varepsilon_{d_{\mathrm{LSQ}}} = d_{\mathrm{LSQ}}/d_{\mathrm{MAHA}}$ to characterize the relative accuracy of LSQ and similarly for KAL. $\varepsilon_{\mathrm{LSQ}} < 1$ means that LSQ is more accurate than MAHA and $\varepsilon_{\mathrm{LSQ}} > 1$ means less accurate.

To obtain reliable statistics, we have done 250 trials for each number of features and we present in fig. (1) the mean value and the standard deviation' of the relative accuracies of LSQ and KAL with respect to MAHA

In these experiments, LSQ and MAHA are initialized with the identity, whereas KAL is initialized with the result of LSQ (the covariance matrix is multiplied by 50 to minimize its influence on the final uncertainty estimation) Indeed, initializing KAL with the identity and a fixed covariance matrix (or any transformation a little too far from the solution) leads to a pseudo-convergence toward a false solution On the contrary, we did not see any influence of the initialization on the solutions of LSQ and MAHA (except for the computation time)

The next observation is that the absolute accuracy of each method is proportional to $1/\sqrt{n}$ (where n is the number of features) but the ratio of the accuracy of different methods does not depend on the number of matches. We found that KAL and MAHA perform very similarly So do QUAT and LSQ On the contrary, we found that KAL and MAHA are 1.2 times to 2 times more accurate than QUAT and LSQ depending on the covariance matrix on frames (more particularly the ratio of the onentation variance over the position variance)

'As the relative accuracy is multiplicative (2 times more accurate means $\varepsilon = 0.5$), we have computed the mean value and the standard deviation of $\log(\varepsilon)$, but we present the result with ratios for an easier visualization.



Relative error of the LSQ and KAL translations

(reference: MAHA error)

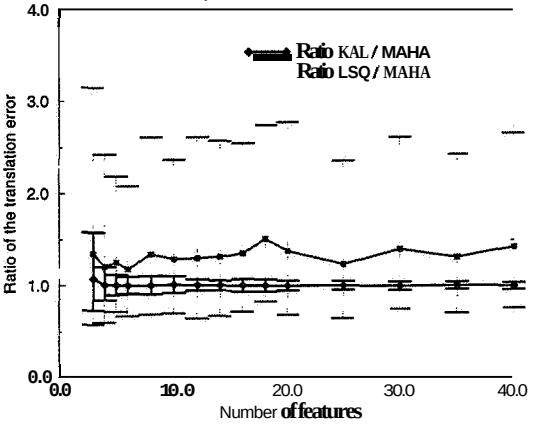


Figure 1 Relative accuracy of the rotation and translation estimated by LSQ and KAL with respect to MAHA

3.2 Accuracy of the uncertainty estimation

3.2.1 The validation index

With synthetic data, we know the exact motion that we are looking for, and we can verify that it is compatible with our transformation estimation modulo its uncertainty. The basic idea of the validation method presented in [PT97] is normalize the residua3 transformation error with respect to the predicted uncertainty In fact, the simplest way do do that is to compute the Mahalanobis distance between the computed and the exact transformation.

$$\mu^2(\mathbf{f},\hat{\mathbf{f}}) = (\hat{\vec{\mathbf{f}}}^{(-1)} \circ \vec{\mathbf{f}})^T \, J_L(\hat{\vec{\mathbf{f}}})^T \, \Sigma^{(-1)}_{\hat{\vec{\mathbf{f}}}\hat{\vec{\mathbf{f}}}} \, J_L(\hat{\vec{\mathbf{f}}}) \, (\hat{\vec{\mathbf{f}}}^{(-1)} \circ \vec{\mathbf{f}})$$

Under the Gaussian hypothesis (justified in [Pen96]), this value should be χ^2_6 distributed if the covariance matrix is exact

We can now repeat this experiment on N pairs of images to obtain N independent values μ_i^2 and verify if it is really χ^2_6 distributed The Kolmogorov-Smirnov (or simply K-S) test [PFTV91] is well adapted to do that but since it only gives a binary answer, we also use the fact that the mean

value of a χ^2_6 distribution is **6** and its variance is **12** We call *validation* ***index*** the estimated mean value of μ_i^2

$$I = \bar{\mu}^2 = \frac{1}{N}\sum \mu_i^2 \qquad \text{[ideal value: 6]}$$

and its variance is computed accordingly with

$$\sigma_I^2 = \frac{1}{N-1}\sum (\mu_i^2 - \bar{\mu}^2)^2 \qquad \text{[ideal value: 12]}$$

This index can be interpreted as an indications on how the estimation method under-estimates **(I > 6)** or over-estimates **(I < 6)** the error on the estimated motion. It is a kind of relative error on the error estimation.

This method can be generalized to the case of real data by splitting the set of matches in two approximatively equal sets on matches. Then one two independent estimates $\mathbf{f}_1$ and $\mathbf{f}_2$ of the same transformation f and their Mahalanobis distance $\mu^2(\mathbf{f}_1, \mathbf{f}_2)$ should once again be χ^2_6 distributed under the Gaussian assumption. The two transformations could then be fused to obtain a better estimation of the transformation and used for other purposes (figure **2)**

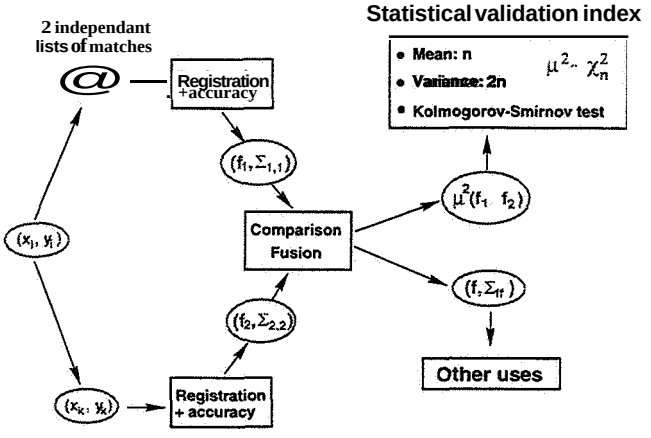


Figure **2:** Validation of the registration uncertainty with real data.

3.2.2 Validation results

As expected, the best accuracy estimation is given by MAHA We present in figure **3** the computed mean and standard deviation of the validation index with MAHA on semi-oriented **frames** with respect to the number of matches The dotted lines represent the theoretical values for a χ^2_6 distribution (a mean of **6** and a standard deviation of $\sqrt{12} = 3.46$**)** The K-S test at **1%** is always accepted The mean validation index over the $25 * 500 = 12500$ registrations in this graphic is $\mathbf{I = 6.05}$ for a variance of $\sigma_I^2 = 11.90$**,** which proves that the uncertainty estimation is very accurate. This indicates **an** underestimation of the covariance matrix of less than **1%.** The same results are observed with other kind of feature.

The graphic for KAL is quite similar, except that the K-S test is one order of magnitude worse. Indeed, the mean validation index is now $\mathbf{I = 6.29}$ for a variance of $\sigma_I^2 = 12.90$**,** indicating an underestimation of **4.6%.**

In the case of LSQ, the hypothesis of a diagonal and isotropic covariance matrix on features does not hold (except for points). Thus, using formula $\Sigma_{\hat{f}\hat{f}} = \sigma_z^2 H^{(-1)}$ leads to a

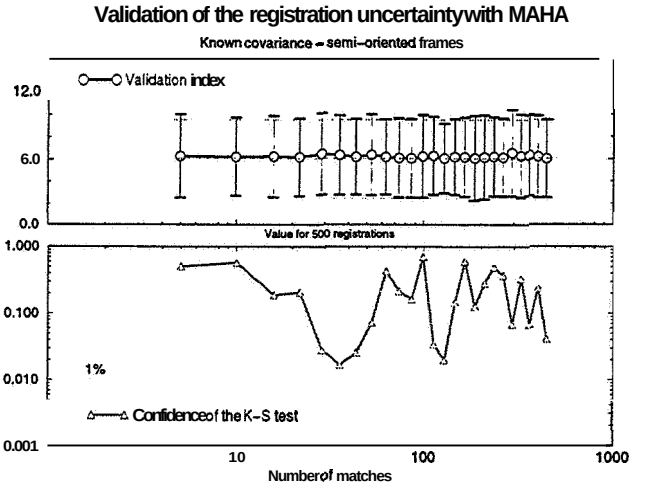


Figure **3:** Validation of the registration uncertainty with a known covariance on semi-oriented **frames** (method MAHA) Top: mean and standard deviation of the validation index with respect to the number of matches used Bottom. **con**fidence value of the K-S test. It validates the uncertainty in all cases with a confidence superior to **1%**

very estimation of the uncertainty. However, We can still use formula **(5)** and in this case the uncertainty is much better estimated. We obtain a mean validation index of $\mathbf{I = 5.82}$ for a variance of $\sigma_I^2 = 11.09$. This means a **3%** overestimation of the uncertainty on the transformation.

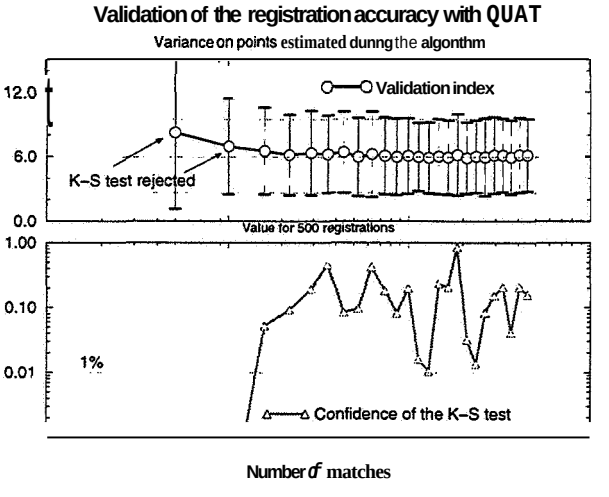


Figure **5:** Validation of the registration uncertainty with QUAT The variance on points is estimated during the algorithm. Top: mean and standard deviation of the validation index with respect to the number of matches used. Bottom: confidence value of the K-S test. It validates the uncertainty with a confidence superior to **1%** for more than **15** matches.

The case of QUAT is more interesting: in most cases, the hypothesis of a diagonal and isotropic covariance matrix **on** points is justified (there is hardly ever an order a magnitude of difference between the **3** directions of the space) Hence,

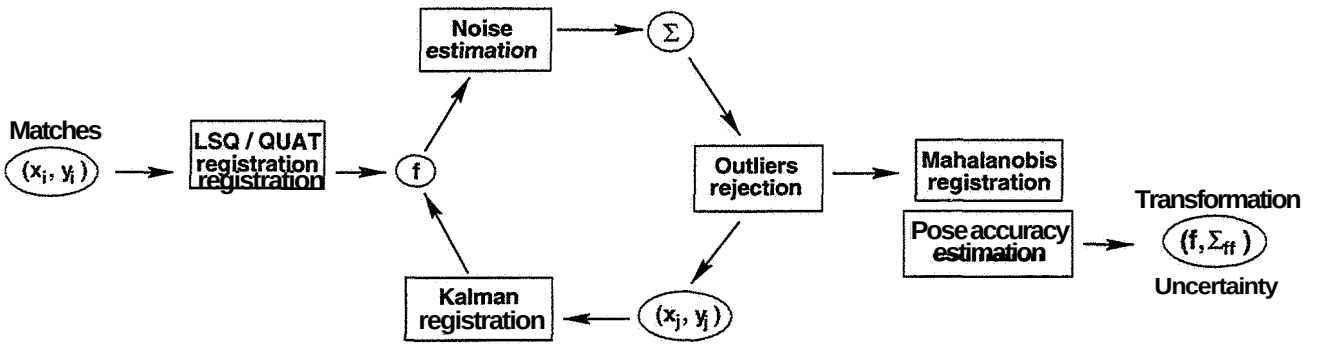


Figure 4: Modular registration algorithm

we *can* use the simplified formula and estimate the variance σ_z^2 on points directly from the data. This make an important difference with the previous experiments where the noise model was **known.** Therefore, we do not expect the uncertainty to be very accurate with a small number of matches. Indeed, we observe in figure (5) that the uncertainty is completely underestimated for less than 15 matches, but is quite correctly estimated (and validated by the K-S test) for more than 15 matches. The mean validation index in this part is $\mathbf{I = 6.06}$ for a variance of $\sigma_I^2 = 12.51$, which corresponds to an underestimation of 1%.

3.3 Computation times

The algorithm LSQ is initialized with the identity but LSQ and MAHA are initialized with the result of QUAT The computation times are approximately linear in the number of matches. Thus, we only give below the computation times for 100 matches on a DEC alpha 3000 166 Mhz (average value on 100 to 500 trials)

QUAT				20 ms
LSQ	694 ms	785 ms	820 ms	275 ms
KAL	216 ms	245 ms	250 ms	87 ms
MAHA	792 ms	880 ms	870 ms	368 ms

3.4 Discussion, use of the algorithms

Due to the choice of the metric (the units for the rotation and the translation), LSQ and QUAT give very similar results. QUAT is much faster and should thus be used when possible, i.e. for more than 3 matches. However, LSQ is still usable for 1 or 2 frame matches (but could be unstable) For the transformation uncertainty, QUAT does not need the covariance matrix on features, but it is only reliable with more than 15 matches. LSQ needs to know the covariance matrix on features is order to produce a reliable uncertainty estimation

KAL and MAHA are quite similar and give a better transformation accuracy. MAHA gives an accurate transformation and uncertainty in all cases, even with very few matches (it is more stable than LSQ). KAL is faster but requires a reliable initialization point and underestimates the uncertainty However, it is incremental and we can forget the matches when they have updated the transformation, which could prove to be a real advantage for the memory complexity. Another combination is to compute the transformation with MAHA for the first (let say 10) matches and then update iteratively the transformation with the other using KAL.

	QUAT	LSQ	KAL	MAHA
Cov. Mat.	No	No	Yes	Yes
Comp. time	++	-	+	- -
Transfo. accuracy	-	-	+	+
Uncert. accuracy		-	- -	+
Sensitivity		Metric	Init. cond.	

3.5 A practical and generic algorithm

At the end of a matching algorithm, we usually have a list of matched features with possibly outliers and no or little information about the noise on features. We have proposed in [PT97] an algorithm that estimates the pose, its uncertainty, the outliers and the noise on features in the case of frames. This algorithm could be easily generalized to generic features and modified to make the best use of each characteristics of the previous algorithms (figure 4).

Since we have no information about the uncertainty of features, we begin by a global QUAT or LSQ registration. Then we can estimate the noise on features, and use a χ^2 test to rule out outliers. The transformation is then up dated using KAL and the process it iterated. When the transformation becomes stable or after a maximum number of iterations (5 to 10), we finally use a MAHA estimation to ensure the accuracy of the uncertainty estimation

4 Estimation of the relative motion of the bones in the pelvis

The goal of the experiment was to measure quantitatively the relative motion of bones in the pelvis during particular motions. Two MRI (T1) images were taken at the Lenval Foundation (Nice, France), one in "open pelvis" configuration and one in "close pelvis" configuration. Images are 256 x 256 x 28 with a voxel size of 1.33 x 1.33 x 5.6 mm (i.e. an anisotropic factor greater than 4 in the z direction!)

The pelvis is principally constituted of the sacrum and the left and right hipbones, forming a quasi-rigid structure. Hence their relative motion is very small and we need to estimate the motions of each bone and their uncertainty to be able to decide statistically if there is really a motion between them.

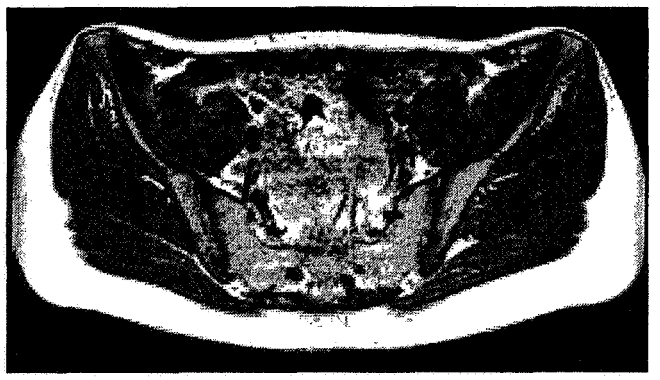


Figure 6: An axial slice of an MR image of the pelvis (T1 weighted). One can see the sacrum in the bottom center and the hipbones on the sides.

4.1 Segmentation

It is quite hard to segment bones in an MRI image. Indeed, the iso-surface method that we usually use for bones in CT-scans or the surface of the brain in MRI images does not work here: the bones appear as an intermediate signal delimited by a narrow band of weaker signal on the border (figure 6) This weak signal has a varying intensity and is moreover asymmetric due to the chemical shift.

Since the other automatic segmentation tools we tried also failed, each bone was manually segmented by a specialist (Pr Dourthe) in both images. The result is presented in figure (7)

4.2 Registrations

To register the bones in the two images, we extracted the crest lines on the bones surface and some particular points on these lines called extremal points [TG93]. These points optimize a differential geometry criterion and can be seen as a generalization of corner points for smooth surfaces. As they are points on a surface, we have not only a point but also the normal to the surface and the two principal directions, i.e. a *semi-oriented frame*. Each bone is thus modeled as a set of semi-oriented frames.

The matches between the extremal points of a bone in one image and their counterpart in the second images were determined using a geometric hashing method described in [GPA97]. We found only 30 matches for the hipbones and 35 for the sacrum Then we used the pose estimation algorithm described in the previous section to compute the transformations. In figure (8) we present the superimposition of the pelvis obtained using the sacrum transformation only Considering that the sacrum is fixed, we can visualize the hipbones motion: the hipbones in the "open pelvis" configuration effectively seem to be slightly more open than in the "closed pelvis" configuration The residual transformation between hipbones confirms that impression since it is a rotation of 1.2 deg around the vertical axis and a translation of 3 to 4 mm that tends to open the hipbones.

4.3 Statistical results

However, when it comes to the uncertainty, the results are much less clear The uncertainty estimated on features is rather large: the standard deviation of the residual transformation is around 25 degrees for the rotation (trihedron) and 1mm for the translation (position) This leads to a very high uncertainty on the estimated transformations: around 3.5 degrees for the rotation and 1.5mm for the translation Comparing the residual transformations (left and right hipbones with respect to the sacrum or left hipbone with respect to the right hipbone) with the identity (rigid hypothesis: no motion inside the pelvis), we found Mahalanobis distances of 16 and 18, which approximately corresponds to the threshold of χ^2 tests at 98%. Thus, we cannot conclude that there is a statistically significant relative motion of bones in the pelvis. In this example, the availability of higher order statistics on the registration prevented us from a possibly wrong conclusion. However, we cannot conclude either that there is no relative motion in the pelvis. In fact, we would need more precise data to be able to conclude statistically. This could be obtained by using images with a better resolution, a better (semi-automatic) segmentation of the bones, and more extremal points on each bone surface.

5 Conclusion

We show in this article how to generalize the classical least squares and least Mahalanobis distance pose estimation algorithms to generic geometric features, and how to estimate the uncertainty of the result. The analysis of the three new algorithms shows that each one has advantages and drawbacks. The uncertainty prediction on the transformation is validated within a bound of 5% for all algorithms, decreasing to less than 1% of inaccuracy for the algorithm MAHA

Then, we show how to combine all the algorithms into a modular and high level registration algorithm to exploit all their particularities The accuracy of the transformations uncertainties is not considered in this article but is perfectly validated in [Pen96, chap. 9]. We believe that these algorithms could be easily embedded in many applications and provide a thorough basis for computing many image statistics.

In the last part, we show an application in medical imaging where the availability of the transformation uncertainties prevents us from a possible wrong conclusion This also stress that the quality of the registration directly depends on the quality and the number of features used. By the way, we could think to evaluate the quality of different feature extraction methods by the accuracy of the estimated pose.

Further improvements could include the generalization of this framework to non rigid transformations, the estimation of a multiple registration of n sets of matched features and the coupling with statistical matching algorithms to provide a complete registration system

References

[AHB87] K.S. Arun, T.S. Huang, and S.D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–700, September 1987

[Aya91] N. Ayache. *Artificial Vision for Mobile robots - Stereo-vision and Multisensor Perception.* MIT-Press, 1991

[DW88a] H.F Durrant-Whyte. *Integration, Coordination and Control of Multi-Sensor Robot Systems.* Kluwer Academic Publishers, 1988.

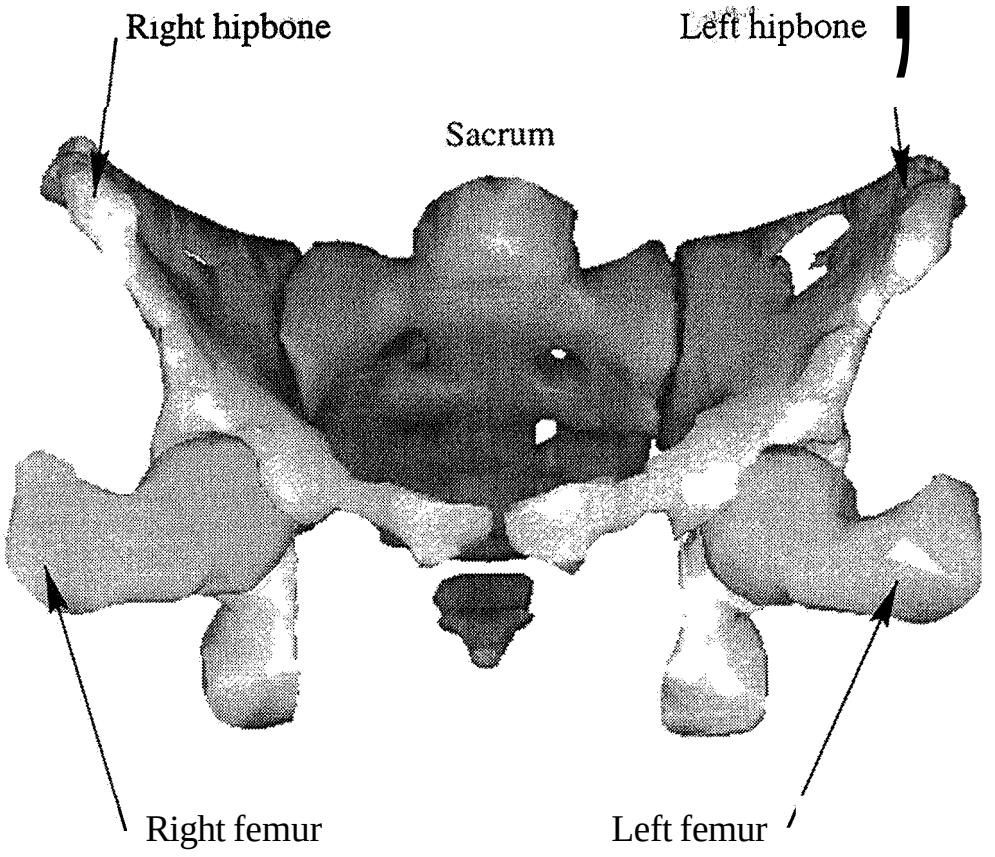


Figure 7 Result of the manual segmentation of one MRI image The surface of each bone is available independently and can be labeled as in this view

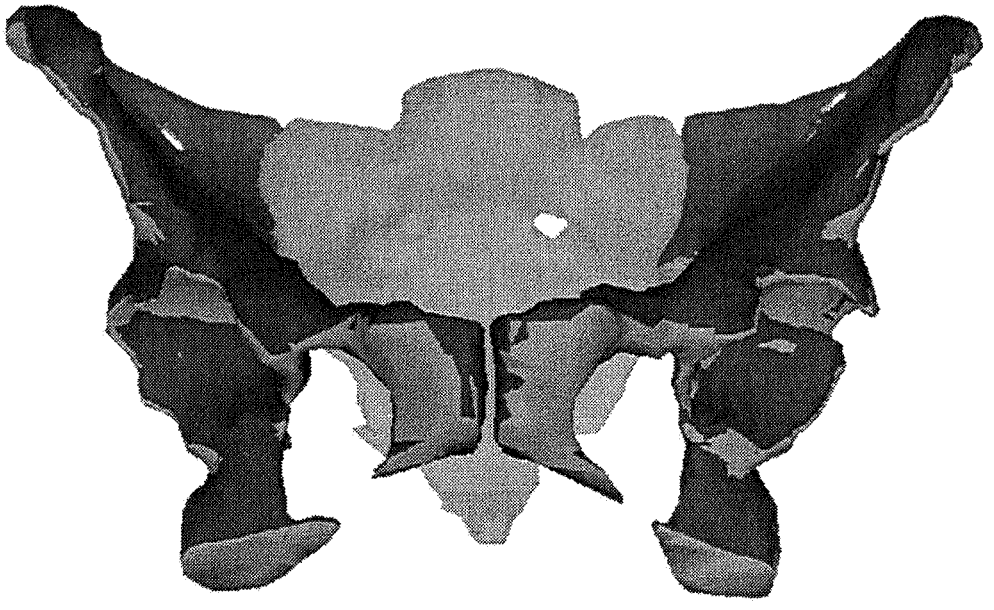


Figure 8: Superimposition of the "open" and "closed" pelvis using the sacrum registration the hipbones in the "open pelvis" configuration (in bright) effectively seem to be slightly more open than in the "closed" configuration (in dark).

[DW88b] H.F Durrant-Whyte. Uncertain geometry in robotics. *IEEE Journal of Robotics and Automation*, 4(1):23–31, February 1988.

[FA96] J. Feldmar and N. Ayache. Rigid, affine and locally affine registration of free-form surfaces *IJCV*, 18(2):99–119, May 1996.

[Fau93] O Faugeras. *Three-Dimensional Computer Vision. A Geometric Viewpoint.* MIT press, 1993.

[FH84] J Fang and T.S. Huang. Some experiments on estimating the 3-d motion parameters of a rigid body from two consecutive image frames. *IEEE Transactions on Pattern Analysis and Machine Intellzgence*, 6(5):545–554, September 1984.

[GMW81] P.E. Gill, W Murray, and M.H. Wright. *Practical optimization.* Academic Press, London and New York, 1981

[GPA97] A Guéziec, X Pennec, and N Ayache. Medical image registration using geometric hashing at INRIA *IEEE Computer Sczence and Engineering, special issue on Geometric Hashing*, 1997 To appear

[Gri90] W.E.L. Grimson. *Object Recognition by Computer The role of Geometric Constraznts.* MIT Press, 1990.

[HHN88] B.K.P Horn, Hilden H.M., and S. Negahdaripour Closed form solutions of absolute orientation using orthonormal matrices *Journal of Optical Society of America*, A-5(7):1127–1135, 1988.

[HJL+89] R.M Haralick, H. Joo, C.N Lee, X Zhuang, V.G Vaidya, and M.B. Kim Pose estimation from corresponding point data. *IEEE transactzon on Systems, Man and Cybernetics*, 19(6) 1426–1446, November,December 1989

[Hor87] B.K.P Horn Closed form solutions of absolute orientation using unit quaternions. *Journal of Optical Soczety of America*, A-4(4):629–642, April 1987

[Jaz70] A.M Jazwinsky. *Stochastic Processes and Filterzng Theory.* Academic Press, 1970.

[Kan93] K. Kanatani. *Geometric Computation for Machine Vision.* Number 37 in The Oxford Engineering science series. Clarendon Press - Oxford, 1993.

[Kan94] K. Kanatani Analysis of 3-d rotation fitting. *IEEE Transactzons on Pattern Analysas and Machine Intelligence*, 16(5):543–549, May 1994.

[LEF95] A Lorusso, D.W Eggert, and R.B. Fisher A comparison of four algorithms for estimating 3-d rigid transformation In *Proc. 6th British Machine Viszon Conference (BMVC'95)*, pages 237–246, 1995.

[PA96] X Pennec and N Ayache. An $\mathcal{O}(n^2)$ algorithm for 3D substructure matching of proteins. In A Califano, I. Rigoutsos, and H.J Wilson, editors, *Shape and Pattern Matching in Computatzonal Biology - Proc. First Int. Workshop, Seattle, Wash, June 20, 1994*, pages 25–40. Plenum Publishing, 1996. also as INRIA Research Report n° 2274.

[PA97] X. Pennec and N. Ayache. Uniform distribution, distance and expectation problems for geometric features processing. *Journal of Mathematical Imaging and Vision*, in press, 1997 Also as INRIA research report n° 2820.

[Pen96] Xavier Pennec. *L'Incertitude dans les Problèmes de Reconnaissance et de Recalage - Applications en Imagerie Médicale et Biologie Moléculaire.* PhD thesis, Ecole Polytechnique, Palaiseau (France), december 1996.
http://www.inria.fr/epidaure/personnel/pennec/These.html.

[PFTV91] W.H Press, B.P Flannery, S.A Teukolsky, and W T Vetterling. *Numerical Recapaces an C.* Cambridge Univ Press, 1991

[PT97] X Pennec and J.P Thirion. A framework for uncertainty and validation of 3D registration methods based on points and frames. *Int. Journal of Computer Vision*, 25(3), 1997

[Sch66] P H Schonemann. A generalized solution of the orthogonal Procrustes problem *Psychometrika*, 31 1–10, 1966.

[Sny89] M.A Snyder The precision of 3-d parameters in correspondence-based techniques: The case of uniform translational motion in a rigid environment *IEEE Transactions on Pattern Analysis and Machzne Intelligence*, 11(5):523–528, May 1989.

[TG93] J-P Thirion and A Gourdon The 3d marching lines algorithm new results and proofs. Technical Report 1881, INRIA, March 1993. accepted for publication in CVGIP in august 1994

[Ume91] S. Umeyama. Least-squares estimation of trans formation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380, April 1991

[WS91] M.W Walker and L. Shao. Estimating 3-D location paramters using dual number quaternions. *CVGIP. Image Understanding*, 54(3):358–367, Nov 1991

[ZF92] Z. Zhang and O. Faugeras. *3D Dynamic Scene Analysas: a stereo based approach*, volume 27 of *Springer series in information science*, chapter 2: Uncertainty Manipulation and Parameter Estimation, pages 9–27 Springer Verlag, 1992.

Image Registration Using Wavelet-Based Motion Model

[1,3]Yu-Te Wu [1]Takeo Kanade [2]Jeffrey Cohn [3]Ching-Chung Li

[1]The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

[2]Dept. of Psychology and Psychiatry
University of Pittsburgh
Pittsburgh, PA 15260

[3]Dept. of Electrical Engr.
University of Pittsburgh
Pittsburgh, PA 15260

Abstract

An image registration algorithm using wavelet approximation as an motion model has been developed to estimate accurate dense optical flow from an image sequence. This wavelet motion model is particularly useful in estimating optical flows with large displacement. Traditional pyramid methods which use the coarse-to-fine image pyramid by image burring in estimating optical flow often produce incorrect results when the coarse-level estimates contain large errors that cannot be corrected at the subsequent finer levels. This happens when regions of low texture become flat or certain patterns result in spatial aliasing due to image blurring. Our method, in contrast, uses large-to-small full-resolution regions without blurring images, and simultaneously optimizes the coarser and finer parts of optical flow so that the large and small motion can be estimated correctly. We compare results obtained by using our method with those obtained by using one of the leading image registration methods, the Szeliski pyramid spline-based method. The experiments include cases of small displacement (less than 4 pixels under 128× 128 image size or equrvalent displacement under other image sizes), and those of large displacement (10 pixels). While both methods produce comparable results when the displacements are small, our method outperforms pyramid spline-based method when the displacements are large.

1 Introduction

This paper presents a method to register two images and estimate optical flow using coarse-to-fine wavelet representation, newly presented by Cai and Wang [5], as a motion model. The wavelet motion model represents motion vectors by a linear combination of hierarchical basis functions. The coarser-scale basis function has larger support while the her-scale basis function has smaller support. Corresponding to these variably sized supports, large-to-small full-resolution regions are used in image matching. In addition, the associated coefficients to be estimated have global and local influences in constructing the motion vectors.

The major feature of Cai-Wang basis functions is to directly transform any function into wavelet coefficients from coarse to fine scale. This differs from the conventional usage of wavelet transform which is carried out from fine-to-coarse for decomposition and then coarse-to-fine for reconstruction. This feature allows us to estimate the coefficients starting from coarsest scale by using the largest full resolution patches. This is particularly useful in estimating the optical flow due to large motion. At each iteration, we reconstruct the motion vectors from the first to the second image based on the estimated coefficients, and use them to warp the firts image. As finer-scale coefficients are estimated by minimizing the sum of squared intensity differences between the warped image and the second image, we obtain more accurate results.

To handle large displacement, a number of coarse-to-fine hierarchical. methods [2, 4, 1] have been developed. The pyramid methods which use coarse-to-fine blurred images sequentially in estimating optical flow often produce incorrect results when large errors occurring in coarser estimates cannot be subsequently corrected in the finer estimates. This happens when regions of low texture become flat or certain patterns result in spatial aliasing due to image blurring. The reason that our method could cope with this difficulty is that our representation hierarchy is different from the typical image coarse-to-fine hierarchy. Our hierarchy is built along motion resolution level, not image resolution level. In addition, we use large-to-small full resolution regions, rather than blurred images, in estimating the coefficients from coarse to fine scales.

2 Mathematical Background

Let $\mathbf{I} \stackrel{\text{def}}{=} [0, L]$ denote any finite interval, where $L \geq 4$ is a positive integer, and $H^2(I)$ denote Sobolev space which contains all continuous functions with finite energy norm up to the second derivative, respectively, i.e.,

$$H^2(I) \stackrel{\text{def}}{=} \{f(x), x \in I | \|f^{(i)}\|_2 = \int_I |f^{(i)}(x)|^2 dx < \infty, i = 0, 1, 2.\}$$

It can be verified that $H^2(I)$ is a Hilbert space with the inner product, $< f, g >= \int_I f^{(2)}(x) g^{(2)}(x) dx$. Hence $|||f||| = < f, f >^{\frac{1}{2}}$ provides a norm for $H^2(I)$ In order to develop a coarse-to-fine wavelet representation in $H^2(I)$, Cai and Wang used the fourth-order B-spline $\phi(x)$ (Figure 1.(a))

$$\phi(x) \stackrel{\text{def}}{=} \frac{1}{6} \sum_{j=0}^{4} \binom{4}{j} (-1)^j (x - j)_+^3 \in H^2(I) \qquad (1)$$

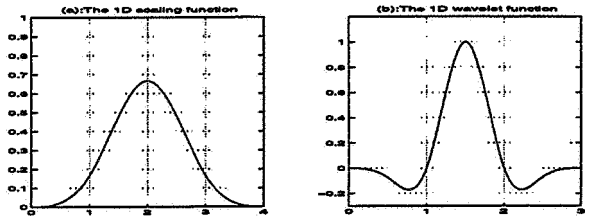


Figure 1: The 1D scaling and wavelet functions.

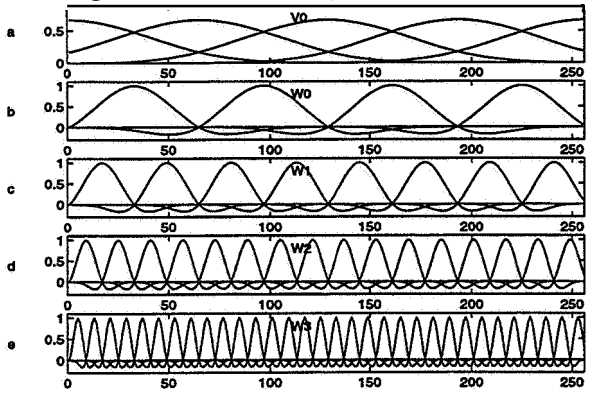


Figure 2: The translation and dilation of basis functions at different resolution scales

as the scaling function, where $x_+^n = x^n$ if $x \geq 0$, and $x_+^n = 0$ if $x < 0$. Using the scaling function, they constructed the wavelet (Figure 1.(b))

$$\psi(x) = \frac{-3}{7}\phi(2x) + \frac{12}{7}\phi(2x-1) - \frac{3}{7}\phi(2x-2) \quad (2)$$

The supports of $\phi(x)$ and $\psi(x)$ are $[0,4]$ and $[0,3]$, respectively, i.e., the values of $\phi(x)$ (or $\psi(x)$) are zeros outside $[0,4]$ (or $[0,3]$). The dilation and translation of $\phi(x)$ and $\psi(x)$ are defined by

$$\begin{aligned} \phi_{j,k}(x) &= \phi(2^j x - k), j \geq 0, \ k = -2, \quad , 2^j L - 2 \\ \psi_{j,k}(x) &= \psi(2^j x - k), j \geq 0, \ k = -1, \quad 2^j L - 2 \end{aligned} \quad (3)$$

For each resolution level j, let V_j and W_j be the closure (under norm $|||f|||$) of hear span of $\{\phi_{j,k}(x), -2 \leq k \leq 2^j L - 2\}$ and $\{\psi_{j,k}(x) | k = -1, \quad , 2^j L - 2\}$, respectively. Cai and Wang have shown that[5] any continuous function $d(x) \in H^2(I)$ can be approximated as closely as possible by a function in V_j for a sufficiently large j which has a unique orthogonal approximation

$$d(x) \approx d_{-1}(x) + d_0(x) + d_1(x) + \quad d_j(x) \quad (4)$$

where

$$d_{-1}(x) = \sum_{k=-2}^{L-2} c_{-1,k}\phi_{0,k}(x) \quad \in V_0 \quad (5)$$

$$d_j(x) = \sum_{k=-1}^{2^j L-2} c_{j,k}\psi_{j,k}(x) \quad E\ W_j \quad \text{for} \quad j \geq 0. \quad (6)$$

Each component d_j is produced by the expansion of translated basis functions at different resolution levels where the coefficients $c_{j,k}$ are appropriately determined. Although the existence of approximation is proved for any continuous function in $H^2(I)$, it has been demonstrated in [5, 8] that it holds for any finite sampled function in the practical applications.

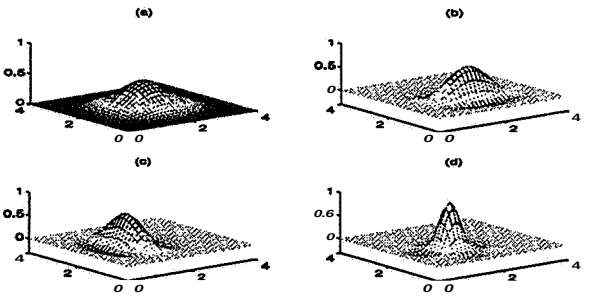


Figure 3: The two dimensional basis functions. (a) $\phi(x)\phi(y)$, (b) $\phi(x)\psi(y)$, (c) $\psi(x)\phi(y)$ and (d) $\psi(x)\psi(y)$.

Influence of coefficients $c_{j,k}$ can be global or local depending on the size of support of the corresponding basis function. To see this, let us assume $L = 4$ for simplicity such that the scaling functions $\phi(x-k)$ in the coarsest scale have support $[0,4]$ (except on the boundary), and the wavelet functions $\psi_{j,k}$ have (narrower) support $[0, \frac{3}{2^j}]$ (except near the boundary), $j \geq 0$. Let us also assume a finite sampled function defined on pixel domain $[0,256]$. We map $[0,4]$ into $[0,256]$ so that the effects are interpreted on pixel unit. The perturbs of $c_{-1,-2}$, $c_{-1,-1}$, $c_{-1,0}$, $c_{-1,1}$ and $c_{-1,2}$ will affect the values $d_{-1}(x)$ globally within the domain $[0,127]$, $[0,191]$, $[0,256]$, $[65,256]$ and $[129,256]$, respectively (Figure 2). In the finer scales (levels) $j \geq 0$, wavelets have support $\frac{192}{2^j}$ pixels within the domain (and $\frac{128}{2^j}$ pixels on the boundaries), and there are total 2^{j+2} shifted wavelets overlapped across $[0,256]$. The perturbs of each wavelet coefficient $c_{j,k}$ will affect $d_j(x)$ locally only inside the domain $[\frac{64k}{2^j}, \frac{64(k+3)}{2^j}]$, and on the boundraries $[0, \frac{128}{2^j}]$ and $[256 - \frac{128}{2^j}, 256]$.

3 Wavelet-Based Image Registration

Suppose image $I_0(x,y)$ and $I_1(x,y)$ are related by horizontal and vertical displacement (flow or motion) vectors $u(x,y)$ and $v(x,y)$ under the intensity constancy constraint [3]. We wish to recover the maximum likelihood estimates $\{u(x,y), v(x,y)\}$ by minimizing the sum of squared intensity differences (SSD)

$$E = \sum_{x,y} [I_1(x + u(x,y), y + v(x,y)) - I_0(x,y)]^2 \quad (7)$$

between the reference image $I_0(x,y)$, and the predicted image $I_1(x+u(x,y), y+v(x,y))$.

3.1 Wavelet-Based Motion Model

We will approximate motion vectors $u(x,y)$ and $v(x,y)$ by using two-dimensional basis functions. A natural extension of one-dimensional to two-dimensional basis functions are using the tensor product. Accordingly, the two-dimensional basis functions are (Figure 3)

$$\Phi_{0,k_1,k_2}(x,y) = \phi(x-k_1)\phi(y-k_2) \quad (8)$$

$$\Psi^H_{j,k_1,k_2}(x,y) = \phi(2^j x - k_1)\psi(2^j y - k_2) \quad (9)$$

$$\Psi^V_{j,k_1,k_2}(x,y) = \psi(2^j x - k_1)\phi(2^j y - k_2) \quad (10)$$

$$\Psi^D_{j,k_1,k_2}(x,y) = \psi(2^j x - k_1)\psi(2^j y - k_2) \quad (11)$$

where the subscripts j, k_1 and k_2 represent the resolution scale, horizontal and vertical translations, respectively, and

the upper subscripts H, V and D represent the horizontal, vertical and diagonal directions, respectively. Similar to the one-dimensional case, any two dimensional motion vector can be expressed in terms of linear combinations of coarsest-scale spline function (8) and horizontal, vertical and diagonal wavelets ((9), (10) and (11)) in finer levels[8]

$$u(x,y) = u_{-1}(x,y) + \sum_{j=0}^{J}(u_j^H(x,y) + u_j^V(x,y) + u_j^D(x,y))$$

$$v(x,y) = v_{-1}(x,y) + \sum_{j=0}^{J}(v_j^H(x,y) + v_j^V(x,y) + v_j^D(x,y))$$

where

$$u_{-1}(x,y) = \sum_{k_1=-2}^{L_1-2}\sum_{k_2=-2}^{L_2-2} c_{-1,k_1,k_2}\Phi_{0,k_1,k_2}(x,y)$$

$$u_j^H(x,y) = \sum_{k_1=-2}^{2^jL_1-2}\sum_{k_2=-1}^{2^jL_2-2} c_{j,k_1,k_2}^H\Psi_{j,k_1,k_2}^H(x,y)$$

$$u_j^V(x,y) = \sum_{k_1=-1}^{2^jL_1-2}\sum_{k_2=-2}^{2^jL_2-2} c_{j,k_1,k_2}^V\Psi_{j,k_1,k_2}^V(x,y)$$

$$u_j^D(x,y) = \sum_{k_1=-1}^{2^jL_1-2}\sum_{k_2=-1}^{2^jL_2-2} c_{j,k_1,k_2}^D\Psi_{j,k_1,k_2}^D(x,y)$$

$$L_1,\ L_2 \geq 4.$$

Functions v_{-1}, v_j^H, v_j^V and v_j^D have similar forms except that c_{-1,k_1,k_2}, c_{j,k_1,k_2}^H, c_{j,k_1,k_2}^V and c_{j,k_1,k_2}^D are replaced by d_{-1,k_1,k_2}, d_{j,k_1,k_2}^H, d_{j,k_1,k_2}^V and d_{j,k_1,k_2}^D, respectively.

3.2 The Algorithm

Given I_0 and I_1, the image registration is now a problem of estimating coefficients vector $\hat{\mathbf{c}} = [(\hat{\mathbf{c}}_{-1})^T \quad (\hat{\mathbf{c}}_j^{H,i})^T \quad (\hat{\mathbf{c}}_j^{V,i})^T \quad (\hat{\mathbf{c}}_j^{D,i})^T \quad \ldots]^T$ where $\hat{\mathbf{c}}_{-1}$ representing the coarsest-scale coefficients, and $\hat{\mathbf{c}}_j^{H,i}$, $\hat{\mathbf{c}}_j^{V,i}$ and $\hat{\mathbf{c}}_j^{D,i}$ representing any finer-scale coefficients in horizontal, vertical and diagonal directions, respectively. This is done by substituting $u(x,y)$ and $v(x,y)$ into (7) and minimizing SSD with respect to E iteratively. The motion vectors are reconstructed using the estimated coefficients. We then use the motion vectors to warp image I_1 toward I_0. In the next finer scale, we estimate the coefficients from coarsest to current scale where the large-to-small regions are utilized. To handle large displacement, we first estimate the coarsest-scale coefficients where the largest regions are used.

Let vector

$$\hat{\mathbf{c}}_{-1}^i = [c_{-1,-2,-2}^i, \ldots, c_{-1,L_1-2,L_2-2}^i, d_{-1,-2,-2}^i, \ldots, d_{-1,L_1-2,L_2-2}^i]^T$$

be the current estimated coarsest-scale coefficients during the ith iteration. From now on, the upper subscript i denotes the ith iteration. The incremental estimate $\delta\hat{\mathbf{c}}_{-1}$ can be obtained by minimizing a quadratic measure Using the differential method. By using the first order Taylor series expansion

$$I(x+u^i+\delta u, y+v^i+\delta v) \approx I(x+u^i, y+v^i, t) + \delta u I_x + \delta v I_y,$$

the incremental quadratic error is

$$E(\delta\hat{\mathbf{c}}_{-1}) = E(\hat{\mathbf{c}}_{-1}^i + \delta\hat{\mathbf{c}}_{-1}) - E(\hat{\mathbf{c}}^i) \approx \delta\hat{\mathbf{c}}_{-1}^T A\delta\hat{\mathbf{c}}_{-1} - 2\mathbf{b}^T\delta\hat{\mathbf{c}}_{-1} \quad (12)$$

where

$$(I_x, I_y) \stackrel{\text{def}}{=} \left(\frac{\partial I_1(x+u^i, y+v^i)}{\partial x}, \frac{\partial I_1(x+u^i, y+v^i)}{\partial y}\right)$$

$$A \stackrel{\text{def}}{=} \sum_{x,y}\mathbf{a}(x,y)\mathbf{a}(x,y)^T, \quad \text{Hessian matrix} \quad (13)$$

$$\mathbf{a}(x,y) \stackrel{\text{def}}{=} [\Phi_{0,-2,-2}I_x \quad \Phi_{0,L_1-2,L_2-2}I_x \quad \Phi_{0,-2,-2}I_y \quad \Phi_{0,L_1-2,L_2-2}I_y]^T, \quad (14)$$

$$\mathbf{b} \stackrel{\text{def}}{=} -\sum_{x,y}(I_1(x+u^i, y+v^i) - I_0(x,y))\mathbf{a}(x,y),$$

$$u^i(x,y) = u_{-1}^i(x,y),\ v^i(x,y) = v_{-1}^i(x,y) \quad (15)$$

To obtain Hessian matrix A and gradient vector $\mathbf{b}$ during each iteration in every resolution scale, we first calculate the warped image $I_1(x+u^i, y+v^i)$ using the updated motion vector $u^i(x,y)$ and $v^i(x,y)$, and bilinear interpolation. The gradient (I_x, I_y) and the residual difference, $I_1(x+u^i, y+v^i) - I_0(x,y)$, are then computed. When $E(\delta\hat{\mathbf{c}}_{-1})$ is minimized using the Levenberg-Marquardt algorithm[6], the resultant coefficients $\hat{\mathbf{c}}_{-1}^i$ are used as the initial guess and propagated into the next finer level.

In the next finer level, we update the coefficients estimated at the previous coarsest level, and estimate three sets of coefficients at the current level one by one. First, we denote the first set of coefficients at current level ($j=0$) by

$$\hat{\mathbf{c}}_0^{H,i} = [c_{0,-1,-1}^{H,i} \quad \ldots \quad c_{0,L_1-2,L_2-2}^{H,i} \quad d_{0,-1,-1}^{H,i} \quad \ldots \quad d_{0,L_1-2,L_2-2}^{H,i}]^T,$$

and estimate $\hat{\mathbf{c}}^i = [(\hat{\mathbf{c}}_{-1}^i)^T \ (\hat{\mathbf{c}}_0^{H,i})^T]^T$ where $\hat{\mathbf{c}}_{-1}^i$ has been initialized from the previous step. The coefficients $\hat{\mathbf{c}}^i$ which have more local influence, are used to reconstruct the motion vectors $u(x,y) = u_{-1}(x,y) + u_0(x,y)$ and $v(x,y) = v_{-1}(x,y) + v_0(x,y)$. The incremental estimate $\delta\hat{\mathbf{c}}$ can be obtained by minimizing the same form of quadratic cost function (12) except that new terms $\Psi_{0,-2,-1}^H I_x \quad \Psi_{0,L_1-2,L_2-2}^H I_x$, $\Psi_{0,-2,-1}^H I_y \quad \Psi_{0,L_1-2,L_2-2}^H I_y$ are included in $\mathbf{a}(x,y)$, and $(u_0^{H,i}(x,y), v_0^{H,i}(x,y))$ are include in $(u^i(x,y), v^i(x,y))$. More precisely, they are

$$\mathbf{a}(x,y) \stackrel{\text{def}}{=} [\Phi_{0,-2,-2}I_x \quad \Phi_{0,L_1-2,L_2-2}I_x \quad \Psi_{0,-2,-1}^H I_x \quad \Psi_{0,L_1-2,L_2-2}^H I_x \quad \Phi_{0,-2,-2}I_y \quad \Phi_{0,L_1-2,L_2-2}I_y \quad \Psi_{0,-2,-1}^H I_y \quad \Psi_{0,L_1-2,L_2-2}^H I_y]^T$$

$$u^i(x,y) = u_{-1}^i(x,y) + u_0^{H,i}(x,y),\ v^i(x,y) = v_{-1}^i(x,y) + v_0^{H,i}(x,y)$$

When $E(\delta\hat{\mathbf{c}})$ attains minimum, the resultant coefficients E are propogated into the next step where the second set coefficients are estimated.

Then, we estimate the second set of coefficients

$$\hat{\mathbf{c}}_0^{V,i} = [c_{0,-1,-2}^{V,i} \quad \ldots \quad c_{0,L_1-2,L_2-2}^{V,i} \quad d_{0,-1,-2}^{V,i} \quad \ldots \quad d_{0,L_1-2,L_2-2}^{V,i}]^T,$$

and update the previous estimated coefficients $\hat{\mathbf{c}}_{-1}^i$ and E simultaneously. The concatenated coefficient vector is $\hat{\mathbf{c}}^i = [(\hat{\mathbf{c}}_{-1}^i)^T \ (\hat{\mathbf{c}}_0^{H,i})^T \ (\hat{\mathbf{c}}_0^{V,i})^T]^T$ The incremental estimate $\delta\hat{\mathbf{c}}$ can

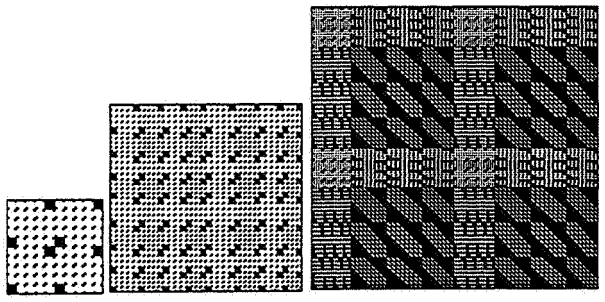


Figure 4 The left, middle and right figures are the forms of hessian matrix when one, two and three levels are used. White and black regions represent nonzero and zero entries, respectively.

be obtained by minimizing the Same form of cost function (12), except that new terms $\Psi^V_{0,-1,-2}I_x \quad \Psi^V_{0,L_1-2,L_2-2}I_x$ and $\Psi^V_{0,-1,-2}I_y \quad \Psi^V_{0,L_1-2,L_2-2}I_y$ are included in previous a(x, y), and $(u_0^{V,i}(x,y), v_0^{V,i}(x,y))$ are included in previous $(u^i(x,y), v^i(x,y))$. When $E(\delta\hat{c})$ attains minimum, the resultant coefficients $\hat{c}$ are propagated into the next step where the third set coefficients are estimated.

Lastly, we estimate the third set of coefficients $\hat{c}_0^{D,i} = [c^{D,i}_{0,-1,-1} \cdots c^{D,i}_{0,L_1-2,L_2-2} \cdot d^{D,i}_{0,-1,-1} \cdots d^{D,i}_{0,L_1-2,L_2-2}]^T$, and update the previous estimated coefficients $\hat{c}^i_{-1}$, $\hat{c}_0^{H,i}$ and $\hat{c}_0^{V,i}$ The concatenated coefficient vector is $\hat{c}^i = [(\hat{c}^i_{-1})^T \ (\hat{c}_0^{H,i})^T \ (\hat{c}_0^{V,i})^T \ (\hat{c}_0^{D,i})^T]^T$ The incremental estimate $\delta\hat{c}$ can be obtained by minimizing the same form of quadratic cost function except, that new terms $\Psi^D_{0,-1,-1}I_x \quad \Psi^D_{0,L_1-2,L_2-2}I_x$ and $\Psi^D_{0,-1,-1}I_y \quad \Psi^D_{0,L_1-2,L_2-2}I_y$ are added to previous a(x, y), while $(u_0^{D,i}(x,y), v_0^{D,i}(x,y))$ are added to previous $(u^i(x,y), v^i(x,y))$.

In general. at level j, we estimate the coefficients from the coarsest to the current level. At each level ($j \geq 0$), three sets of coefficients are estimated one by one. The estimation procedure is the same as we described above. When $E(\delta\hat{c})$ attains minimum, we check the SSD prediction error and decide to either stop or continue the process.

3.3 Variably Sized Supports and Hessian Matrix

The structure of Hessian matrix reveals the usage of global and local information and how they are interactive with each other. The structure of Hessian matrices using one, two and three levels are shown in Figure 4. The nonzeros and zeros elements are presented as white and black regions, respectively. We see that each Hessian matrix can be divided into four patterns, say A_{11}, A_{12}, A_{21} and A_{22} representing left upper, right upper, left lower and right lower patterns, respectively. Let us denote the first and second half parts of vector $\mathbf{a}$ by $\mathbf{a_1}$ and $\mathbf{a_2}$ which are associated with coefficients c_j's and d_j's, respectively. Patterns A_{11} and A22 are formed by $\sum_{x,y} \mathbf{a_1 a_1}^T$ and $\sum_{x,y} \mathbf{a_2 a_2}^T$, respectively, pattern A_{12} is formed by $\sum_{x,y} \mathbf{a_1 a_2}^T$, and A12 is equal to A_{21}. A_{11} and A_{12} are used to calculate coefficients c_{j,k_1,k_2}'s, and A_{21} and A_{22} are used to calculate Coefficients d_{j,k_1,k_2}'s. Each entry in the Hessian matrix is produced by the multiplications of any two basis functions with each other, and with two image gradients. As a result, the entry is nonzero when the associated two basis functions are overlapped and two image gradients are nonzeros. In Figure 4, we observe that as the resolution scale moves up, the basis functions have narrower supports and fewer overlaps with functions in coarser and current scale. Therefore, the white regions becomes sparser. Since the block diagonal nonzero entries of any pattern are produced by any two overlapped basis functions at the same scale and the remaining blocks are produced by the overlaps of different basis functions from coarsest to current scales. This implies that in the finer scales not only smaller but also larger patches are optimally used in the sense that the coarser- and finer-scale coefficients minimize SSD.

4 Experimental Results

This section compares the results produced by wavelet-based and spline-based methods.

4.1 Synthetic Image Pairs Containing Small Displacement

In the first experiment, we use the synthetic images in paper [1], where the ground truth of optical flow vectors is provided. In the pyramid spline-based method, the low-pass Gaussian filter with mask size 3 x 3 pixels is successively applied to create the coarse-to-fine blurred images. Two error measurements we use are the angular measure θ_e [1] and the magnitude measure m_e, defined by

$$\theta_e = \arccos\left(\frac{uu_t + vv_t + 1}{\sqrt{u^2+v^2+1}\sqrt{u_t^2+v_t^2+1}}\right) \quad (16)$$

$$m_e = |\sqrt{u^2+v^2} - \sqrt{u_t^2+v_t^2}| \quad (17)$$

where $(u_t(x,y), v_t(x,y))$ is the correct motion vector For each image pair, we compute the averaged angle and magnitude errors ($\overline{\Delta\theta_e}$ and $\overline{Am}$), using the flow vector at each pixel. We have tested the image pairs of Sinusoidal 1, Translating Tree, Diverging Tree and Yosemite. Figures 5-8 show the estimated motion vectors using our method. To compare our method with pyramid spline-based methods, the number of unknowns is designed to be the same. We name the flow estimates produced by spline-based and by our methods as spline and wavelet flow, respectively. The quantitative comparisons between pyramid splin-based and wavelet-based methods are shown in the Table. In the Sinusoid 1image pair, the one level spline flow is more accurate than the one level wavelet flow but less accurate than the three level wavelet flow The three level spline flow has large errors, $(71.37^0, 19.043)$, which are propagated from the coarsest estimates due to the significant texture change. In the Translating Tree and Diverging Tree image pairs, one level spline flow is less accurate than the wavelet flow because of incorrect estimates around the poor texture regions. In the Yosemite image pair, the three level spline flow is better than the wavelet flow However, we obtained better results when 4 levels were used. In general, both flow estimates are accurate and comparable, and both methods outperform other methods in [1] and method in [7]. (see [4] and [8]).

4.2 Synthetic Image Pair Containing Large Displacement

In the second experiment, the input image (Figure 9) is created by cropping part of the Sinusoidal 1 and the Tree images where the sinusoidal pattern has a displacement of 1.585 pixels upward and 0.863 pixel rightward, and the tree

Image Pairs	Spline-based method		Wavelet-based method	
	$\Delta\theta_e$	Δm_e	$\Delta\theta_e$	Δm_e
Sinusoid 1	0.22^0 (1)*	0.54 (1)	0.40^0 (1)	0.03 (1)
	71.37^0 (3)	19.043 (3)	0.056^0 (3)	0.0021 (3)
Translating	13.27^0 (1)	0.79 (1)	0.45^0 (1)	0.026 (1)
Tree	0.59^0 (3)	0.0210 (3)	0.85^0 (3)	0.030 (3)
Diverging	4.23^0 (1)	0.065 (1)	1.33^0 (1)	0.0254 (1)
Tree	1.52^0 (3)	0.0254 (3)	2.49^0 (3)	0.030 (3)
Yosemite	4.38^0 (3)	0.136 (3)	4.63^0 (3)	0.137 (3)
			3.54^0 (4)	0.097 (4)

*The number inside the parentheses denotes the image level(s) used in the spline-based method, or the motion resolution level(s) in wavelet-based method.

pattern has 10 pixels displacements to the right. The results in the left column of Figure 10 show that large errors occurring at level 1 due to spatial aliasing are propagated to the final result. The results using wavelet method are shown in the right column of Figure 10. The flow vectors of the sinusoidal 1 and Tree patterns are well recovered.

4.3 Real Image Sequences

The results of three real image sequences, SRI, NASA, Hamburg Taxi sequences in [1] are presented in Figures 11, 12, and 13, respectively. Flow of ten image frames is presented for each image sequence. Overall, the results look reasonable. Finally, the left and right columns in Figure 14 show the computed optical flow representing surprised and smile expressions with head movement using the wavelet method. The optical flow results of 6 basic facial expressions can be seen in www.cs.cmu.edu/afs/cs.cmu.edu/user/ytw/www/facial.html and [8].

5 Conclusions

We have used the wavelet model to develop a multiple-window, coarse-to-fine registration algorithm. The wavelet basis functions play triple roles in our algorithm, allowing us to construct motion vectors from coarse-to-he, select multiple large-to-small regions for image matching and impose smoothness. Since the motion vector at each pixel is obtained by simultaneously matching large-to-small full resolution regions, our algorithm produces robust and accurate results. The advantage has been demonstrated by comparing the results obtained by our method with those obtained by Szeliski pyramid spline method. This method has been applied successfully to the motion estimation of facial expressions[8], and we are in the process of extending our algorithm to other applications.

Acknowledgements

This research is supported by NIMH grant R01MH51435. We would also like to thank Marie Elm for reviewing the manuscript.

References

[1] Barron, J.L., Fleet, D.J., and Beauchemin, S.S. Performance of optical flow techniques, International Journal of Computer Vision, vol. 12, 1994, pp. 43–77

[2] Bergen, J R., Anandan, P., Hanna, K. J., and Hingorani R. Hierarchical model-based motion estimation, in: G. Sandini, ed., Computer Vision-ECCV '92, Springer, Berlin, 1992, pp. 237-252.

[3] Horn, B.K.P and Schunck, B.G. Determining optical flow: A retrospective, Artificial Intelligence, vol. 17, 1981, pp. 185–203.

[4] Szeliski, R. and Coughlan, J. Spline-Based Image Registration International Journal of Computer Vision, 22(3), 1997, pp. 199-218.

[5] Cai, W and Wang J Adaptive multiresolution collocation methods for initial boundary value problems of nonlinear PDEs, SIAM NUMER. ANAL. Vol. 33, No. 3, pp. 937-970, June 1996.

[6] Press, W H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W T Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, Cambridge, 2nd, 1992

[7] Xiong, Y and Shafer, S A Moment and Hypergeometric Filters for High Precision Computation of Focus, Stereo and Opticai Flow, International Journal of Computer Vision, vol. 22(1), 1997, 25-59,

[8] Yu-Te Wu, Image Registration Using Wavelet-Based Motion Model and Its applications, PhD Dissertation, University of Pittsburgh, 1997

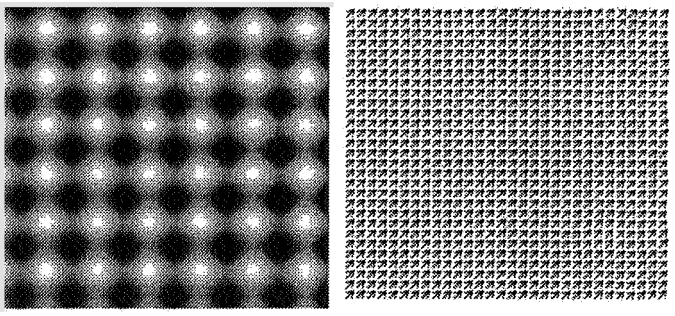


Figure 5: Sinusoidal 1

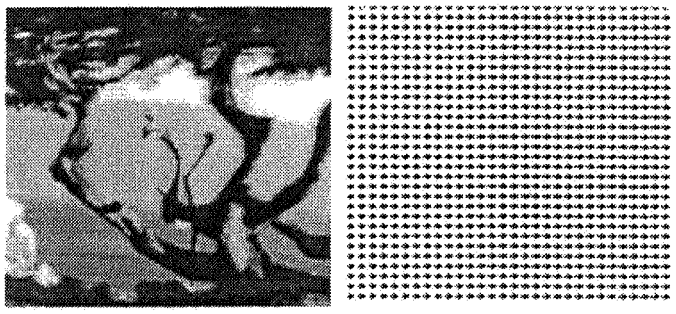


Figure 6: Translating Tree

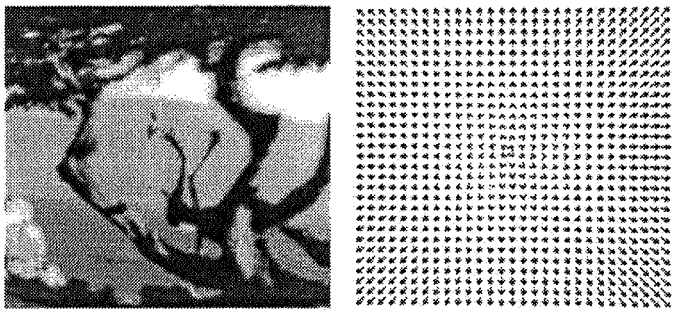


Figure 7: Diverging Tree

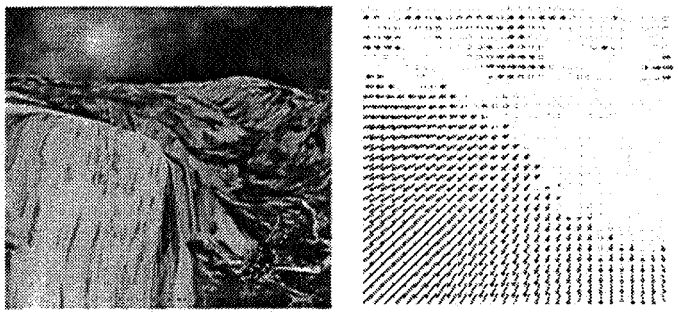


Figure 8: Yosemite

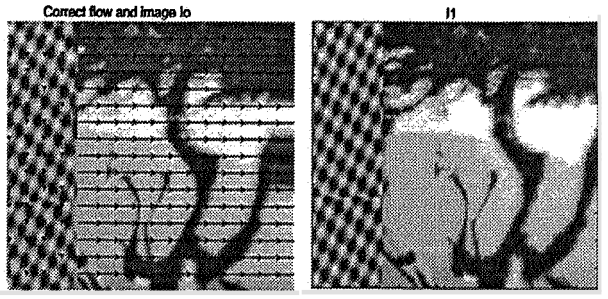


Figure 9: Snythetic image pair (128× 128) containing large and small motions. The sinusoidal pattern (left) has displacement 1.585 pixel upward and 0.863 pixel rightward, and the tree pattern (right) has 10 pixels displacements tc the right.

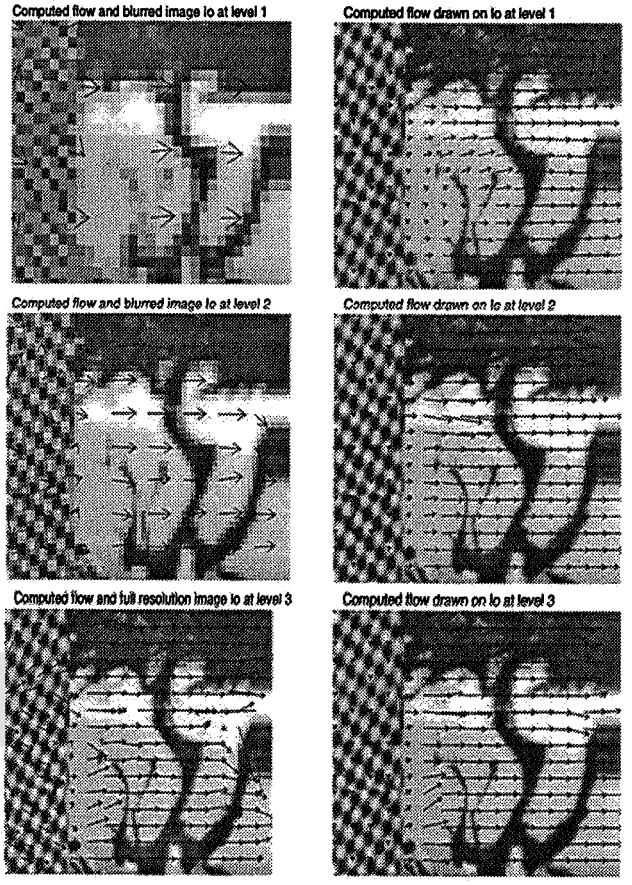


Figure 10 The comparison of pyramid spline-based (left column) and wavelet-based methods (right column). Observe that in left column the large errors which occurred in the left Sinusoidal 1 textured region at level 1 have been wrongly propagated to finer estimates. In the right column, the optical flow are recovered correctly.

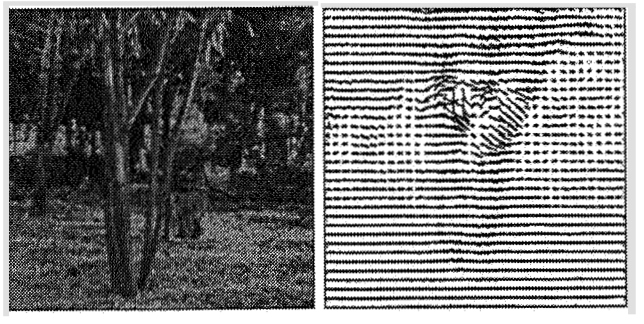


Fig. 11: SRI Trees.

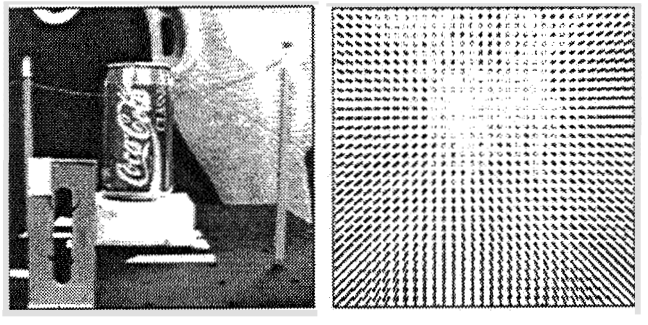


Figure 12: NASA

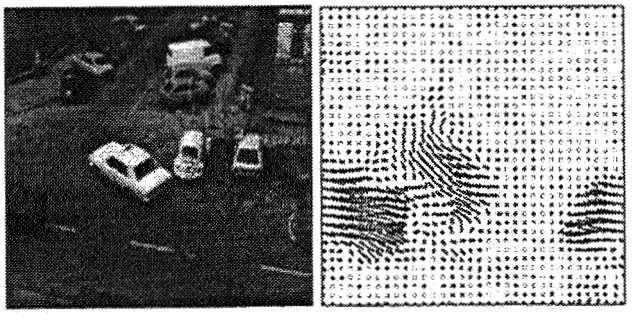


Figure 13: Hamburg Taxi.

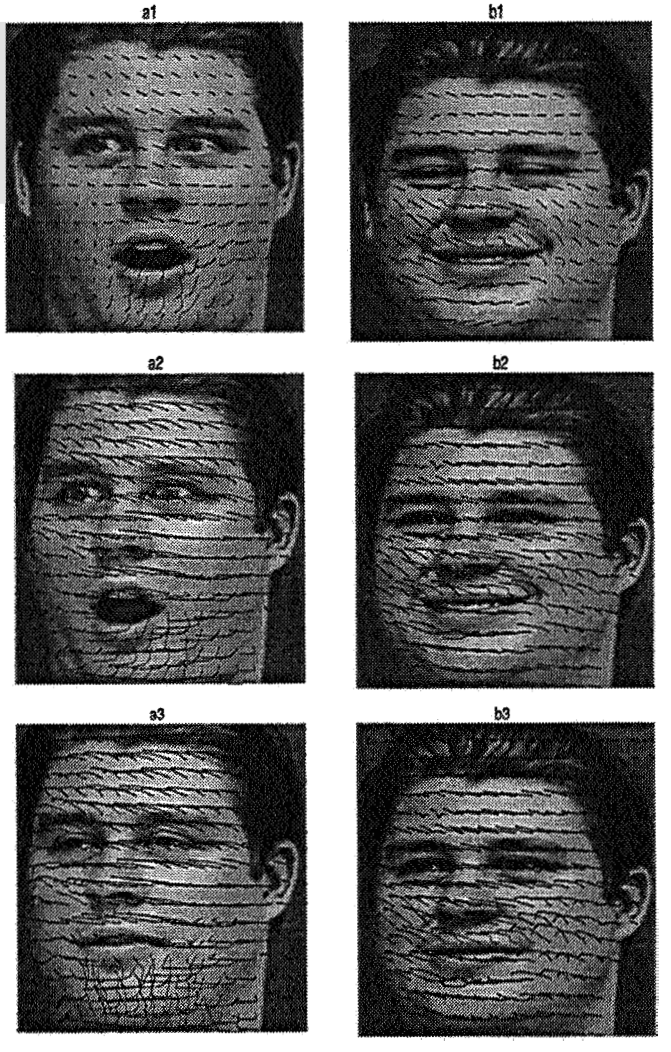


Figure 14 Two image sequences (left and right columns) containing large displacements. Each sequence has 25 frames. Three selected frames *are* shown from each sequence.

Recovery of motion parameters from distortions in scanned images

Jeffrey B. Mulligan
NASA Ames Research Center

S31-61
247979
p.12

Abstract: Scanned images, such as those produced by the scanning-laser ophthalmoscope (SLO), show distortions when there is target motion. This is because pixels corresponding to different image regions are acquired sequentially, and so, in essence, are slices of different snapshots. While these distortions create problems for image registration algorithms, they are potentially useful for recovering target motion parameters at temporal frequencies above the frame rate. Stetter, Sendtner and Timberlake (Vision Res, vol. 36, pp. 1987-1994, 1996) measured large distortions in SLO images to recover the time course of rapid horizontal saccadic eye movements. Here, this work is extended with the goal of automatically recovering small eye movements in two dimensions. Eye position during the frame interval is modeled using a low dimensional parametric description, which in turn is used to generate predicted distortions of a reference template. The input image is then registered to the distorted template using normalized cross correlation. The motion parameters are then varied, and the correlation recomputed, to find the motion which maximizes the peak value of the correlation. The location and value of the correlation maximum are determined with sub-pixel precision using biquadratic interpolation, yielding eye position resolution better than 1 arc minute (Mulligan, Behavior Research Methods, Instmments and Computers, vol. 29, pp. 54–65, 1997). This method of motion parameter estimation is tested using actual SLO images as well as simulated images. Motion parameter estimation might also be applied to individual video lines in order to reduce pipeline delays for a near real-time system.

1. Introduction

Video image sequences are often used to track object motion. Unless a special high frame-rate camera is used, the recovered motion is usually sampled in time at the video frame rate (50-60 Hz). While low resolution sampling is adequate for many applications, documentation of high-speed events often requires higher temporal resolution. For images obtained with a scanned system, in which individual pixel values are acquired at different times, it is possible to obtain higher temporal resolution for the motion of extended targets. The sequential nature of the scanning process introduces geometric distortions in the image of a moving target. By measuring these distortions, high temporal resolution information about the target motion can be recovered This technique is especially useful when *a priori* knowledge about the possible target motions permits a concise description using low-dimensional parametric models, because this reduces the space of possible distortions which must be searched. In the following sections, expressions for the precise form of the motion-induced distortions will be derived.

Address for author correspondence: MS 262-2, **NASA** Ames Research Center, **Moffett** Field, **CA, 94035-1000. Email** jbm@vision.arc.nasa.gov *An* electronic version *of this paper* with additional **multimedia figures** is available **at** http://vision.arc.nasa.gov/~jbm/papers/irw97.html.

11. Raster scanning

x, y	position in image plane
$s_x(t), s_y(t)$	position of scan at time t
f_L	line frequency (~ 15 kHz)
f_F	frame rate (~60 Hz)
i_L	index of current line
t_S	start time of current line
t_L	time in current line, $t - t_S$
$v_{S,x}, v_{S,y}$	scan velocities

Some imaging systems, using an electronic or mechanical shutter, can simultaneously capture all of the pixels in an image. In a scanned system, however, only a single point is sensed at a given time, and the location of this point is swept over the image area by electronic or mechanical means. Here we present some definitions and conventions that will allow us to precisely describe the scanning process.

The imaging area is defined to be a rectangle indexed by normal Cartesian coordinates x and y The raster is defined by two scan functions, $s_x(t)$ and $s_y(t)$, which represent the instantaneous beam position. These functions are approximated by sawtooth waveforms (see figure 1). By convention, the horizontal dimension is scanned at a relatively high frequency, called the *line frequency*, f_L, while the slower vertical frequency determines the *frame rate*, f_F.

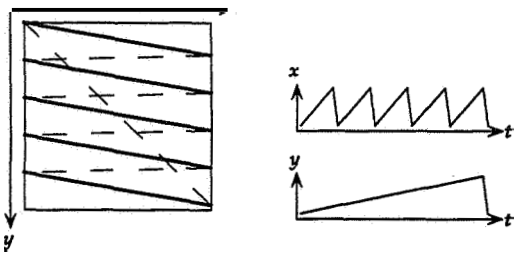


Figure 1 Diagram of raster pattern on the left, with the active portion of each line shown as a heavy solid line, and retrace as a dashed line (see appendix). On the right, the scan functions are shown over time.

Time $t=0$ in our temporal coordinate system is the beginning of the current frame. By convention, numbering of raster lines begins with 1, The index of the current line, i_L, is

$$i_L = \lfloor t\, f_L \rfloor \tag{1}$$

We define t_S to be the time of the start of the current line, and t_L to be the time relative to the start of the current line:

$$t_S = \frac{i_L}{f_L}, \quad \text{and} \quad t_L = t - t_S. \tag{2a,b}$$

These quantities are illustrated graphically in figure 2.

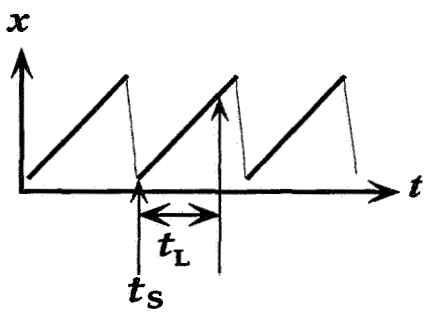


Figure 2: Raster waveform diagram, indicating the current time, t, the start time of the current line, t_S, and the time within the current line, t_L

The *scan velocities*, $v_{S,x}$ and $v_{S,y}$, describe the rate at which the scanning beam traverses the image plane. When expressed in units of image widths per second, these are approximately equal to the scan frequencies, f_L and f_F (see appendix for details). We can write simple expressions for the instantaneous scan position in terms of the scan velocities. The horizontal scan position $s_x(t)$ is:

$$s_x(t) = t_L\, v_{S,x}\,, \tag{3a}$$

In most scanning systems, the vertical scan is continuous (partly due to "mechanical" constraints), and

$$s_y(t) = t\; v_{S,y} \tag{3b}$$

12. Effects of object motion

P	a target point
$p_x(t), p_y(t)$	instantaneous position of P
$\dot{p}_x(t), \dot{p}_y(t)$	instantaneous velocity of P
$\ddot{p}_x(t), \ddot{p}_y(t)$	instantaneous acceleration P
x_0, y_0	position of P at time $t=0$
x_P, y_P	position of P in scanned image
t_P	time P is scanned

We consider a fiducial point on the target located at coordinates (x_0, y_0) at time 0. Let the position at time t be expressed by the functions $p_x(t)$ and $p_y(t)$. These positions can be expressed using Taylor series, where $\dot{p}_x(0)$ is the x velocity at time 0, $\ddot{p}_x(0)$ is the acceleration, and so on:

$$p_x(t) = x_0 + \dot{p}_x(0)\, t + \tfrac{1}{2}\ddot{p}_x(0)\, t^2 + \tag{4a}$$

$$p_y(t) = y_0 + \dot{p}_y(0)\, t + \tfrac{1}{2}\ddot{p}_y(0)\, t^2 + \tag{4b}$$

We wish to know the position of the given point, (x_P, y_P), in the acquired image. When the point's trajectory intersects the raster, the time at which the point is scanned, t_P, will be:

$$t_P = \frac{y_P}{v_{S,y}} + \frac{x_P}{v_{S,x}} \tag{5a}$$

$$\approx \frac{y_P}{v_{S,y}} \tag{5b}$$

By making the approximation, we ignore the dependence on horizontal position. This is justified on the grounds that $v_{S,x}$ is large, and so this term will be small. By definition, $y_P = p_y(t_P)$, and so the value of t_P obtained in equation 5b may be substituted into equation 4b, which can then be solved for yp. The result can then be used to evaluate equation 4a to obtain x_P.

In general, the raster will not pass directly over the point, and features of finite size will often be represented in more than one scan line. We assume that little target motion occurs during a single line time, so the position of a feature located between two scan lines can be accurately determined by interpolation, and results obtained for points lying directly on the raster will hold for all points.

1.3. Example: constant object velocity

$v_{T,x}, v_{T,y}$	target velocity, $p_x(t) = v_{T,x}$

We can use the results of the preceding section to generate simulated distorted images for various motions. We consider first the simple case where the target moves with constant velocity, $(v_{T,x}, v_{T,y})$:

$$p_x(t) = x_0 + v_{T,x} t \ , \tag{6a}$$

$$p_y(t) = y_0 + v_{T,y} t \tag{6b}$$

Following the strategy outlined above, we construct the following equation for y_P:

$$y_P = y_0 + \frac{v_{T,y}\, y_P}{v_{S,y}} \tag{7a}$$

$$= \frac{y_0\, v_{S,y}}{v_{S,y} - v_{T,y}} \ , \tag{7b}$$

$$= y_0 + \frac{v_{T,y}\, y_0}{v_{S,y} - v_{T,y}} \tag{7c}$$

We can use this result to derive a corresponding expression for x_P:

$$x_P = x_0 + \frac{v_{T,x}\, y_0}{v_{S,y} - v_{T,y}} \tag{8}$$

Several important points may be noted from these equations: first, the deviation in feature position for each component is proportional to y_0, the vertical position of the feature in the image, and to the corresponding component of object velocity We also notice that when $v_{T,y} \geq v_{S,y}$ (object moving faster than the raster), the solution corresponds to a negative value of t, and does not correspond to a point in the current frame.

Distortions arising from a target speed of $\frac{v_{S,y}}{4}$ are illustrated in figure 3. The left-hand patch shows the image obtained when a square grid target is moved at the right, while the right-hand patch shows the image resulting from upward motion.

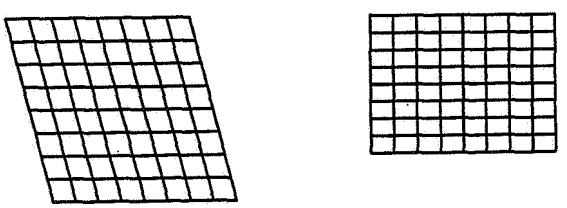


Figure 3: Image distortions of a regular grid for constant velocity motion to the right (left) and upwards (right).

1.4. Example: constant object acceleration

a_x, a_x	target acceleration, $\ddot{p}_x(t) = a_x$

We assume the object accelerates from rest at time 0 with accelerations a_x and a_y

$$p_x(t) = x_0 + \tfrac{1}{2} a_x t^2 \ , \tag{9a}$$

$$p_y(t) = y_0 + \tfrac{1}{2} a_y t^2 \tag{9b}$$

We first consider the case where $a_y = 0$, i.e. a purely horizontal motion. In this case, the vertical position of the fiducial point will not be changed, and the raster will scan the point at time $t_P = \frac{y_0}{v_{S,y}}$ Substituting this value into equatlon 9a, we obtain.

$$p_x(t_P) = x_0 + \frac{a_x\, p_y^2(0)}{2\, v_{S,y}^2} \tag{10}$$

Equation 10 is quite similar to equation 8, except that here the deviation is proportional to the square of the vertical position. This case is illustrated on the left side of figure 4

The case of vertical accelerations is more complex, due to the interaction between the accelerating motlon with the vertical scan. As we did above with equation 7a, we begin by constructing an equation in y_P:

$$y_P = y_0 + \frac{a_y\, y_P^2}{2\, v_{S,y}^2} \ , \tag{11a}$$

which after application of the quadratic formula yields:

$$y_P = \frac{v_{S,y}\ (\ v_{S,y} \pm \gamma\ )}{a_y} \ , \tag{11b}$$

where

$$\gamma = \sqrt{v_{S,y}^2 - 2\, a_y\, y_0} \tag{12}$$

The smaller of two solutions corresponds to the first coincidence of the raster and the point, while the larger only exists when the acceleration is so large that the point subsequently overtakes the raster When the acceleration is so large that the raster never encounters the pomt, γ is imaginary

After more algebra, we can also obtain this result for x_P:

$$x_P = x_0 + \frac{v_{S,y}^2 - a_y\, y_0 \pm v_{S,y}\, \gamma}{a_y} \tag{13}$$

These results are used to compute the images in figure 4. For horizontal acceleration, we see that the vertical grid lines become curved, while for vertical accelerations the grid is compressed nonuniformly

In general we observe that horizontal motions generate horizontal shearing of the image proportional to the instantaneous velocity, and that vertical motion similarly generates vertical compression. Figure 5 illustrates the case of a rapid smooth displacement during the middle of the scan. Mathematical details are omitted in the interest of brevity

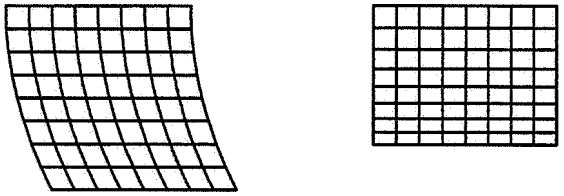


Figure 4 Image distortions resulting from constant acceleration (from rest) of the target to the right (left panel) and upwards (right panel).

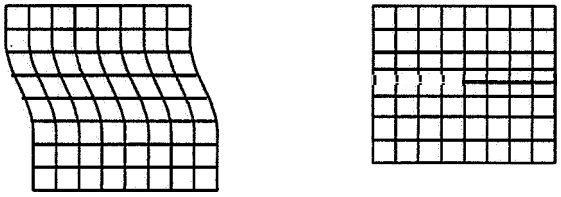


Figure 5: Image distortions resulting from a smoothed step motion to the right (left panel) and upwards (right panel).

2. Methods

2.1. Basic registration methods

The basic registration method, described in detail by Mulligan[1], was developed to process retinal images obtained from a camera-based video ophthalmoscope. It assumes the existence of a *template,* which is a large image of the object. It is assumed that all input images can be registered wholly within the template. Each input image is registered to the template by finding the maximum of the normalized cross correlation. The cross-correlation is computed by taking the product of the Fourier transform of the input and the complex conjugate of the template transform, then taking the inverse Fourier transform. The resulting correlation image is then multiplied by a normalizing image which accounts for the fact that the energy in the template image varies in space. The normalizing image is computed by convolving the pixel-wise squared image of the template with a mask corresponding to the valid input area.

Subpixel interpolation of the correlation maximum is performed by first locating the maximum value in the correlation image, and then performing biquadratic interpolation on the 3 by 3 neighborhood of pixels centered on the maximum. A band-pass filter applied to the input images blurs the the cross-correlation, allowing accurate interpolation. The method has been shown empirically to produce errors less than 0.1 pixel[1,2] The peak value of the correlation (which occurs between the sample points) is interpolated using the parameters of the best-fitting quadratic surfaces, computed using the singular value decomposition.

2.2. Motion parameter estimation

Estimation of the target motion profile is done by computing the distortions of the template image resulting from a set of sample guesses. For each guess, the normalized cross correlation between the input image and the distorted template is computed, and the parameter space is searched to find the motion whch produces the largest normalized correlation C with the input. We can think of the difference $1-C$ as representing an "error" between our guess and the true state of the world, although this may not reach a value of 0 even for the correct motion. The key to a practical solution is minimizing the number of dimensions of the space of possible motions. For example, assuming that the target moved with constant velocity, then there are only two unknown parameters of the warp, $v_{T,x}$ and $v_{T,y}$

A straightforward but expensive approach is to finely sample the parameter space, and compute the cross-correlation for each candidate motion. The peak value of each correlation is stored in an array, indexed by the motion parameter values. This array can then be examined to find the maximum, corresponding to the best-fitting motion parameters. The space can then be resampled on a finer grid, if desired, to obtain a more precise estimate. The feasibility of such an approach largely depends on the error surface character If it is smooth, the initial sampling can be quite coarse and yet still provide a good estimate of the maximum through interpolation.

Figure 6 shows velocity space images of the correlation maxima, for uniform translation of a retinal template (shown ahead in figure 8). It can be seen that, at this resolution, the values decrease monotonically with distance from the true parameter values. In

such a case, we can obtain reasonable estimates at greatly reduced cost by sampling more coarsely

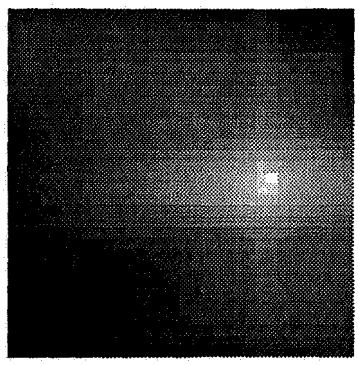


Figure 6: Velocity space image showing the peak value of the normalized cross correlation, evaluated on a 32x32 grid of trial velocities used to warp the template. Input image was warped in accordance with a simulated horizontal motion of 1.25 image widths per frame, sample grid spans the range ±2 image widths per frame. The gray level range has been scaled to span the range of correlation values (0.114 - 0.940).

2.3. Successive refinement

ε_i	ith estimation error
Δv_i	ith parameter sample spacing

After obtaining an initial estimate of the target motion from a coarse sampling of the parameter space (see section 2.2), we may wish to reduce the estimate error by resampling the parameter space more finely in the neighborhood of our current estimate. In the current implementation, all estimates are obtained by using the biquadratic interpolation procedure (developed to localize cross-correlation peaks) to 3x3 sample arrays from the parameter space error surface. The array is sampled at the current estimate (initially 0), and the flanking samples are separated by Δv. After the initial estimate is obtained, a new 3x3 sampling grid is placed at the location of the current estimate. The sample spacing of the new grid is equal to the previous sample spacing, times a fraction β, which can be thought of as the reciprocal of the "zoom" factor. If beta is large, i.e. close to 1, the error in the new estimate may not be significantly reduced. If it is too small, however, then the true maximum will lie outside of the region spanned by the sample array. A one-dimensional example with a value of β=0.5 is shown in figure 7

Experimentation has shown that the interpolation procedure can be unstable when the maximum sample value occurs at the array edge; the samples can be fit with a hyperbolic surface, or an ellipsoidal surface which is concave up. Therefore, if the maximum value is not obtained at the center sample, the array is shifted by one sample until the maximum does occur at the center. In two parameter estimation, each lateral or vertical shift requires the computation of 3 additional correlations, while diagonal shifts require 5. (This problem does not arise when interpolating cross-correlation images because the entire correlation image is available: we can perform an exhaustive search for the maximum sample value, and insure that it falls at the center of the interpolation array.)

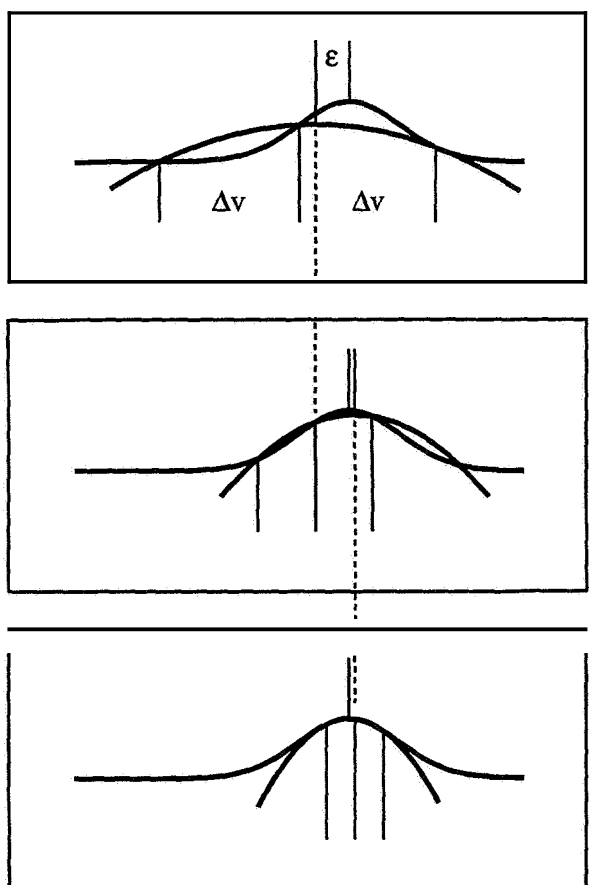


Figure 7: Estimation of the extremal position by quadratic interpolation from 3 samples, illustrated in one dimension. The true underlying error function is a Gaussian. In the initial estimate (top), the quadratic fit is poor, and localization error is significant. This initial estimate is used as the center of a new array of samples, at half the spacing (center). The error is reduced, although the fit is still poor. In the third iteration (bottom), the fit is good and the error is small.

Ideally, we would like for the new sample spacing, Δv_{i+1}, to be approximately equal to the estimation error on the previous iteration, ε_i. While we do not in general have *a priori* information about the nature of ε as a function of Δv, when the error function is smooth the quadratic fit will improve as Δv decreases, and ε will approach zero, or more precisely some small value determined by the numerical precision of the machine and the noise level in the input images. For a particular template image, the dependence of ε on Δv may be precomputed for a set of representative velocities, and the results used to construct a table of β's.

When there are no local extrema on the error surface, this procedure works well. This is an unrealistic assumption, however, because numerical imprecision, input noise, and the structure of the template autoconelation all add complexity to the error surface. What is needed is the analog of an antialiasing filter in the parameter space domain, Remember that when a signal is sampled at frequency f_s, the signal must first be prefiltered to remove frequencies above the *Nyquist frequency,* $f_s/2$. Failure to do so causes super-Nyquist frequencies to .appear as low-frequency "aliases," which cannot be distinguished from the actual low frequencies in the input.

When we sample the parameter space at some spacing Δv, we would like to insure that there are no bumps and wiggles occurring between our sample points which will corrupt the interpolation process and cause the procedure to become trapped in a local extremum. The basic idea is similar to the "coarse-to-fine" approach, proposed for stereo correspondence[3] and motion estimation[4] In those cases, blurring is performed in the image domain, but here what we would like here is a template transformation which will result in blurring of the error surface. Averaging together the distorted templates corresponding to a neighborhood of the parameter space is not exactly correct because of the (nonlinear) maximum operation that we perform in the construction of the error surface, but something similar may produce a useful result. Rucklidge has described a related approach[5] which is immune to some of the problems of coarse-to-fine strategies, while offering similar computational savings.

2.4. Exploiting features of the cross-correlation

While the successive refinement described in the previous section performs well while computing many fewer correlations than would be required by exhaustive search at the smallest sample spacing, it is still fairly computationally expensive. To reduce the amount of computation required, an intriguing possibility is suggested by the appearance of the correlation images. Figure 8 (top row) shows a portion of a retinal template which we use as our test image, together with a distorted version produced by a simulated eye movement. The third row of figure 8 shows the autocorrelation of the template on the left, together with the cross correlation of the template with the distorted version on the right.

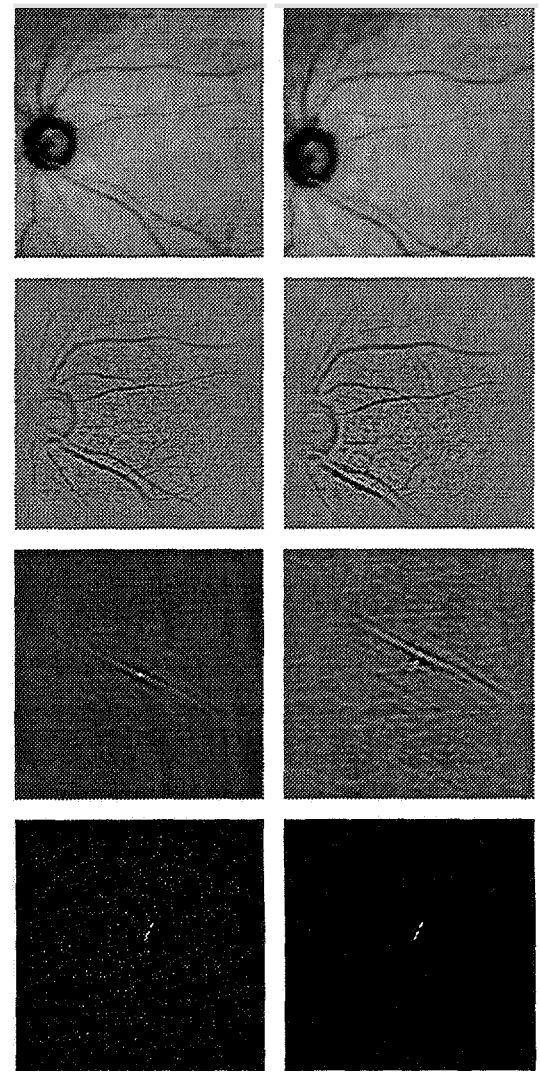


Figure 8: Top row a portion of a retinal image template (left), together with a version distorted by a simulated eye movement of constant velocity. Second row: the images from the first row are band-pass filtered to accentuate the retinal blood vessels, and windowed to reduce edge artifacts. Third row the autocorrelation of the template (left), and the cross correlation of the template and the input. Because the gray level range has been matched to the image extreme values, more detail is seen in the cross-correlation image on the right, which has a lower peak value. The long diagonal streak is related to the prominent vessel to the lower right of the optic disk. Bottom row the cross-correlation is deconvolved with the template autocorrelation (left). The sharp diagonal feature is related to the motion trajectory, and may be identified by thresholding (right).

In the cross-correlation image, the central peak is smeared out in the direction of motion. To see why this should be so, recall that when scanning a moving object, the position in the top of the frame is different from the position in the bottom of the frame. The peak in the cross-correlation image indicates the relative location of the object; as the object moves during the scan, this peak shifts accordingly When the autocorrelation of a template image patch is roughly independent of the its location (e.g. high-pass noise), the cross correlation image can be thought of as the convolution of the autocorrelation with the motion trajectory For less uniform images, the trajectory may show gaps, corresponding to portions of the template in which there is little or no energy

We can attempt to solve for the trajectory by deconvolving the cross-correlation image with autocorrelation of the template; the result of this operation is shown in the bottom row of figure 8. The bottom row of figure 8 also shows the result of clipping the deconvolved image from below at 0, rescaling the positive values to the range 0-1, and thresholding at a value of 0.4. A simple machine vision algorithm mght look for a line segment in this image, and obtain a rapid estimate of $v_{T,x}$ and $v_{T,y}$ from the location of the endpoint.

3. Application to SLO images

The registration methods described were developed to track eye movements using images obtained with a camera-based video ophthalmoscope[1], and were subsequently applied to images obtained from a scanning-laser ophthalmoscope (SLO)[6] The full-field correlation method differs from previous efforts to obtain eye position data from SLO images[7], which relied upon the identification and localization of small discrete features, such as vessel bifurcations. Stetter *et al.*[8] computed high temporal resolution profiles of horizontal saccades from the profiles of a few major blood vessels oriented roughly vertically in the image, exploiting the shear distortion produced by the interaction between the vertical scan and horizontal eye movements. In this section we will begin by considering the problems encountered applying full-field correlation techniques to these images, and discuss the use of image distortions to recover complete two dimensional motion trajectories.

3.1. Template construction

In the preceding section, we assumed the existence of a template Image, e.g. a large image of the target object constructed as a mosaic of a large number of input images. Various techniques have been proposed to automate the construction of image mosaics from sets of fundus images[9-11], but these have generally been more concerned with visualization of *gross* anatomical features than the preservation of metric structure. Here, templates are constructed from an input sequence using a bootstrap procedure. All images are first band-pass filtered; the high-pass component of the filter serves to accentuate the retinal blood vessels, and remove low-frequency artifacts due to non-uniform illumination of the retina, while the low-pass component attenuates high-frequency camera noise. These images are optionally windowed with a Gaussian-blurred rectangle slightly smaller than the input to reduce edge artifacts. The first preprocessed image is used as the initial template, and the second frame is registered with respect to it. The computed displacement is then applied to the second frame to form a scrolled copy. Sub-pixel translations are performed by appropriate phase shifts[12] in the Fourier domain. The shifted image is then summed into an accumulation buffer, while a similarly shifted mask of 1's is summed into a pixel count buffer The current template is a weighted average of all of the previously registered frames, computed as the pixel-wise quotient of the accumulation and count images. Frames whose normalized correlation with the template is below a threshold (typically around 0.5) are excluded. A representative template image is shown in figure 9.

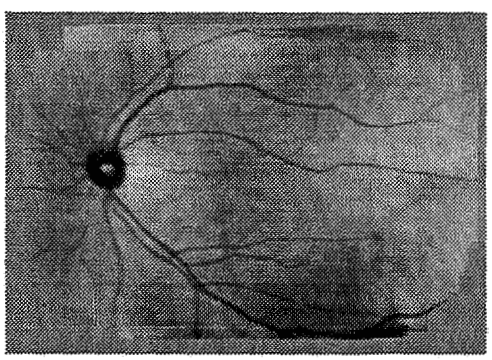


Figure 9: Registration template constructed from 120 SLO field images.

Once a template has been constructed, the registration quality can be assessed by creating a video sequence in which the input images are overlaid on the template, after having been translated to provide the best match to the template. When this is done well, what is seen is a stationary template pattern, with a smaller noisy window moving over it. Registration errors are manifested as motions of the template structure within the moving window, while the position of the window is related to the computed direction of gaze. A static version of this is presented in figure 10, where a single nominally registered input image is superimposed over the template shown in figure 9.

Figure 10 Template image shown in figure 9 overlaid with a single input image. Misregistration can be observed at the upper border, while registration at the **right** and lower edges appears to be good.

When the quality of registration is visualized as described above, non-rigidity of the template structure is often observed, indicating the occurrence of registration errors. There are two potential sources of these errors. The first is distortion introduced by a moving eye (the topic of this paper). The second results from template construction errors. Stetter ***et al.*** measured the static **SLO** distortion by holding a sheet of graph paper in front of the instrument and observing the image transformation resulting from the **SLO** optics. They used the measured distortion to correct the image feature locations used in their method.

If not corrected for, distortions such as that observed by Stetter ***et al.*** will corrupt the template construction process, because globally correct matches will not be obtained. The resulting template distortion depends not only on the inherent imaging system distortion, but also on the positions which are sampled and the order in whch they used to compute the template. Another potential source of distortion results from the fact that the retina is a spherical surface, which is projected to a plane by the imaging system. This distortion is negligible for excursions of less than 4 degrees[13], but must be incorporated into the registration process for accurate construction of large templates. Error-free templates are critical for obtaining eye movement estimates which are limited only by the noise in the images.

3.2. Parametric models of eye movements

Before attempting to estimate high temporal resolution eye movements from **SLO** images, it is useful to first consider what is known about eye movements (an excellent survey is provided by Kowler[14]). The highest velocities reached by the eye are during the execution of ***saccadic*** eye movements, which will therefore be of most relevance to the present work. Saccades are rapid, ballistic changes in fixation, which can be superimposed over other smooth movements. Small saccades which occur during "steady" fixation are sometimes referred to as ***microsaccades***. Lawful relations are observed between saccadic amplitude, duration, and velocity[15-18], so that saccades in a given direction can be described by a family of functions. A good analytic description of these functions is provided by the density function of the gamma distribution[19], whch has three parameters which are related to duration, peak velocity, and skewness or asymmetry. (The situation is more complicated for oblique saccades, where asynchronies between different muscle group activations can produce "looping" saccades.)

Because the largest saccadic velocities are well below the scan velocities of the **SLO,** reasonable template registration is obtained even during saccades. Thus, the positions sampled at the frame rate (obtained by correlating the uncorrected images with the template) will allow us to identify the occurrence of large saccades, and for large saccades whose durations spans several frames we can make good predictions about average velocity and acceleration within each frame, speeding the initial portion of the parameter search. Microsaccades are too brief to be detected in this way, but may be revealed by dips in the value of the correlation of the raw input with the template.

3.3. Ocular torsion

In addition to rotating about horizontal and vertical axes to redirect gaze, the eye is also capable of rolling around the line of sight, a movement which is known as ***ocular torsion.*** There is a systematic variation of torsion with deviation of gaze[20], as well as small fluctuations during fixation[21] The occurrence (and measurement) of torsional eye movements complicates the registration of fundus mages, although a search procedure similar to that described above for finding image warps has been able to recover simulated torsions with an accuracy of around 0.1 degree[1], whch is comparable to results obtained from video images of the iris[22-24] Ott and Eckmiller[25] used **SLO** images to measure ocular torsion during smooth pursuit by measuring the change in the slopes of lines connecting image features, They do not mention the existence and correction of the scan induced shear in the moving target image, which is significantly smaller than the measured effects, but significantly larger than the attainable measurement precision.

Fortunately, torsional eye movements are relatively slow, so that there is very little rotation during the scan of a single frame. Furthermore, this slowness means that the torsional state of the eye can be well predicted from the state in the preceding frames. Therefore, the only distortions which have been con-

sidered in this paper are those resulting from pure translation of the entire target, although the methods are completely general.

3.4. Superresolution

When the target contains frequencies above the Nyquist limit of half the sampling rate which are passed by the point spread function of the imaging system, ***aliasing*** will occur, This term refers to the fact that samples from a signal above the Nyquist frequency are indistinguishable from samples from a corresponding signal composed entirely of frequencies below the Nyquist frequency This is true for single images, but not strictly so when we consider multiple images with slight positional offsets. When the target moves by a small amount, all of the sub-Nyquist frequencies in the image move by the same amount. Aliases of super-Nyquist frequencies, on the other hand, have a different motion, like that of a Moire pattern. Reconstructing the super-Nyquist frequencies from multiple samples of the aliases is known as ***superresolution***[26-28] (the term has also appeared in the optics literature to describe physical situations in which the resolution is increased by a factor of *2* or $\sqrt{2}$ above the nominal diffraction limit[29]).

Can we apply superresolution techniques to SLO images? In other words, are there any aliases of retinal image frequencies above the Nyquist limits imposed by the SLO raster? The following anecdotal observation suggests that the answer is yes: Stevenson[30] has noted that the individual raster lines of the SLO can be resolved by the eye (of the subject). This means that the spot size of the illumination beam is smaller than the line spacing. The retinal image is effectively low pass filtered by a filter whose kernel is the spot profile of the laser beam convolved with the eye's optical point spread function. This low-pass filter functions as an effective anti-aliasing filter only when its width is comparable to the line spacing. When the raster lines are clearly resolved, two images whose vertical position differs by half the line spacing can be viewed as interlaced components of an image with twice the vertical resolution, and by analogy a sequence of slightly-misregistered images can be combined to produce a high-resolution image, provided the input images can be accurately registered.

3.5. Application to laser surgery

There is a current need for real-time retinal tracking and stabilization in the medical community Laser surgery consists of delivering a brief, high energy pulse to a small retinal area. The target area is selected by the physician, who presses a button indicating that an aiming cursor has been placed over the target area in an ophthalmoscopic retinal image. There is a small, but finite, time window (a few hundred milliseconds) between the physician's decision that the cursor is placed, and the actual delivery of the laser pulse. During this interval, the patient could make an eye movement causing the pulse to be delivered to the wrong part of retina, possibly injuring the fovea or optic nerve head! A safer system would would track eye movements and permit the physician to indicate the target location on a stabilized image, while stabilization optics keep the laser accurately targeted in the presence of patient eye movements.

A nearly real-time stabilizer might be constructed by registering individual SLO raster lines to the template as they become available from the instrument, eliminating the 16 msec pipeline delay needed to buffer an entire video field. Because the video line rate of 15 kHz is much higher than the bandwidth of natural eye movements, it should be possible to make accurate predictions by correlating each line with just a few lines of the template in the neighborhood of the expected position. Computing a small number of 1-D correlations between selected lines is considerably less computation than the complete 2-D cross correlation, and is likely to be tractable with modem digital signal processing circuits. The problem is complicated somewhat if one wishes to measure ocular torsion, although this may not be necessary to produce a useful surgical stabilizer

4. Conclusions

While more work is needed to refine motion parameter estimation to extract *high* time resolution eye movement data from SLO image sequences, these results suggest that the technique is practical. Increased temporal resolution, combined with the high precision made possible by sub-pixel interpolation, make this approach competitive with other high-resolution techniques, such as the more invasive search coil[31]

5. Acknowledgements

Special thanks to Tina Beard for comments on early drafts of the manuscript and help with the figures. This work was supported by NASA grant 199-06-12-39, and NASA Aeronautics RTOP's 505-64-53 and 548-60-12. SLO imagery was kindly provided by David Grosof, Washington University and Central Institute for the Deaf, whose work received support from the J. Epstein Foundation, Jewish Hospital Research Fund, St. Louis MO, and an unrestricted grant from Research to Prevent Blindness to Washington University Ophthalmology

6. Appendix: Minutiae of *raster* scanning

τ_A	active time
τ_R	retrace duration
α_L	line scan duty cycle, $\tau_{AL} f_L$
α_F	frame scan duty cycle, $\tau_{AF} f_F$

In our discussion of raster scanning, we ignored the fact that raster retrace requires a finite duration, and stated that the scan velocities (in image widths per second) were approximately equal to the scan frequencies in Hz. In this appendix we will present a more precise formula for the scan velocities, and discuss some additional issues concerning raster geometry

Typically, image data are acquired, or displayed, only during a portion of the scan periods, called the *active time,* of duration τ_A, the remainder of each scan period is spent returning the beam to the starting position, known as *retrace,* of duration τ_R. These quantities are shown with respect to a typical scan waveform in figure 11 In the raster diagram in figure 1 (section 11), the active portions of the scan lines are shown as solid lines, while the retrace portions are shown as dashed lines. The *duty cycle* α of the active segment for each scan is defined as the product of the active time and corresponding scan frequency

$$\alpha_F = \tau_{AF} f_F \ , \quad \text{and} \quad \alpha_L = \tau_{AL} f_L \qquad (14a,b)$$

The *scan velocities,* $v_{S,x}$ and $v_{S,y}$, are equal to the corresponding scan frequency divided by the corresponding duty cycle times the image width (equal to 1 due to our choice of units):

$$v_{S,x} = \frac{f_L}{\alpha_L} \ , \quad v_{S,y} = \frac{f_F}{\alpha_F} \qquad (15a,b)$$

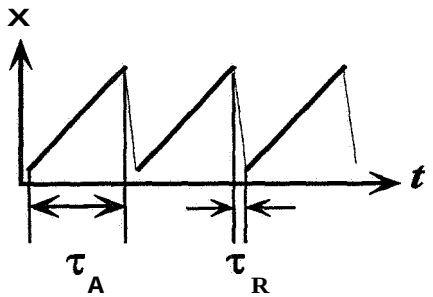


Figure 11 Diagram of raster waveform showing active time τ_A and retrace time τ_R.

In the television standard defined by the National Television Standards Committee (NTSC)[32,33], the scan frequencies have a value of $f_F \simeq 60$ Hz, and $f_L = 262.5\, f_F$. The significance of the horizontal frequency being a half-integral multiple of the frame rate is that two successive "frames" are vertically offset by half the line spacing, known as *vertical interlace (see* figure 12). These two half-frames are referred to as *odd* and *even fields,* determined by the parity of the line numbers. Treating fields as "frames" is equivalent to superimposing vertical square wave motion on the target at half the field rate, with amplitude of 0.5 line, and so may easily be corrected *post* hoc.

The raster shown in figure 1 depicts a continuous vertical scan, resulting in the image samples being taken from slightly oblique lines. Images manipulated in a digital frame buffer, however, are often assumed to be samples obtained from a rectangular sampling grid, corresponding to a discontinuous vertical scan (see figure 13). In this case,

$$s_y(t) = t_S \, v_{S,y} \qquad (16)$$

The small errors introduced by this approximation vanish as $f_L \gg f_F$; when necessary they may be corrected by resampling an appropriately sheared version of the image.

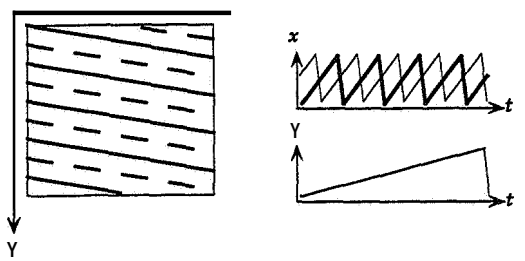


Figure 12: Raster diagram similar to figure 1, showing an interlaced scan.

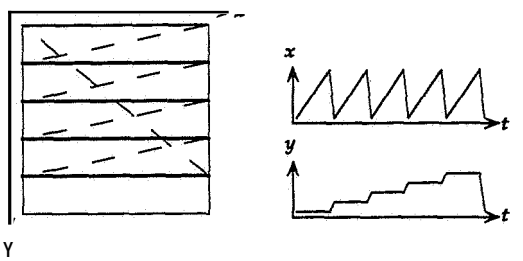


Figure 13: Diagram showing a rectangular raster approximation, and the associated waveforms.

References

1 Mulligan, J. B., "Image processing for improved eye tracking accuracy," *Behav. Res. Methods Instrum. Comput.,* vol. **29,** pp. **54-65, 1997**

2. Mulligan, J. B. and Beutter, B. R., "Eye-movement tracking using compressed video images," in ***Vision Science and its Applications,*** vol. **1,** pp. **163-166, 1995** OSA Tech. Dig. Series, Optical Society of America, Washington, D.C., **1995.**

3 Marr, D. and Poggio, T., "A computational theory of human stereo vision," ***Proc, Roy. Soc. Lond. B,*** vol. **204,** pp. **301-328, 1979.**

4. Anandan, P., "A computational framework and an algorithm for the measurement of visual motion," ***International Journal of Computer Vision,*** vol. **2,** pp. **283-310, 1989.**

5. Rucklidge, W., "Efficient guaranteed search for gray-level patterns," in ***Proc. CVPR,*** pp. **717-723,** IEEE Computer Society Press, Los Alamitos, **CA, 1997**

6. Webb, R. H., Hughes, ***G. W.,*** and Delori, F C., "Confocal scanning laser ophthalmoscope," ***Applied Optics,*** vol. **26,** pp. **1492-1499, 1987**

7 Wornson, D. P., Hughes, G. W., and Webb, R. H., "Fundus tracking with the scanning laser ophthalmoscope," ***Applied Optics,*** vol. **26,** pp. **1500-1504, 1987**

8. Stetter, M., Sendtner, R. A., and Timberlake, ***G.*** T., **"A** novel method for measuring saccade profiles using the scanning laser ophthalmoscope," ***Vision Research,*** vol. **36,** pp. **1987-1994, 1996.**

9 Peli, E., Augliere, R. A,, and Timberlake, G. T., "Feature-based registration of retinal images," ***IEEE Transactions on Medical Imaging,*** vol. **MI-6,** pp. **272-278, 1987**

10. Cideciyan, A. V., "Registration of ocular fundus images," ***ZEEE Engineering in Medicine and Biology,*** pp. **52-58, 1995.**

11. Mahurkar, **A.** A., Vivino, M. **A.,** Trus, B. L., Kuehl, E. M., III, M. B. Datiles, and Kaiser-Kupfer, M I., "Constructing retinal fundus photomontages," ***Investigative Ophthalmology & Visual Science,*** vol. **37,** pp. **1675-1683, 1996.**

12. Bracewell, **R.** R., ***The Fourier transform and its applications,*** McGraw-Hill, New York, **1965.**

13. Ott, D. and Daunicht, **W** J., "Eye movement measurement with the scanning laser ophthalmoscope," ***Clinical Vision Science,*** vol. **7,** pp. **551-556, 1992.**

14. Kowler, E., "The role of visual and cognitive processes in the control of eye movement," in ***Eye movements and their role in visual and cognitive processes,*** ed. E. Kowler, pp. **1-70,** Elsevier Science, Amsterdam, **1990.**

15. Ditchburn, R. W., ***Eye-movements and visual perception,*** Clarendon Press, Oxford, **1973.**

16. Boghen, D., Troost, B. T., Daroff, R. B., Dell'Osso, L. F., and Birkett, J E., "Velocity characteristics of normal human saccades," ***Investigative Ophthalmology,*** vol. **13,** pp. **619-622, 1974.**

17 Bahill, A. T., Clark, M. R., and Stark, L., "The main sequence, a tool for studying human eye movements," ***Mathematical Biosciences,*** vol. **24,** pp. **191-204, 1975.**

18. Bahill, **A.** T and Stark, L., "The trajectories of saccadic eye movements," ***Scientific American,*** vol. **240,** no, **1,** pp. **108-117, 1979**

19. Opstal, A. **J.** Van and Gisbergen, J A. M., "Skewness of saccadic velocity profiles: **A** unifying parameter for normal and slow saccades," ***Vision Research,*** vol. **27,** pp. **731-745, 1987**

20. Nakayama, K., "A new method of determining the primary position of the eye using Listing's law," ***American Journal of Optometry and Physiological Optics,*** vol. **55,** pp. **331-336, 1978**

21 Rijn, L. J. van, Steen, J. van der, and Collewijn, H., "Instability of ocular torsion during fixation: cyclovergence is more stable than cycloversion," ***Vision Research,*** vol **34,,** pp. **1077-1087, 1994.**

22. Curthoys, **I.** S., Moore, **S.** T., McCoy, **S** G., Halmagyi, G. M., Markham, C. H., Diamond, **S.** G., Wade, **S.** W., and Smith, **S.** T., "VTM -- a new method of measuring ocular torsion using image-processing techniques," ***Annuls of the New York Academy of Sciences,*** vol. **656,,** pp. **826-828, 1992.**

23. Bos, J. E. and *Graaf,* B de, "Ocular torsion quantification with video images," ***IEEE Transactions on Biomedical Engineering,*** vol. **41,** pp. **351-357, 1994.**

24. Groen, E., Nacken, P F M., Bos, J E., and Graaf, B. de, "Determination of ocular torsion by means of automatic pattern recognition," ***IEEE Transactions on Biomedical Engineering,*** vol. **43,** pp. **471-479, 1996.**

25. Ott, D. and Eckmiller, R., "Ocular torsion measured by TV- and scanning laser ophthalmoscopy during horizontal pursuit in humans and monkeys," ***Investigative Ophthalmology & Visual Science,*** vol. **30,** pp. **2512-2520, 1989.**

26. Irani, M. and Peleg, S., "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, pp. 231-239, 1991.

27 Mann, S. and Picard, R., "Virtual bellows: constructing high quality stills from video," *Proc. ZCZP*, Austin, Texas, 1994.

28. Cheeseman, P., Kanefsky, B., Kraft, R., Stutz, J., and Hanson, R., "Super-resolved surface reconstruction from multiple images,'' in *Maximum Entropy and Bayesian Methods*, ed. G. R. Heidbreder, pp. 293-308, Kluwer, the Netherlands, 1996..

29. Cox, I. J. and Sheppard, C. J. R., "Information capacity and resolution in an optical system," *Journal of the Optical Society of America A*, vol. 3, pp. 1152-1158, 1986.

30. Stevenson, S. B., personal communication, 1995.

31. Robinson, D. A., "A method of measuring eye movement using a scleral search coil in a magnetic field," *ZEEE Transactions on Biomedical Electronics*, vol. BME-10, pp. 137-145, 1963.

32, Lougbren, A. V., "Recommendations of the National Television System Committee for a color television signal,'' *J. Soc. Motion Picture and Television Eng.*, vol. 60, pp. 321-336, 596, 1953

33. Poynton, C. A., *A technical introduction to digital video*, John Wiley and Sons, New York, NY, 1996.

ASSESSMENT OF NEUROLOGICAL FUNCTION THROUGH THE MULTIDIMENSIONAL INTEGRATION OF INVASIVE AND NON-INVASIVE MODALITIES OR SENSORS

Luc Bidaut† Nicolas de Tribolet‡. Theodor Landis□ Jean-Raoul Scherrer¥ François Terrier"
†Laboratory of Functional and Multidimensional Imaging - ¥Medical Informatics Division - *Radiology Department
‡Neurosurgery □Neurology.
Geneva University Hospitals. 24 rue Micheli-du-Crest, CH-1211 Geneva 14, Switzerland

Abstract: All digital information available at various stages of neurological investigation has been totally integrated through the implementation of IMIPS (Integrated Multidimensional Imaging and Processing System) in a common multisensor space which can be navigated at will All modern modalities (Computed Tomography Magnetic Resonance Imaging. Single Photon Emission Computed Tomography Positron Emission Tomography functional MRI, and also Electro-Encephalo-Graphy and Electro-Magnetic Tomography) can be combined through our system for the investigation of normal or abnormal brain morphology and function Although the potentral is obviously tremendous for the study of normal brain, we initially targeted the pre-surgical evaluation of drug-resistant epilepsy and the planning or monitoring of subsequent surgery Compared to the studying of nonnal function, the complete handling of this pathology drove us to also incorporate in our approach invasive electro-physiological investigation means (such as depth electrodes and strips or grids) and also slides taken during surgery

Introduction:

Modem neurological investigation techniques span a wide area of methods, equipment and disciplines. Initially. neurology has mostly been handled through neuro-psychological testing of senses and cognitive functions. Besides interview-like tests and EEG. modern and commercially available medical imaging modalities now allow us to have a direct look at in-vivo morphology (Computed Tomography Magnetic Resonance Imaging), metabolism (Single Photon Emission Computed Tomography Positron Emission Tomography) and function (PET functional MRI) While none of these techniques individually provides enough information for a complete look at the brain's function, they all provide information of a complementary nature. In such a context, and for such reasons, it is then natural to merge the broadest scope of information at hand in order to get the best composite picture **of** what is going on in the brain under normal or dys-functional behavior This goal is what we targeted for research but also for clinical applications. Although mapping of brain function or complete objective (i.e quantitative) assessment of any brain disorder is still a challenge in itself we planned to initially address the needs of the pre-surgical evaluation of pharmaco-resistant epilepsy. and of the subsequent planning and monitoring **of** surgery This choice allowed us to focus on a true clinical application and to incorporate invasive investigation (depth electrodes, strips, grids, intervention pictures), without really simplifying the requirements or complexity of our implementation.

Material and Methods:

For understandable ethical reasons, normal brain function is studied - in humans at least - through non invasive techniques such as neuropsychology. EEG/EMT (and also Magneto-Encephalo-Graphy), fMRI and PET

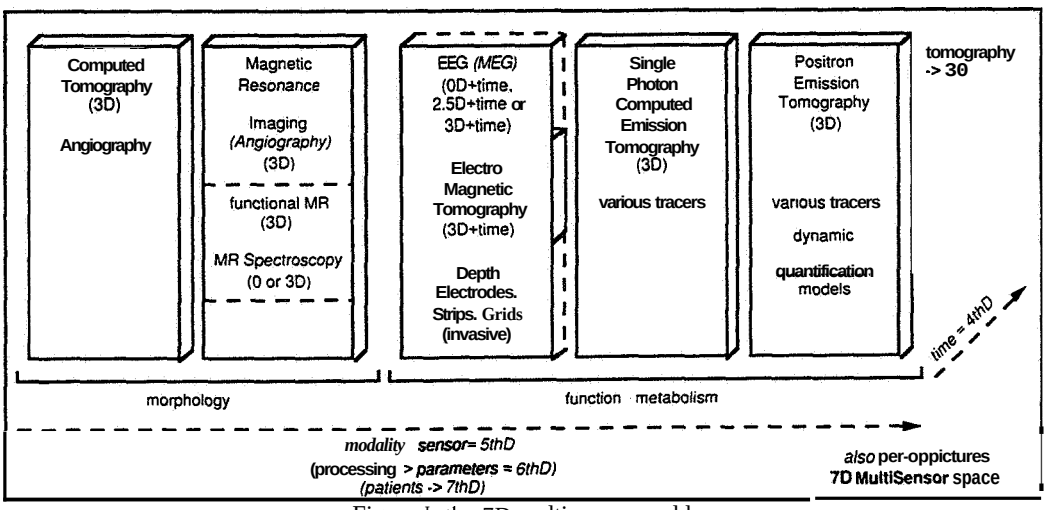


Figure 1 the 7D multisensor world

Modern protocols for the assessment of pathology and epilepsy are quite complex by the number of modalities they currently involve [1 2]. initially. an EEG is performed and because of its time sampling capacities - this is where the abnormal electro-physiologic behavior is best exhibited through direct reading or frequency analysis. Recently. 3D reconstruction EMT techniques have been developed [3] which allow for the reconstruction of current generator equivalents in 3D and through time in ordcr to nicely complement the dircct scalp measurements. **As** the spatial resolution of such techniques is still quite limited. we need to rely

on other modalities in order to localize the most likely epileptic focus(-ci). MRI datasets exhibit an cxccllcnt soft tissue contrast and may show lesion(s) Interictal PET (with Fluoro-Deoxy-Glucose a sugar analog which uptake is linked to energy consumption) should show an hypometabolism at the focus' location [1]. The comparison between intcrictal and ictal SPECT with Tc99m-Ethyl Cysteinate Dimer (ECD, a perfusion tracer) should show an ictal increase of tissue perfusion at the location of the focus [2]. This information does not often appear all together but the combining of data - and of the more classical neuropsychological investigation generally permits the most likely focus to be pointed at with enough accuracy for subsequent surgery If not and if there **is** still enough ground to refine the investigation. a more direct look at the physiology of the brain can be taken through depth electrodes or strips/grids which are placed directly in contact with the cortex or the brain structures under scrutiny

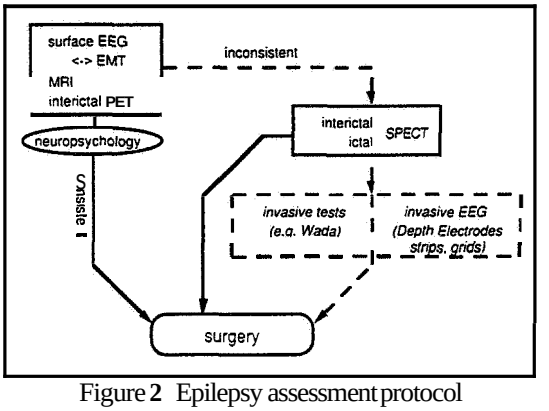


Figure 2 Epilepsy assessment protocol

Among all the modalities previously mentioned, only MRI and CT provide the good anatomical clues which are mandatory to any proper planning or monitoring of surgery When the epileptic focus is not a clear focal lesion identified on the morphological data sets. the intervention will have to be guided by the results from the other modalities. Most metabolic or functional modalities are of low spatial resolution and/or with a poor anatomical content, and this is why the need to register them to modalities with better anatomical contrast (e.g. CT or MRI) was initially clearly identified in order to locate the structures underlying a specific metabolic or functional behavior In our approach, we go beyond the simple localization aspects because the whole multimodality volume can be navigated along any dimension (the 3 Cartesian ones. time. and also the axis of modalities and parameters) for diagnosis but also for planning or monitoring of therapy Our developments led us to integrate all aspects in a virtual hybrid system which we dubbed IMIPS (Integrated Multidimensional Imaging and Processing System, [4] and figure 3) which was implemented on top of a commercial multiplatform module-based development software (Explorer SGI - Numerical Algorithm Group, Oxford. UK) and which provides interfacing with many different data formats. visualization paradigms and systems in order to accommodate its many applications.

The major entities our system is based upon arc data retrieval and interfacing. processing. registration. visualization and navigation. All data from our imaging or non-imaging equipment arc first retricvcd and converted bcforc being registered through manual. point. surface. or voxel similarity based engines to any reference space (e.g. the anatomical MRI) and retrofitted to our PACS (Picture Archiving & Communication System)[5] for review or processing on any of its workstations.

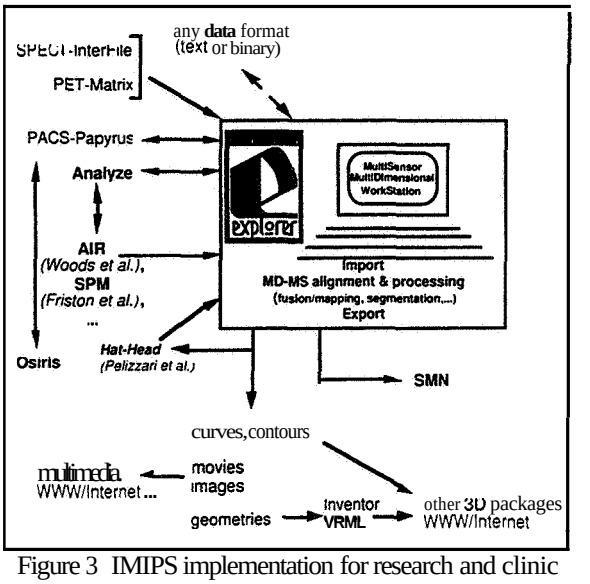


Figure 3 IMIPS implementation for research and clinic

For brain study processing can range from segmentation of specific structures to pharmaco-kinetic modeling of dynamic data sets and statistical processing of sequences of functional acquisitions. While the latter type can rely on largely distributed. inherently validated and continuously improved Statistical Parametric Mapping [6], the former processes were mostly locally developed. Segmentation of medical images generally ranges from direct editing/tagging of the data to morphology based or statistical operators. Whole brain segmentation is now best performed through morpho-topological techniques which isolate the biggest connected object in a head scene [7]. Pharmaco-kinetic modeling is performed for PET by fitting the content or exchange parameters of a compartmental model to experimental datasets (blood/plasma curve and tracer's uptake in the tissues) [8]. Many additional cross-modality processing and refinements are always being implemented or developed in order to improve specific aspects and also move them further in the clinical world by tuning them to specific protocols.

As further discussed in [4] and many other references, registration of modalities which exhibit various characteristics can only be achieved through a hybrid combination of various techniques, the scope of which covers a wide area. simple inanual techniques [9], point based registration [10], registration of various graphical primitives [11 surface based techniques [12 13], voxel similarity ones [14]. Besides the primitives they use. the pre-processing steps they require. and their level of integration all these techniques are also characterized by the calculation of their alignment cost which provides a value to rank the quality of rhe alignment for a given inter-sct geometrical transformation Although most techniques implement simple linear transformations. which are often enough to register datasets from the same patient. others can also include non linear warping which allows for the registration of datasets across patients or with a generic atlas. Automated iterative cost minimization is often performed by modifying the transformation parainetcrs until the cost reaches a given threshold. In the common linear case. the measurable accuracy of most techniques is on the order of one voxel or better While implementation and iterative engine can differ for a specific cost function. it is important to remark that the techniques can also be divided by the primitives they rely on for their cost calculations (images. points. densities. other features, etc). Not unlike the multimodality rationale. this is the reason

why. while none of the techniques is general cnoiigh to solve all registration problems. hcing able to use any of thcin (at least one from every family) allows for the solving of most.

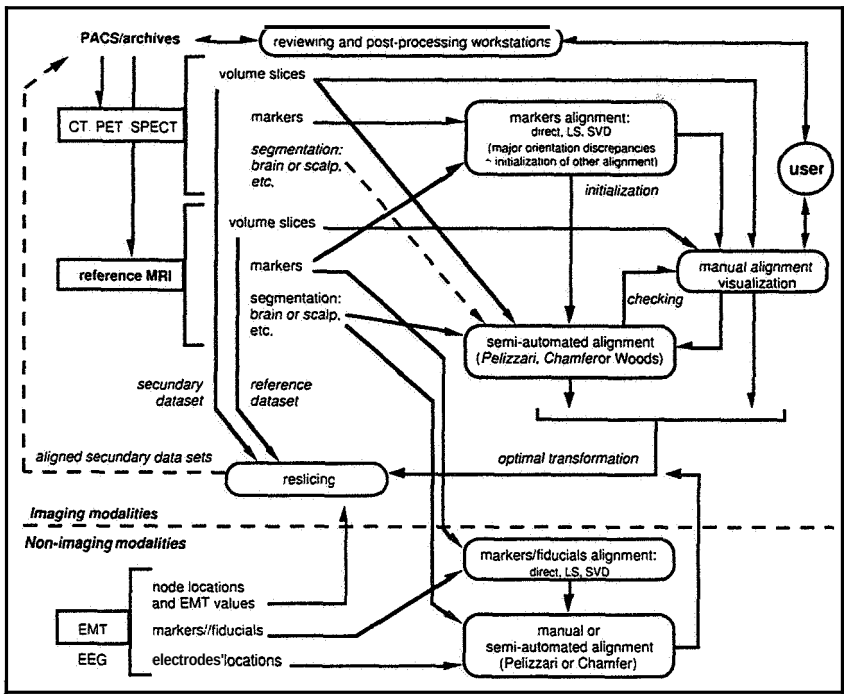


Figure 4 Multisensor registration within **IMIPS**

Because of the extra informahon it may provide to the all-digital world. another registration approach was also identified as the need to register 2D pictures taken during the intervention with either pre- or post-surgical digital volumes. In this case. 2D images of the **3D** volumes are generated with imaging parameters similar to the ones the 2D pictures were taken with (orientation. perspective). Then similar points/landmarks are identified on both datasets (for example electrodes, fiducials or anatomical landmarks) and they are registered to each other with a 2nd or 3rd order global polynomial fit which warps one **2D** image to the other one [15].

Once the datasets have been registered and/or segmented, visualization and navigation can take place in various modes which range from the simple 2D display of slices selected from the $3D^{+}$ volume to the visualization in **3D** of $3D^{+}$ entities (points, contours, surfaces, etc) extracted from the $3D^{+}$ dataset **As** visualization can also be made dependent on the specific entity under scrutiny data from one initial datasets may be projected onto entities extracted from another for example, functional modalities can be projected on morphological surfaces or planes. etc Figures *6* to 8 show examples of such visualization techniques which. as they take place in **3D**, also allow for the navigation within and around the space encompassed by the data. e.g. in parallel to anatomical features or along neurosurgery trajectories. Movies or VRML (Virtual Reality Modeling Language [16]) objects can also be generated for decreasing the dataset storage requirements or to ease data transfer to or manipulation on - other architectures.

Results, Discussion:

In its current - but always evolving state. IMIPS has been successfully used to assess the condition of many epileptic patients and also to plan or to confirm treatment choices. The success of this clinical application both proves the interest and feasibility of moving such complex approaches from the research world to the clinical one

Besides more classical implementations, which mainly deal with standard imaging modalities. our approach has been extended to also incorporate scalp EEG. electro-magnetic tomography and now invasive (i.e. surgically implanted, which means not really used for the investigation of normal function in humans) depth electrodes. strips or grids which record electrical data directly in contact with the structures or tissues under scrutiny From the images generally produced for monitoring the success of the implantation, major structures (e.g the brain) and the electrodes' contacts locations can be extracted by combination of threshold and morpho-topological operations. If there is not too much soft tissue deformation (e.g. because of mechanical stress. subdural hematoma. etc) the major structures are subsequently registered to their pre-surgical equivalent. Any electrode contact can then directly **be** located within the pre-surgical muitisensor world. This allows for disruptive behavior on a specific contact or group of contacts to be compared to the muitisensor signature at the same location(s). If there is too much deformation for accurately registering the "invaded" anatomy to the pre-surgical one. it is generally possible to specify which structures are interacting with which contacts and to point at the corresponding locations in the pre-surgical inultisensor world. For the registration of surgical slides to **3D** digital data, a more refined (i.e local) tit can always be designed but our approach is already providinp us with results which nicely compare with the true 3D solutions (figure 8. 3rd row) **An** interesting corollary we already noted about this approach. that was initially devclopcd to locate cortical grids **on** the **3D** datasets. is that we can also check the extent of the pcrformcd surgery and its agreement with the functional modalities (figure 8. rightmost 3rd row)

In the scope of our ongoing IMIPS dcvclopments, several display modes (including 3-4D objects or hybrid 2-3D images) have been created or modified to merge the dircct or processed information from very diffcrcnt sensors in a manner comprehensible to even non-tcchntcal investigators. Example of current processing and visualization arc presented in figures 6 to 8 In order to finalize the chain of data from acquisition to planning and monitoring of surgery IMIPS can directly be interfaced with commercial neurosurgical microscope navigators (such as the Zciss SMN system) through the delivering of the pre-processed data in a compatible file format. While preliminary planning can be performed with IMIPS alone, final planning and interventional monitoring can then be achieved through the commercial software available on the machine.

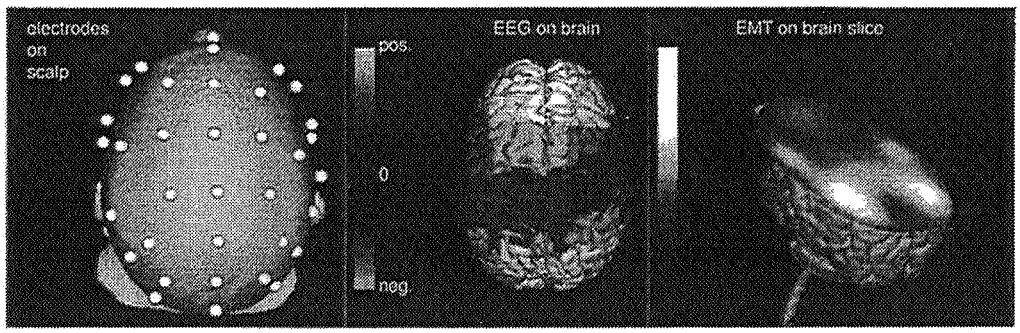


Figure 6: Electrophysiology on anatomy (EEG and EMT)

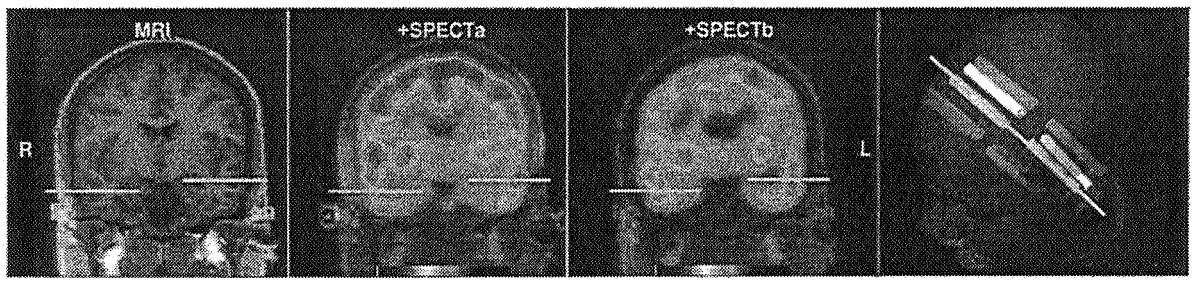


Figure 7: depth electrodes inside a presurgical multisensor space (anatomical MRI, perfusion SPECTs)

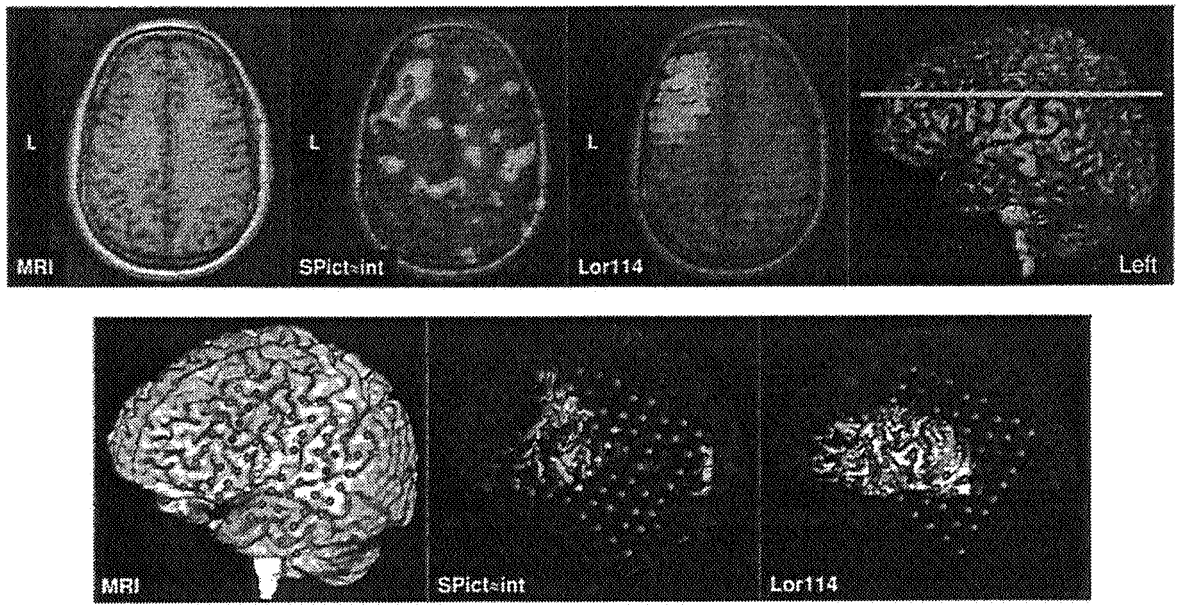


Figure 8: multisensor epilepsy investigation.
first row slices of MRI. SPECT comparison and EMT
second row projection of the same data on the cortex + location of the invasive grid contacts
third row projection of surgical pictures (grid and surgery) on the digital multisensor space and grid contacts

Conclusion:

Aside from more standard imaging modalities for which the interest of multimodality registration has already been well proven. extending the multisensor approach to electrophysiology and more invasive protocols or pictures has been found to be of primary importance for the complete and proper evaluation of complex brain dysfunction such as pharmaco-resistant epilepsy Integrating all the clinical (and digital) investigation techniques now available allows for proper calculations to be performed and for the effective comparison of direct (e.g. invasive) hands-on measurements with preliminary or follow-up (e.g. non-invasive) results From the clinical point of view it is expected that such approaches will improve diagnostic accuracy and eventually minimize the actual need for invasive investigations (or at least help in targeting them better). From the research point of view such global approaches - without the invasive parts - will clearly improve our understanding of brain (dys)function by allowing localization. assessment and study from many and various complementary viewpoints.

Acknowledgments: MRI and CT data courtesy of the Division of Radiodiagnosis, SPECT and PET images courtesy of the Division of Nuclear Medicine. EEG-EMT(Loreta) data courtesy of the laboratory of electrophysiological brain mapping - Clinic of Neurology. epilepsy investigation under the responsibility of the EEG and Epilepsy Unit - Clinic of Neurology, invasive surgery performed in the Clinic of Neurosurgery. all from the Geneva University Hospitals.

References:

1 Gaillard WD. White S, Malow B. et al FDG-PET in children and adolescents with partial seizures: role in epilepsy surgery evaluation. Epilepsy Res. 20(1):77-84. 1995

2 Nagata T Tanakn F Yonekura Y et al: Limited value of interictal brainperfusion SPECT for detection of epileptic foci: high resolution SPECT studies in comparison with FDG-PET Ann Nucl Med. 9(2):59-63. 1995

3 Pascual-Marqui RD. Michel CM. Lehmann D **Low** resolution electromagnetic tomography a new method for localizing electrical activity in the brain. Int J of PsychoPhysiology 18:49-65. 1994.

4. Bidaut, L. M. Pascual-Marqui R.. Delavcllc. J.. et al Three- to five-dimensional biomedical multisensor imaging for the assessment of neurological dys)function. J of Digital Imaging. 9(4):185-198. 1996.

5. Ratib O, Ligier Y, Scherrer J-R Digital image management in medicine. Comput Med Imaging Graph 18-73-84. 1994.

6. Friston KJ. Frith CD, Liddle PF et al Comparing functional (PET) images the assessment of significant change J of Cereb. Blood Flow Metab. 11:690-699. 1991

7 Mangin J-F Frouin V Bloch I. et al: Fast nonsupervised 3D registration of PET and MR images of the brain. J of Cereb. Blood Flow Metab 14.749-762, 1994

8. Jacquez JA. Compartmental analysis in biology and medicine. The University of Michigan Press, Ann Arbor 1988.

9 Pietrzyk U, Herholz K, Fink G. et al An interactive technique for three-dimensional image registration: validation for PET SPECT MRI and CT brain studies J Nucl Med, 35(12):2011-2018. 1994

10. Arun KS. Huang TS. Blostein SD: Least square fitting of two 3D point sets. IEEE Trans Pattern and Mach Intell 9:698-**700,** 1987

11 Meyer CR. Leichtinan GS. Brunberg **JA,** et al Simultaneous usage of homologous points, lines and planes for optimal 3D linear registration of multimodality imaging data IEEE Trans. Med Imaging 14(1):1-11 1995.

12. Pelizzari **CA,** Chen GTY Spelbring DR, et al Accurate three-dimensional registration of CT PET and/or MR images of the brain. J. of Comput. Assist. Tomogr 13:20-26, 1989.

13 van Herk M, Kooy HM: Automatic three-dimensional correlation of CT-CT CT-MRI. and CT-SPECT using chamfer matching. Med. Phys. 21 1 163-1178, 1994.

14 Woods R, Mazziotta JC Cherry S MRI-PET registration with automated algorithm. J of Comput Assist. Tomogr 17:536-546, 1993.

15. Appel R.D. Palagi P., Walther D. et al MelanieII a third generation software package for analysis of two-dimensional electrophoresis images - I Features and user interface Electrophoresis 18. 1997(in press).

16 Ames AL. Nadeau DR. Moreland JL The VRML source book. Wiley. New York. 1996.

Poster Session

Building 28 - Atrium

Image Registration By Parts

Prachya Chalermwat
Dept of Electrical Engineering and Computer Science
The George Washington University
Washington, DC 20052

Tarek El-Ghazawi
Computer Engineering and Computer Science
Florida Institute of Technology
Melbourne, FL 32901-6975

Jacqueline Le Moigne
Center of Excellence in Space Data and Information Sciences
Code 930.5, NASA GSFC
Greenbelt, MD 20771

Abstract

In spite of the large number of different image registration techniques, most of these techniques use the correlation operation to match spatial image characteristics. Correlation is known to be one of the most computationally intensive operations and its computational needs grow rapidly with the increase in the image sizes.

In this article, we show that, in many cases, it might be sufficient to determine image transformations by considering only one or several parts of the image rather than the entire image, which could result in substantial computational savings

This paper introduces the concept of registration by parts and investigates its viability It describes alternative techniques for such image registration by parts and presents early empirical results that address the underlying trade-offs,

1 Introduction

Given a reference and an input images, image registration is the process which determines the most accurate match between these two images [1, 3]. In remote sensing, mage registration is crucial as successive observations are collected from moving satellites. Due to the tremendous amount of data generated from the MTPE remote sensing satellite systems and the amount of images to be registered, the time required to register input images to existing reference images is quite large. This prompted the need for fast computational methods to conquer the lengthy computation time [2, 4, 5]

Even on high performance computers, comparing two images at their full resolution takes considerable amount of time [2] One straightforward solution is to cut the computational cost by basing the registration on only one or several significant parts of the image. This gives rise, however, to many important questions. For instance it is not clear how the derived registration would differ depending upon the used parts of the image, and if so, then how to select the best subimages for such purpose.

This paper is organized as follows. Section two describes the registration by parts. Section three details early experimental results. Conclusions and future directions are given in section four

2 Registration by Parts

Image registration by parts saves computational cost by correlating corresponding subimages in the reference and the input images rather than using the entire images [1] It could be considered in the following cases:

(1) When it is known that the transformation between reference and input images is global and uniform instead of the images, only one part of the image can be utilized. Registering only one part of the image over the entire image may also be a trade-off between computation time and registration accuracy.

(2) When knowledge about local distortions is not known, a choice must be made about the optimal number and location of subimages to correlate. This choice can be eased by the use of known geographic landmarks or features in the image. Utilizing several well-distributed subimages enables to save computation time and to simultaneously verify the consistency of the computed transformations over the image. If the same transformation is found over all subimages, the final result is given with a higher confidence. If we are looking for a rigid transformation and significantly different transformations are locally found, the registration strategy should be changed to finding a more flexible polynomial, or non-linear transforma-

[1] This is under the assumption that the transformation of an input image does not exceed the size of the subimage

tion, or perform a larger number of local transformations. In this last case, a "seam problem" will have to be solved at the intersection of all image parts. More generally, we can imagine an iterative registration scheme which starts by registering a small number of well-distributed image parts and depending on the consistency of the results will choose or not to increase the number and decrease the size of new image parts to correlate.

Consider reference and input images of size N x N which are subdivided into subimages of size M x M each There exists three possible ways of considering the subimages, each of which provides a different level of computational savings. At one extreme, we can only select one subimage for the registration process, which provides a computation savings in the order of $O(\frac{N^4}{M^4})$, where $O(M^4)$ operations become required instead of $O(N^4)$ Another possibility is to consider only k of the subimages, which produces a speed up gain of $O(\frac{M^4}{k})$, where k is an integer between 1 and $\frac{N^2}{M^2}$ for non-overlapping subimages, and between 1 and $\frac{N^2}{(M-p)^2}$ for p-pixel overlapping subimages. Even if all partitions are used, still a computational saving by a factor of $O(\frac{N^4}{\frac{N^2}{M^2}\times M^4}) = O(\frac{N^2}{M^2})$ would result.

With these options, a number of possible algorithmic requirements arise:

Case I: Using only 1 Partition

In this case, there is a need to develop appropriate criteria and a method for selecting the best part to use for the correlation. Criteria may include the energy content, the edge content of the selected part, or some knowledge of known geographic landmarks or features. Clearly such criteria need to be selected for a specific class of images, as a result of an experimental investigation to determine the best criteria. Another important issue is the overhead cost associated with finding the subimage which is likely to yield accurate registration. It is important for that overhead to be maintained within limits that are much lower than the computation savings derived from the use of registration by parts. *However, it should be noted that if we are selecting parts from the reference image, then evaluating the image to select the right parts can be done off-line prior to receiving new input images from satellites or any other input processes This eliminates the cost of the overhead for such a method from the operational viewpoint.*

Case 11: Using k partitions

In this case, the registration results from the k partitions can be used in several ways. As was described previously, different decisions will be taken depending on the consistency of the correlation results in each partition. If the results mainly agree with each other, the simplest solution is to average the correlation function from all selected partitions and use the best correlation point to indicate registration parameters. Another way is to use a voting scheme, in which the results from these partitions are clustered and the result from the largest cluster is selected. If the results are significantly different from each other, the corresponding correlations must be considered to check if all results have similar confidence levels. If not, the previous voting scheme can be utilized. If confidence levels are similar, two decisions can be made: looking at smaller, more numerous image parts, or considering local transformations.

Case III: Using all partitions

In this case, all $\frac{N^2}{M^2}$ partitions are used. Clusters can be treated in the same way as case II.

The alternate routes to implement the registration by parts leave a number of open questions to carry the research to the next level. These include:

1. Determining the criteria for selecting M, or the size of subimages in a given class of images
2. Selecting the set of subimages to use (1, k, or all)
3. Selecting the location of the subimages and the criteria for doing so (e.g. energy or edge contents)

One possible algorithm to select the location and size of a subimage, which is likely to yield good registration results, would be to follow a quad tree search. The algorithm looks for the smallest subimage that meets a pre-specified criteria such as the energy content or the edge content.

Note that, in registration by parts, two images can be compared at their full resolution partially. Figure 2 illustrates selected 16 subimages of the reference image and the input image. We perform correlation between subimages of the reference and input images

2.1 Rotation, Translation, and Transformation

Most transformations found in satellite images can be modeled by an affine transformation As a first step

to find such transformations, we will be looking at the composition of rotations and translations, which form the basic components of affine transformations. Figure 1illustrates the orientation of our rotation and translation. Rotation is implemented clock-wise and the degree is in integer format. The translation is also in integer format. Positive X-axis translation is to the right of the center Positive Y-axis translation is below the center

We use a *transformatzon index* to plot the curves for correlation, because it would take too many curves of the correlation values for all combination of rotations and translations separately The combination of rotation and translation forms an array of *transformations.* The *transformation* is simply an integer representing index of each combination of θ, Tx, and Ty The transformation is invertible. In the results of this paper, we will refer to the graph plotting the correlation V.S. the transformation index as the transformed registration graph.

2.2 Correlation Function

The following equation shows the correlation function that we use to measure the similarity of the two images, A and B of size $M \times M$ A_i is a pixel value of the i^{th} pixel in image A

$$correlation(A,B) = \frac{\sum_{i=1}^{N^2}(A_i - \overline{A}) \times (B_i - \overline{B})}{(\sqrt{\sum_{i=1}^{N^2}(A_i - \overline{A})^2}) \times (\sqrt{\sum_{i=1}^{N^2}(B_i - \overline{B})^2})} \quad (1)$$

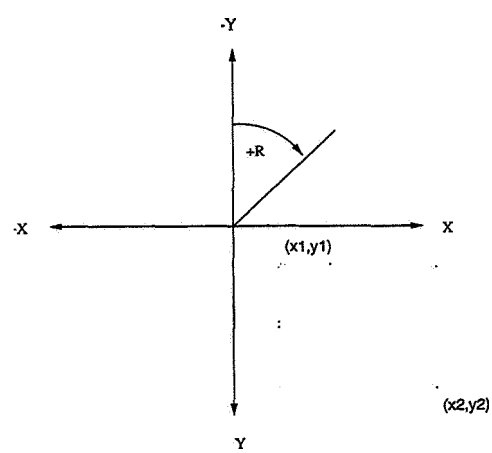


Figure 1 Orientation of rotation and translation

3 Early Experimental Results

3.1 Experiment Setup

The following reference and input images are used [2]

Reference	Input	Size
girl-ref	girl.r5	512x512
tm.ref	tm.r4tx5ty2	512x512
goes8_101615.ch1	goes8_102015.ch1	1024x1024

The girl.ref, tm.ref, and goes8_101615.ch1 are reference images. The girl.r5 is a 5-degree rotated input image. The tm.r4tx5ty2 is an input image with 4 degrees rotation, 5 pixel x-translation, and 2 pixel y-translation. The goes8_102015.ch1 is an input image with unknown transformation

For the registration by parts of each pair of images, we experimented using three different numbers of subimages: 16, 64, and 256. The search scope is θ ={-5,-4,..,0,..4,5} for rotation, and tx, ty = {-10, -8,..,0,..,8,10} for translation The translation has a 2-pixel increments. Hence, the total number of transformations is 11x 11x 11 = 1331 Thus, for each degree of rotation, there are 11x 11=121x-y translations Note that the vertical grid on x axis are divisible by 121.

For each correlation graph, the correct transformation is marked as "reference" with a vertical dot line. For example, in figure 3, the rotation of 5 degrees has the transformation(5, 0, 0) at 1270.

3.2 Correctness of Registration by parts

Table 2 shows the results of registration by parts using 16 subimages. Let assume that we allow an error of ± 1 degree or ± 2 pixels for rotation and translation respectively. [3] Let C be an average of the correctness of the registration by parts. Its value ranges from 0 to 1 C = 0 when none of subpart registration are correct and C = 1when all subpart registrations are correct. As shown in table 2, all rotational registration are 100% correct except the "GOES" While the "girl" gives about 81% and 88% for x and y translation respectively, the x and y translation for the "TM" (thematic mapper) yield 100% and 88% correctness. Under the assumption that the "GOES"

[2] Available from anonymous ftp at cesdis.gsfc.nasa.gov under directory
/pub/people/lemoigne/outgoing/TOOLBOX_TESTING

[3] This is because we use an increment of two for translation

images transformation $(\theta, x, y) = (0, 0, 0)$, **63%** of x and y translation are correct.

		Girl---		--TM--			--GOES--		
		X	Y	R	X	Y	R	X	Y
0	5	0	-10	4	4	2	1	-10	-4
1	5	0	0	3	6	-2	0	0	0
2	5	**0**	0	4	4	2	1	0	8
3	4	**-4**	-4	4	4	2	-1	10	-10
4	4	2	-2	4	4	2	0	0	0
5	4	0	-2	4	4	2	0	0	0
6	5	0	0	4	4	2	0	0	0
7	5	**0**	0	4	4	2	2	10	0
8	5	0	0	4	4	2	0	0	0
9	4	0	0	4	4	2	1	-10	10
10	4	-2	0	4	4	2	-1	10	-8
11	4	-4	0	4	4	2	0	2	0
12	5	0	0	4	4	2	0	0	0
13	4	0	2	4	4	2	1	10	-4
14	5	0	-2	5	6	-2	0	0	0
15	4	-4	2	4	4	2	0	2	0
C	1	.81	.88	1	1	.88	.94	.63	.63

Table **2:** Registration by parts results

Figure **2:** Girl. **512x512** reference and 5-degree rotated input image

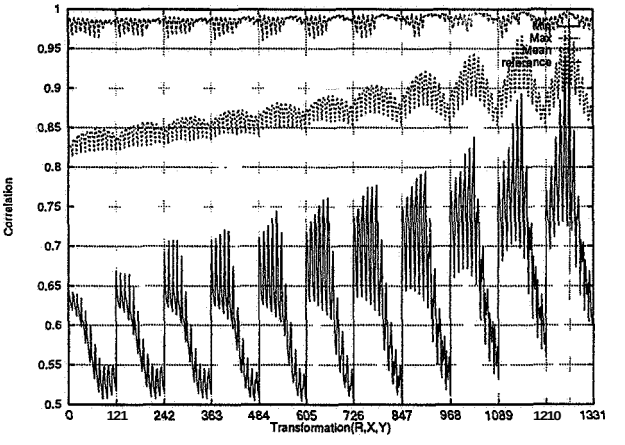


Figure **3.** Girl: Overall correlation values of **16 128x128** parts

3.3 Results for Girl, TM, and GOES

Figure **2** illustrates a **16** partitioned reference image and an input image. Figure **3, 4,** and **5** show the correlation values of the registration by parts using subimages of size **128** x **128, 64** x **64,** and **32** x **32** respectively For each transformation on **x** axis, we plotted three values of correlations of **all** subimages: minimum, maximum, and mean. The minimum is taken from the smallest correlation among all correlation of all subimages. Similarly, the maximum is taken from the largest correlation among all correlation of all subimages. The mean is an average of all subimage correlations Figure **8,** 9, and **10** show the correlation values of the registration by parts using subimages of size **128** x 128, **64** x **64,** and **32** x **32** respectively. For **1024** x **1024** size of image, figure **13, 14,** and **15** show the correlation values of the registration by parts using subimages of size **128** x **128, 64** x **64,** and **32** x **32** respectively.

3.4 Characteristics of Subparts Correlation

As an example, we take a registration of the thematic mapper images and observe the correlation values of **all 16** parts of size **128** x **128.** Figure **16-31** demonstrates the correlation values of each subpart from the registration by parts. We observed that all subregistration of subimages can lead us to correct registration However, this may not be true for all cases because the correlation is data dependent.

4 Conclusions and Future Directions

In this **work** we have defined the notion of registration by parts, the relevant algorithms, and the performance issues. The performance gain in all possible implementations was found to range from a factor of $\frac{M^4}{k}$ to M^4, where k is the number subimages of size M x M Computational overhead associated with identifying the parts can be practically eliminated by using the reference images to pre-determine these parts off-line.

The early experimental results have shown strong evidence that sufficient information for efficient regis-

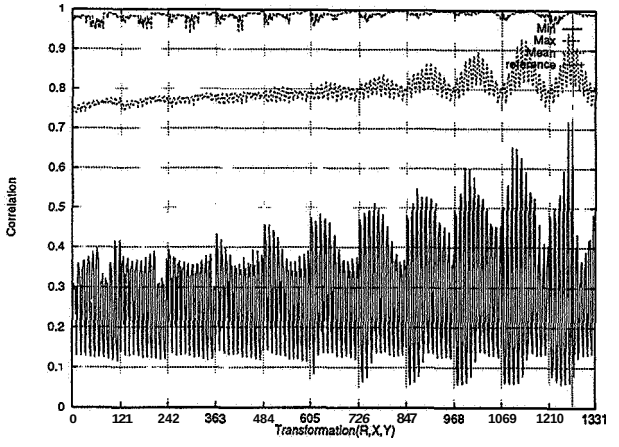


Figure **4:** Girl: Overall correlation values of **64 64x64** parts

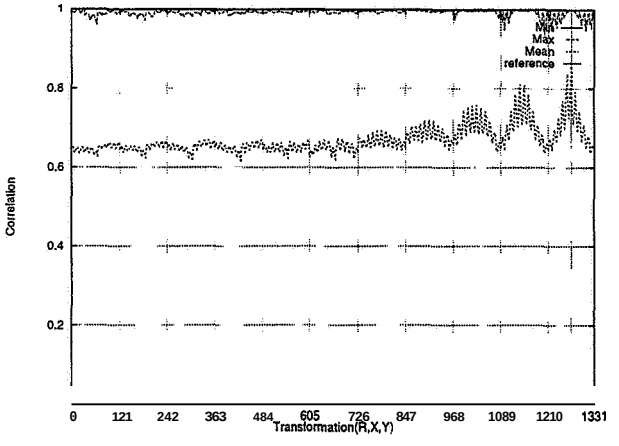


Figure **5:** Girl: Overall correlation values of **256 32x32** parts

tration are retained in most subimages, **as** the subimage sizes were decreased to **32** x **32** in images **as** large as **1024** *x* **1024.** The average correlation from these subimages was quite consistent over the different subimage sizes. Furthermore, the majority of the subimages produced the correct registration parameters The video image "girl" (figure **2)** gave the best results across **all** subimages due to the strong features and minimal noise. The thematic mapper (figure **6** and **7)** gave the second best performance under the registration by parts, followed by the GOES images (figure **11** and **12)** In the case of the GOES images, **60%** of the subimages were still capable of producing correct results using **256** x **256** subimages.

Future direction includes experimental investigations leading to the best method for selecting the subimages, and for the the experimental evaluations of registration by part algorithms **(as** described

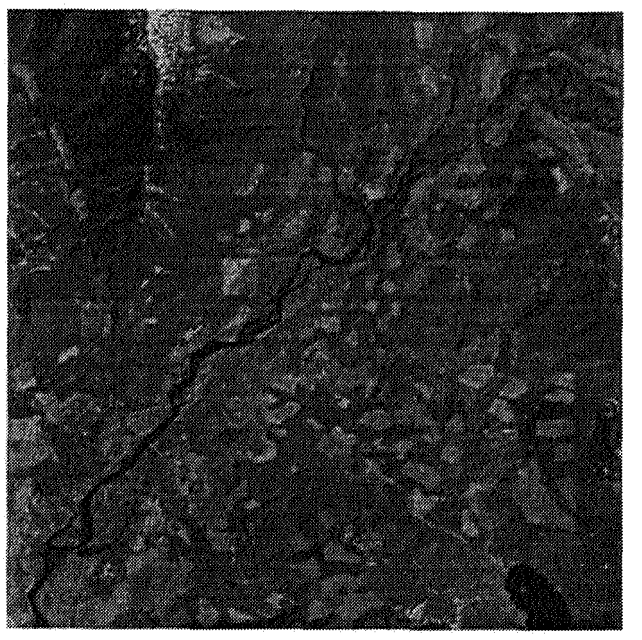


Figure **6:** Thematic Mapper· **512x512** reference image

above) across a large number of NASA images For further execution time improvements, parallel image registration by parts will be also investigated and evaluated experimentally over different high-performance computers

References

[1] M. Corvi and G Nicchiotti, "Multiresolution Image Registration," *in Proceedings of 1995 IEEE International Conference on Image Processing*, Washington, D.C., Oct, **23-26, 1995.**

[2] T El-Ghazawi, **P** Chalermwat, J LeMoigne, "Wavelet-Based Image Registration on Parallel Computers," *an Proceedings of SuperComputing* 97, San Jose, Nov **15-21, 1997**

[3] E. Kaymaz, B-T Lerner, J F Pierce, W.J Campbell, and J. Le Moigne, "Multi-Resolution Geo-Registration of Remote Sensing Imagery Employing the Wavelet **Transform,"** *SPIE 25th AIPR Workshop*,Washington, DC, October **16-18,1996.**

[4] J Le Moigne, "Towards a Parallel Registration of Multiple Resolution Remote Sensing Data," *in Proceedings* of *IGARSS'95*, Firenze, Italy, July **10-14, 1995.**

[5] J Le Moigne, W.J Campbell, and R.F Cromp, "An Automated **Parallel** Image Registration

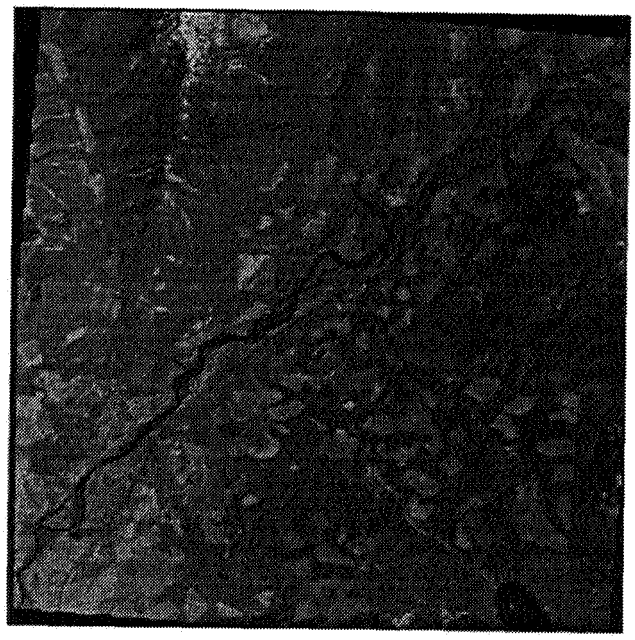


Figure 7: Thematic Mapper. 4 degrees rotated input image (tm.r4tx5ty2)

Technique of Multiple Source Remote Sensing Data," accepted for publication *in IEEE Bans-actions on Geoscience and Remote Sensing,* 1998.

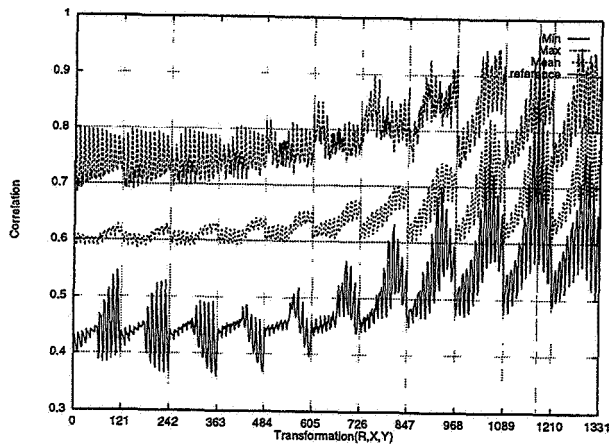


Figure 8: Thematic Mapper. Overall correlation values of 16 128x128 parts

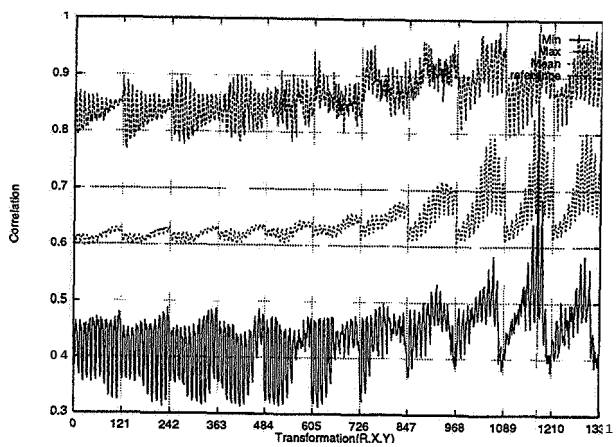


Figure 9: Thematic Mapper- Overall correlation Values of 64 64x64 parts

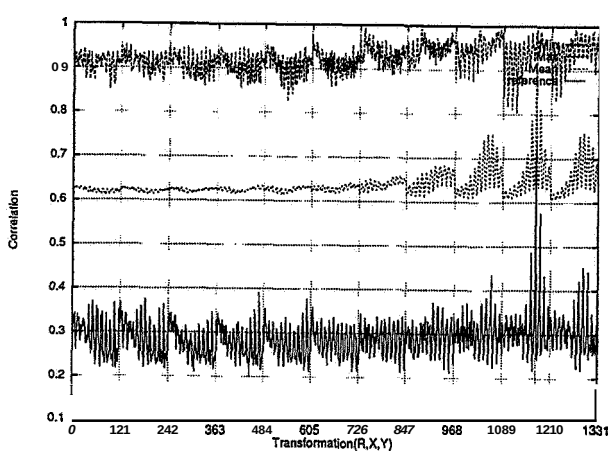


Figure 10: Thematic Mapper. Overall correlation values of 256 32x32 parts

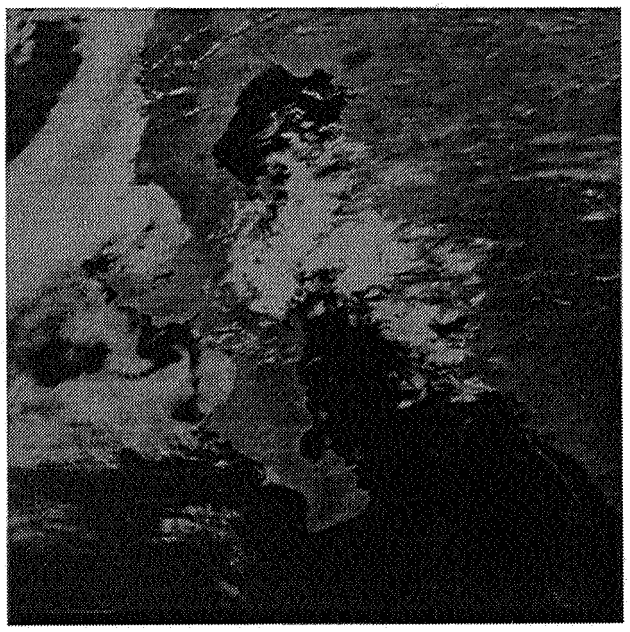


Figure 11 GOES: 1024x1024 reference image

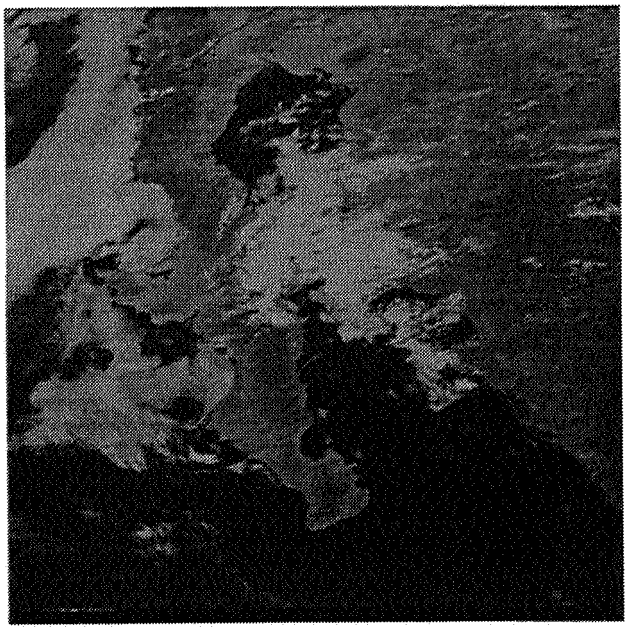


Figure 12: GOES: 1024x1024input image

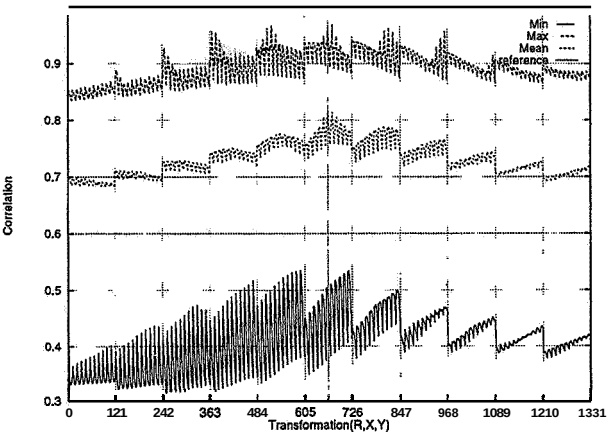


Figure 13: GOES. Overall correlation values of 16 128x128parts

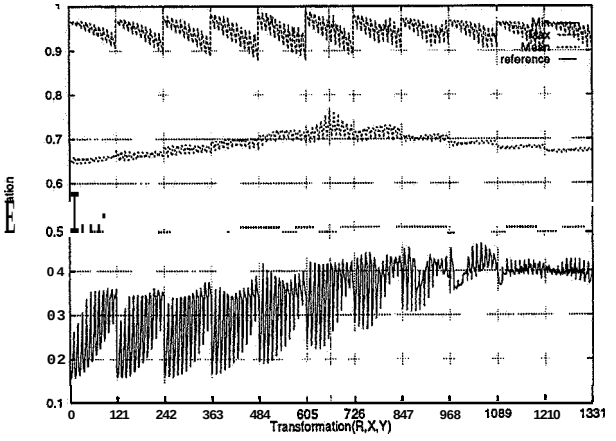


Figure 14: GOES: Overall correlation values of 64 64x64 parts

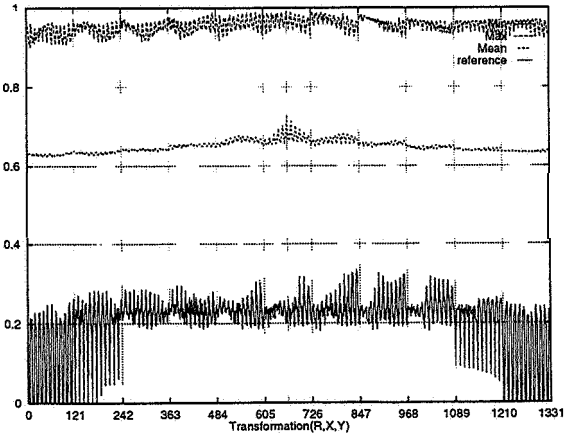
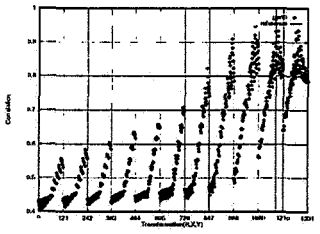


Figure 15, GOES: Overall correlation values of 256 32x32 parts



Fimre 16: Part 1

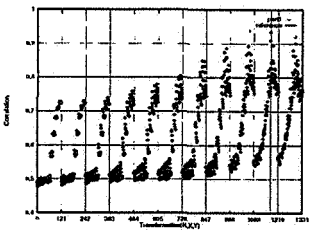


Figure 1 7 Part 2

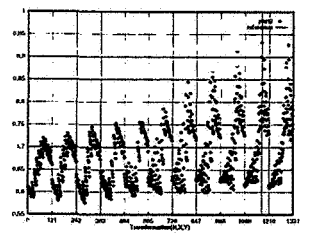


Figure 18: Part 3

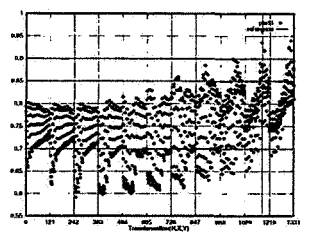


Figure 19: Part 4

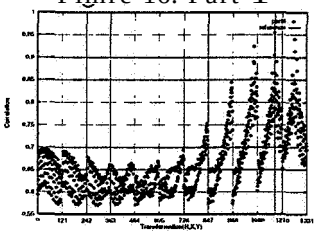


Figure 20: Part 5

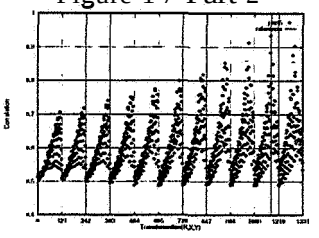


Figure 21. Part 6

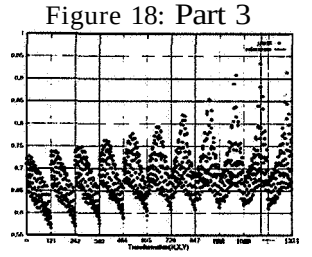


Figure 22: Part 7

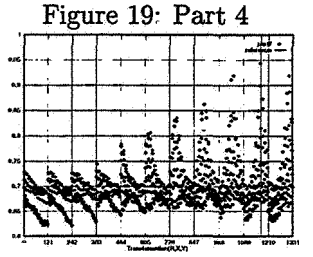


Figure 23: Part 8

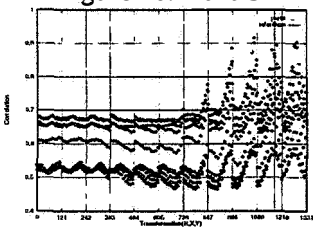


Figure 24. Part 9

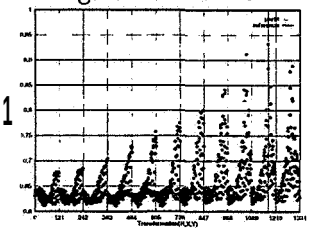


Figure 25: Part 10

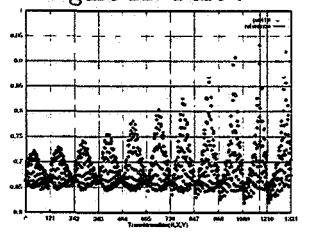


Figure 26: Part 11

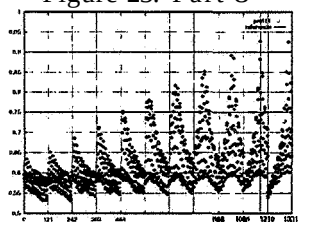


Figure 27. Part 12

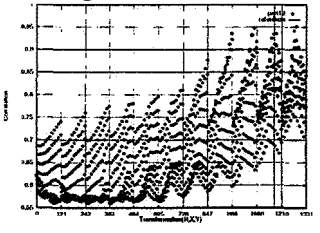


Figure 28. Part 13

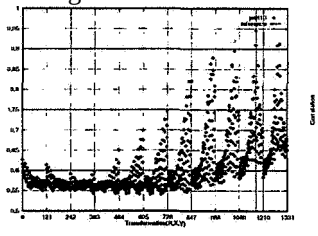


Figure 29: Part 14

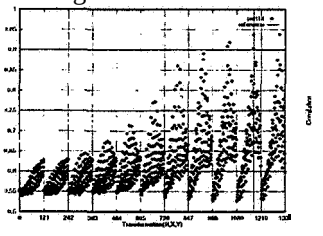


Figure 30: Part 15

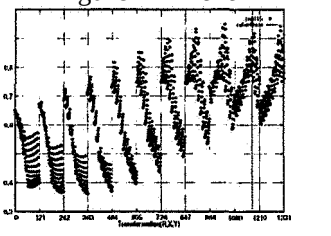


Figure 31. Part 16

Figure 32: Thematic Mapper, Correlation values of subimage 1-16 of size 32 x 32

Towards an Intercomparison of Automated Registration Algorithms for Multiple Source Remote Sensing Data

Jacqueline Le Moigne [1], *Wei Xia* [2], *Samir Chettri* [3], *Tarek El-Ghazawi* [4], *Emre Kaymaz* [5], *Bao-Ting Lerner* [5], *Manohar Mareboyana* [6], *Nathan Netanyahu* [1,7], *John Pierce* [5], *Srini Raghavan* [2], *James C. Tilton* [8], *William J. Campbell* [8], *Robert F. Cromp* [8]
lemoigne@cesdis.gsfc.nasa.gov

(1) USRA/CESDIS; (2) Hughes STX, (3) GST; (4) GWU; (5) KTT; (6) BSU; (7) UMD;
(8) NASNGSFC - Contact: Code 930.5, Goddard Space Flight Center, Greenbelt, MD 20771

Abstract

The first step in the integration of multiple data is registration, either relative image-to-image registration or absolute geo-registration, to a map or a ked coordinate system. As the need for automating registration techniques is recognized, we feel that there is a need to survey all the registration methods which may be applicable to Earth and space science problems and to evaluate their performances on a large variety of existing remote sensing data as well as on simulated data of soon-to-be-flown instruments. In this paper we will describe: 1) the operational toolbox which we are developing and which will consist in some of the most important registration techniques, and 2) the quantitative intercomparison of the different methods, which will allow a user to select the desired registration technique based on this evaluation and the visualization of the registration results.

1. Introduction

In studying how our global environment is changing, research in the Mission to Planet Earth (MTPE) program involves the comparison, fusion and integration of multiple types of remotely sensed data at various temporal, radiometric and spatial resolutions. Results of this integration will be utilized for global change analysis, as well as for the validation of new instruments or of new data analysis.

The first step in the integration of multiple data is registration, either relative image-to-image registration or absolute geo-registration, to a map or a ked coordinate system. Currently, the most common approach to registration is to extract a few outstanding characteristics of the data, which are called *control points* (CP's), *tie-points,* or *reference points.* The CP's in both images (or image and map) are matched by pair and used to compute the parameters of a geometric transformation. Most available systems follow this registration approach, and because automated procedures do not always offer the needed reliability and accuracy, current systems assume some interactive choice of the CP's. But such a point selection represents a repetitive, labor- and time-intensive task which becomes prohibitive for large amounts of data. Also, since the interactive choice of control points in satellite images is sometimes difficult, too few points, inaccurate points, or ill-distributed points might be chosen thus leading to large registration errors. A previous study [1] shows that even a small error in registration may have a large impact on the accuracy of global change measurements. For example, when looking at simulated 250m spatial resolution MODIS (Moderate Resolution Imaging Spectrometer) data, a one pixel misregistration can produce 50% error in the computation of the Normalized Difference Vegetation Index (NDVI). So, for reasons of speed and accuracy, automatic registration is an important requirement to ease the workload, speed up the processing, and improve the accuracy in locating a sufficient number of well-distributed accurate tie-points.

As this need for automating registration techniques is recognized, we feel that there is a need to survey all the registration methods which may be applicable to Earth and space science problems and to evaluate their performances on a large variety of existing remote sensing data as well as on simulated data of soon-to-be-flown instruments.

Although automatic image registration has been extensively studied in other areas of image processing, it is still a complex problem in the framework of remote sensing. Given the diversity of the data sources, it is unlikely that a single registration technique will satisfy all different applications. We propose to: 1) develop an operational toolbox which will consist of some of the most important registration techniques, and 2) provide a quantitative intercomparison of the different methods, which will allow a user to select the desired registration technique based on thii evaluation and the visualiition of the registration results. The algorithm intercomparison will be based on the following criteria.

- accuracy
- applicability and restrictions (e.g., adaptability to multiple sensors)
- level of automation
- computational requirements (and adaptability to large datasets).

These criteria will be computed for each algorithm utilizing several datasets such as: (a) Multi-temporal, intra-sensor data, e.g., NOAA/AVHRR, Landsat TM and MSS, GOES, and hyperspectral such as ASAS data, and (b) Inter-sensor data, e.g., MODIS Airborne Simulator (MAS) vs. Landsat/TM data, and AVHRR vs. Landsat/TM

We have chosen the Khoros environment as the framework for the implementation of these techniques. Khoros is an object-based data analysis, data visualization, and application development environment. In Khoros, a "toolbox" is a collection of programs and libraries that is handled as a single object. In that sense, our registration toolkit is also composed of the various registration routines, each of which can be handled as an object. Khoros is also an open software system and will enable us to widely distribute the toolbox and to get feedback from a variety of users. Having the toolbox developed in Khoros also makes it easy to distribute with the software developed by the Applied Information Sciences Branch at NASA/Goddard Space Flight Center for the Regional Validation Centers (RVC's) program [2]. The RVC's will receive remote sensing data by direct readout from various satellites and the users will utilize this software to process in real-time data needed for their regional applications (e.g., monitoring regional change, storm prediction, etc.). Successive versions of the toolbox will be integrated with this software and will also provide us with feedback fiom the remote sensing community This feedback will enable us to develop, or help the users develop, new tools which would be more adapted to their particular applications.

2. A Registration Toolbox

The distortions described in [3] for Landsat-MSS data are general enough to account for most sensors directed at the Earth. Distortions are defined by the combined effects of sensor operation, the Earth's rotation, orbit and attitude anomalies, and atmospheric and terrain effects. These distortions, which are noticeable when dealiig with only one sensor, are even larger when considering multiple sensors carried by multiple platforms on different orbits, or when looking at airborne sensors [4]. As a first approximation, the distortions corresponding to orbit and attitude anomalies mainly correspond to an affine transformation. A few other small continuous nonlinear distortions due to altitude, velocity, yaw, pitch, and roll can be handled by global or local polynomial or elastic transformations of a higher degree.

We assume that any new incoming *sensed* image is being registered relative to a known *reference* image or map. We also assume that *sensed image* and *input image* are synonymous. According to Brown [5], image registration can be viewed as the combination of four components:

(1) *a feature space,* i.e., the set of characteristics used to perform the matching and which are extracted fiom reference and input data,
(2) *a search space,* i.e., the class of potential transformations (or deformation models) that establish the correspondence between input data and reference data (e.g., rigid, affine, or polynomial transformations);
(3) *a search strategy,* which is used to choose which transformations have to be computed and evaluated;
(4) a similarity metric, which evaluates the match between input and transformed reference data for a given transformation chosen in the search space.

Correlation measurement is the usual similarity metric [6], although it is computationally expensive and noise sensitive when used on original gray level data. Using a pre-processing such as edge detection or a multi-resolution search strategy enables large reductions in computing time and increases the robustness of the algorithms. Other similarity metrics are described in [7], and a comparison of isometric (which preserves distances) versus polynomial transformations is given in [8]. Although high-order polynomials are superior to isometric transformations when the deformation includes more than a translation and a rotation, isometric transformations are more accurate when a smaller number of points is available and are less sensitive to noise and to largely inconsistent tie-points. Several feature extraction and feature matching methods have been integrated and compared in a system developed by Rignot et al [9]. Another way to classify registration methods is adopted by Manjunath et al. [10], where twelve recent registration methods are described according to the type of their feature extraction. Automatic feature extraction is either area-based or feature-based, the feature-based process occurring either in the spatial domain or in the transform domain.

Another viewpoint is taken in a study performed by Le Moigne et al [11] and describes a majority of the registration methods found in the literature since 1972 which have the potential for being applied to remote sensing images (see Tables 1&2). As previously described, the usual human/manual registration first determines control points and then performs a matching point by point before computing the deformation model. Automatic methods can be classified into two types, those which follow a human approach, and those which take a different, more global approach. See [11] for references and a more detailed discussion of all these techniques.

This large variety of techniques which could be utilized for remote sensing images leads us to the development of a toolbox of registration techniques, where each technique will be associated with a set of parameters and working conditions.

Reference	Feature Space	Search Space	Method	Application
Cracknell et al, 1989	Binary Shapes after Smoothing and Thresholding	Rigid Transformations	Automatic Method: * Shape Extraction and Cross-Correlation with Chips (Id. or Higher Resol.)	Registration of AVHRR Data with Sub-Pixel Accuracy
Corvi et al, 1995	Maxima and Minima of MRA Wavelet Coefficients	Rigid Transformations	Multiresolution with Initial Estimate/LMS point matching	Banknotes, Mechanic. Pieces, Aerial Imag., Multispectral and SAR
Djamdji et al, 1993	Maxima of Wavelet Coefficients	Polynomial Transformations	Iterative Matching Maximum to Maximum	Multisensor Images:SPOT MSS,SPOT/Landsat
Fiore, 1995	Feature Points (Assume already Extracted)	Rotations and Translations	Matching the two sets of Points by Relative Distances and Angles	Not Defined
Flusser et al, 1992&94	Regions with Gravity Centers and Affine-Invariant Moments	Affine Transformations	Match Regions by Features then by Distances of Gravity Centers + Adaptive Mapping	SPOT with TM Imagery with Sub-Pixel Accuracy
Goshtasby&al, 1986&88	Gravity Centers of Middle-Size Regions (Iterative Thresholding)	Rigid and Polyn. Transformations	Iteratively Refine Regions and Gravity Centers; mapping with 2 surface splines(X&Y)	Landsat TM,MSS,MSS2, Simulator;Text
Li, Manjunath, and Mitra 1995	Gravity Centers of Closed Contours and Corners of Open Contours	Rigid Transformations	Match Contours and CP's (Elastic Contour Extraction for ***SAR)***. Iterative registr with consistency checking	Aerial and Satellite Visible/Visible and Visible/SAR (multiband, multisensor)
Li &Zhou 1995&96	High-Interest Points, e.g. from multi-resolution wavelet contours	Affine Transformations	Multi-Resolution Iterative Matching of High-Interest Points	Visible/SAR Aerial Images
Devereux et al 1990	Selected GCP's + CP's from Triangulation	Locally Affine Transformations	Iterative Triangulation of GCP's to add new CP's + Matching of Triangles	Airborne Thematic Mapper (ATM) Imagery
Lee&Chen 1990	High-Interest Points Predicted by Voronoi-Delaunay Diagram	Affine Transformations	3 initial pts; Predict new CP's by Voronoi Diag., Interest operators;LMS;Consist.check	Aerial Imagery
Ton et al, 1989	Centroids of Water and Pad Regions	Translations and Rotations	Point Matching Using Relative Distances and Areas in Relaxation Scheme	Landsat-TM
Ventura et al, 1990	Structures with Areas, Main Axis Angle, Ellipticity, Distances	First-Order Polynomials	Labelling Process by Logic System (Multi-Value Logical Tree)	Registration of SPOT/Landsat-TM and Landsat/map
Zheng&Chellappa 1993	Local Curvature Discontinuities Based on a Gabor Wavelet	Rigid Transformations	Rotation by Illuminant Direct. Estimation; Shift and Scale by Features Match (Multi-Resol.)	Aerial and Satellite Imagery

Table 1

Automatic Registration Methods With Point to Point Matching (See [11] for references)

Reference	Feature Space	Search Space	Method	Application
Allen et al, 1993	Laplacian or Wavelet Coefficients of Decomposed Curvature	Translations	Mutiresolution/Multipath Search/ Cost Function	Registration of 1-D Coastlines
Anuta, 1970	Gray Levels or Gradient Edges or Region Borders	Rigid Transformations	Global Search/ Phase Correlation/ Least Mean Squares	Multispectral and Multitemporal Aerial Imagery
Barnea, 1972	Gray Levels	Translations	Sequential Similarity Detection Algorithm, SSDA (Extension of Correlation)	Registration of Weather Satellite NOAA/ITOS-1 Data
Chen et al, 1994	Output of Fourier-Mellin transform	Rigid Transformations	Find Maximum of Phase-Only Matched Filtering	Various Applications (Medical, Remote Sensing, Fingerprint, Multiobject Identif.)
Lucas et al, 1981 and Irani et al, 1991	Original Image and Derivatives	Translations and Rotations	Iterative Optimization with Derivatives and Assuming Smaller and Smaller Displac.	Stereo Vision and Visual Test Images
Kamgar-Parsi& al, 1991 [29]	Contours of Constant Range	Rotations + Translations	Local by Matching Pairs of Contours; Global by Optimization	Register Mapping of Ocean Floor
Olivo et al 1995	Non Overlapping Blocks in Multi-Resolution Wavelet Images	Translations	Multi-resolution Block Matching	Objects on Background, e.g., Medical Images
Stockman et al, 1982	Edge Segments or Vectors Linking Feature Points	Rigid Transformations	"Hough Accumulator" to Measure All Possible Transf. Between Features	Aerial Images/Map Airplane Detection Industrial Inspection
Unser et al, 1993 and Thévenaz et al, 1995	Multiresolution Spline Pyramid Images (2-D and 3-D)	Affine Transformations	Multi-Scale Iterative Marquardt-Levenberg Optimization	Medical Images
Wu et al, 1990	Multi-Resolution Scale-Space Edge Boundaries of Coastlines	Rigid Transformations	* Approx from Inflection Pts * Consistency Check * Accurate Chamfer Matching of Distance Maps	Registration of Seasat-SAR against SPOT Images

Table 2

Automatic Registration Methods With Global Point Matching (See [11] for references)

Depending on the type of sensors, the desired registration accuracy, the computer availability, and the speed requirement, one or another of the techniques will be chosen. Multiple resolutions of the sensor data will be dealt with by the different approaches included in the toolbox. A multi-resolution approach such as the one taken by a wavelet-based technique deals with the multiple spatial resolution data. Other content-based approaches such as an object-oriented registration might be more appropriate for dealing with multi-temporal or multi-radiometric approaches. This toolbox will be implemented on several architectures, new parallel architectures as well as the DEC-Alpha workstation, under a widely-used software package such as the Khoros image processing package.

The first version of the toolbox includes the following techniques:

- *Semi-manual registration* where pairs of corresponding control points are manually selected, followed by the transformation computation (with choice of polynomial, rotation, translation, rigid or affine transformations).
- *Correlation-Based Methods* including phase correlation [14] and spatial correlation [15],
- *Feature-Based Methods* with edge-, corner-, region- and wavelet-based methods [13,16,17,18,19],
- *Post-Processing Tools*, such as interpolation of the matching functions, matching based on moment invariants, as well as robust matching of points [30], given, for example, by a region-based method.

Figure 1 shows the top level of the Khoros graphic user interface of the current registration toolbox, and the methods included in the first version of the toolbox are detailed in sections 3 and 4 The next versions of the toolbox will include a larger choice of techniques from the ones described in Tables 1&2, as well as newly developed algorithms. Future methods will also include cloud masking [20].

3. Known Technologies

3.1. Semi-Manual Registration

The semi-manual tool is similar to the method most commonly utilized for registration; a human operator "manually" selects *Ground Control Points* (GCP's) in *two* images, and these points become the input to compute the deformation model between the two datasets, often chosen as a polynomial transformation. In our implementation, users select GCP's fiom the displayed reference image and the image to be registered respectively Zoom capabilities are available to help users choose the GCP's more accurately Then a choice of transformations is provided to the user' rotation, translation (e.g., shift), rigid, affine, and polynomial transformations. The GCP's are then used to calculate the parameters of the chosen transformation: either the rotation angle, or the translation shifts, or the transformation coefficients for rigid, affine, or polynomial transformations.

3.2. Phase Correlation

Phase correlation is a mathematical technique that was developed to register images in which the misregistration is only a translation [14]. The technique *can* be described as follows: given a reference image, g_R and a sensed image g_S, with 2-D Fourier transforms G_R and G_S, respectively, the cross-power spectrum of the two images is defined as $G_R G_S^*$ and the phase of that spectrum as

$$e^{j\phi} = \frac{G_R G_S^*}{|G G^*|}$$

The phase correlation function, d, is given by the inverse Fourier transform of that spectrum:

$$d = F^{-1}\{e^{j\phi}\}$$

The spatial location of the peak value of the phase correlation function, d, corresponds to the translation misregistration between g_R and g_S. The innovation in the implementation demonstrated here is in the finding of the peak of the correlation function, d. Instead of looking for an interpolated peak of d, the center of mass of the peak of d is found. We have found that this gives a more robust result than searching explicitly for the peak.

4. New Technologies

The methods described in this section are all feature-based. Since features are more reliable than

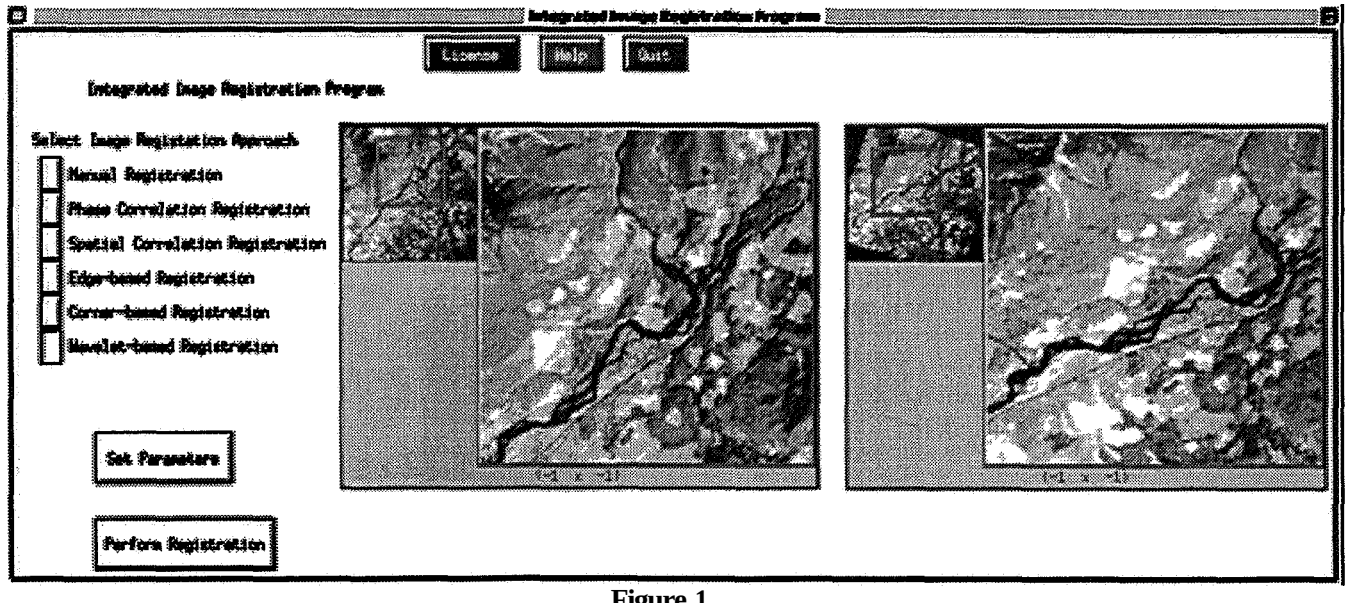


Figure 1
Khoros Graphic User Interface of the Current Registration Toolbox

intensity or radiometric values, feature-based methods are usually more accurate. The only situation where a feature-based approach will fail is when the scene is mainly homogenous or densely textural in which case a combination of feature- and intensity-based methods will yield better results. In the current feature-based methods included in the toolbox, the features correspond either to edges, comers, regions, or wavelet characteristics. Edge or edge-like features are very useful to highlight regions of interest such as coastlines [21,22]. Regions can be used either to extract contours or to use their gravity centers as control points. In more structured images containing, for example, city grids, comers might be of interest.

4.1. Edge-Based Methods

An edge detection computes the gradient of the original gray levels and highlights the pixels of the image with higher contrast. Since edge values are less affected by local variations in the intensity or time-of-day condition than original gray levels, edge features are more reliable in any registration scheme.

4.1.1. Iterative Edge Matching

This first edge-based method performs the registration in an iterative manner, first estimating independently the parameters of the deformation transformation, and then iteratively refining these parameters. For this implementation, we chose to model the transformation as a combination of a scaling in both directions (dsx,dsy), a rotation (dΘ), and a shift or translation (dtx,dty)in both directions. This algorithm is based on the assumption that rotation angle and scaling parameters are small (within 5 degrees for the rotation angle and within [0.9,1 1] for the scaling parameters). At each iteration, the five parameters are retrieved by computing the cross-correlation measures for all successive values of the parameters taken at incremental steps [23].

4.1.2. Matching of Line Segments

This algorithm is described fir the simple case of a similarity transfirmation (composition of a rotation, translation, and scaling). The input to our technique are line segments extracted from both images. The line segments are firmed from the edges using a noise-robust edge detector such as the Canny operator. Since the end points of the line segments are not reliable due to noise, low contrast, and interactions between adjacent objects, we make use of the local orientation and the intercept ofthe line segment. Line segments in both images are utilized to find the rotation component of the transformation by performing a Hough transfirm. Knowing this rotation, "virtual triangles" are firmed and utilized to find the remaining parameten ofthe transformation.

We note that the similarity transformation is a constrained version of the general affine transfirmation. The general affine transform has six parameters whereas the similarity transfirm has four parameters to be estimated. In the case of the similarity transfirm, the scaling is the same in the x and y direction whereas in the case of the affine transformation the scaling is non-uniform across the image. The difficulty with estimating the general affine parameten is that the rotation parameter cannot be easily separated as before. If the rotation cannot be separated in such a way that the peak bin of the histogram indicates the corresponding line pairs, it then becomes difficult to solve for the parameten of the generic affine model. One way to get around this difficultyis to assume that the rotation is the largest source ofalignment error (at least locally). With this assumption, it is possible to hypothesize a set of bins closer to the peak bin as the best possible matches and then use the virtual triangles to estimate the affine parameters.

4.2. Region-Based Methods

4.2.1. Region Boundary Spatial Correlation

Automatic iterative parallel region growing (IPRG,[24]) is followed by spatial correlation [15] on the region boundaries. The advantage of spatial correlation on region boundaries over spatial correlation on edge features is that the feature outlines which are generated (outlines of islands, lakes, coastlines, etc.) are *complete* outlines. In many cases edge features provide very fragmented outlines. Figure 2 shows results of the IPRG algorithm producing a segmentation of a GOES 8 image over a portion ofthe Gulf of California.

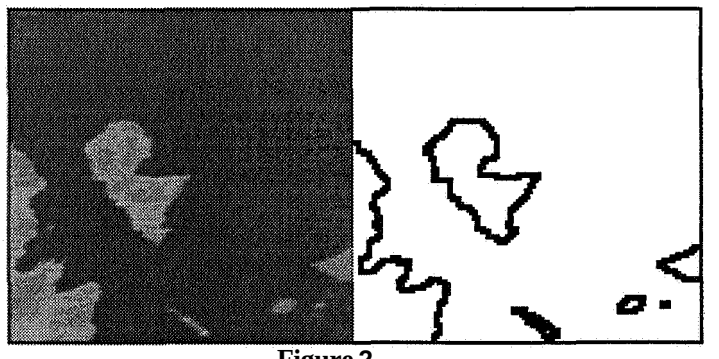


Figure 2
Original GOES8 Image Over the Gulf of California — *Results of IPRG Segmentation*

4.2.2. Moment Invariant Matching

This method uses unsupervised clustering (K-means) to find clusters within each image to be registered. The total number of clusters is user-determined but should not exceed 10 for practical computational reasons. Within each image we extract clusters of one type and treat this as a binary image. Thus we will have, for example, two images of cluster type "A", two images of cluster type "B", etc. Now within each binary image we search for enclosed clusters that have closed boundaries (for example fields, lakes, etc.). We then compute the moment

invariants of these closed objects in each image and match the clusters that have the same or similar vector of moment invariants. Objects are replaced by their centroids which form the tie points necessary to compute the warping of one image relative to the other

4.3. Wavelet-Based Methods

Similarly to a Fourier transform, wavelet transforms provide a time-frequency representation of a signal, which can be inverted for later reconstruction. However, the wavelet representation allows a better spatial localization as well as a better division of the time-frequency plane than a Fourier transform, or than a windowed Fourier transform. In a wavelet representation, the original signal is filtered by the translations and the dilations of a basic function, called the "mother wavelet" For the algorithms described below, only discrete orthonormal basis of wavelets have been considered and have been implemented by filtering the original image by a high-pass and a low-pass filter, thus in a multi-resolution fashion. At each level of decomposition, four new images are computed; each of these images is half the size of the previous original image and represents the low frequency or high frequency information of the image in the horizontal or/and the vertical directions; images LL (Low/Low), LH (Low/High), HL(High/Low), and HH (High/High) of Table 3. Starting again from the "compressed" image (or image representing the low-frequency information), the process can be iterated, thus building a hierarchy of lower and lower resolution images. Table 3 summarizes the multi-resolution decomposition.

Each of the three wavelet-based methods which are included in the toolbox represents a three-step approach to automatic registration of remote sensing imagery The first step involves the wavelet decomposition of the reference and input images to be registered. In the second step, we extract at each level of decomposition domain independent features fiom both reference and input images. Finally, we utilize these features to compute the transfirmation function. For each method, the search strategy follows the multiresolution approach provided by the wavelet decomposition.

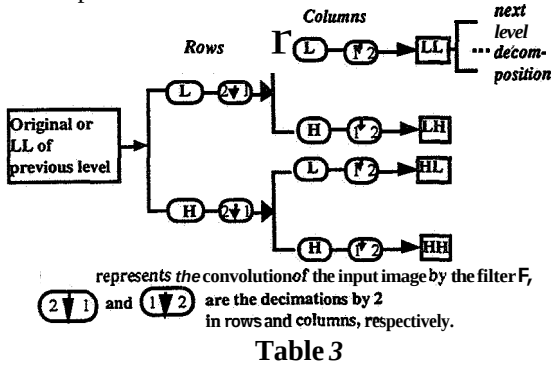


Table 3
Multi-Resolution Wavelet Decomposition

4.3.1. Wavelet Maxima Matching

The first wavelet-based registration algorithm [16,19] is based on the high-frequency infirmation extracted iiom the wavelet decomposition (images "LH" and "HL" from Table 3). Only those points whose intensities belong to the top x% of the histograms of these images are kept (x being a parameter of the program whose selection can be automatic); we call these points "maxima of the wavelet coefficients," and these maxima form the feature space.

We assume the transformation to be either a rigid or an affine transformation. Both types of transformations include compositions of translations and rotations; therefore, as a preliminary study, our search space is composed of 2-D rotations and translations, and will be extended later to rigid and affine transformations. The search is performed iteratively from the deepest level of decomposition (where the image size is the smallest), until the first top level of decomposition. At each level, the transformation is found with an accuracy A and is refined at the next level up with an accuracy $\Delta/2$. See [13] for more details on this algorithm.

4.3.2. Wavelet Edges Matching

This method utilizes the same search strategy as the one described in the previous section, but with a different feature space extracted from the low-frequency component of the wavelet decomposition (image "LL" of Table 3). In the second step, the feature points are extracted fiom the low-low (LL) components of the wavelet decompositions of the reference and input images employing the Lerner Algebraic Edge Detector (LAED, [26]). The low-low wavelet coefficients are preprocessed by the LAED and a mean square error (mse) based similarity metric computes the transfirmation hction. Currently, fir this method, we limit our search space to 2-D rotations.

The LAED is based on a highly structured and compact representation of gray-level images and is an inherently robust, generic mathematical technique which uses polynomial algebraic operations to produce gray-level image enhancement, edge detection, and gradient component information. The LAED is unique in that it is based on a global shiff operator which combines an image with displaced versions of itself and is rotationally invariant to feature orientation.

4.3.3. Wavelet Corners Matching

In this method, the comer points from the LL, HL, and HH components are extracted at each level of decomposition starting from the lowest level Then, windows of size 15x15 typically and centered at each comer point are extracted fiom the LL component of

the input image and correlations are computed with the respective LL image of the reference image. The maximum of the correlation curve corresponds to the matching comer point in the reference image. The comer point from the input image test and its matching comer point in the reference image are projected to the next higher level and these matching points are then refined by search in a 3x3 neighborhood. This process is repeated all the way to the top of the decomposition pyramid. When matching points for all comer points is performed, interpolation and smoothing techniques are utilized to obtain pixel-to-pixel correspondence between the two given images, or to compute a more general transformation (see [27] for more detail).

4.4. General Tools

4.4.1. Subpixel Accuracy Utilities

The outcome of all correlation-based methods is a set of parameters that is expressed with an *accuracy* equal to the one given by the algorithm (e.g., one pixel for shiffs or one degree for rotations). The tool described in this section will provide the means for estimating the registration coordinates with subpixel accuracy from the computed correlation coefficients. When only looking at shift and rotation, one can view the correlation function of two images to be registered as a 3-D function in x-shiff, y-shift, and rotation angle. By projecting the correlation function on each one of these dimensions and using a cubic spline interpolation, one can find the correlation peak in that dimension, with subpixel accuracy The projected peak points are then integrated to form the estimated registration position with subpixel *accuracy* Future developments will include the generalization of this method to more general transformations.

4.4.2. Statistical Robust Point Matching

This tool focuses on registration methods based on the principle of point mapping with feedback [5]. Specifically, given a set of control points in the reference image and a corresponding set of points in the sensed image within a prespecified transformation (e.g., rigid, affine), our method attempts to derive a computationally efficient algorithm to match these point patterns. In principle, our proposed method is similar to the ones pursued by [12,28,29], whereby matching is achieved via cluster detection in transformation space. Whereas all of the above methods are computationally intensive (their running time is O(N^4), where N denotes the number of control points), our proposed methodology is expected to yield much faster variants. See [30] for more detail.

5. Algorithm Intercomparison

As stated earlier, it is unlikely that a single registration technique will satisfy all different applications. Although automated registration has been developed for a few Earth science applications, there is no general scheme which would assist users in the selection of a registration tool. By providing the results of an intercomparison of multiple registration techniques, we will enable the users to choose the method which is the most appropriate for their particular application. Having all the algorithms implemented in a single toolbox will reinforce their ease of use and will provide the visualization capabilities that will facilitate this choice.

5.1. Definition of the Criteria

Defining intercomparison criteria is a relatively difficult task, since each application might have different requirements and the importance of the criteria might vary fiom one application to the next. Affer a first evaluation, the toolbox and the results of the evaluation will be made available to the scientific community From comments and feedback, new evaluation criteria as well as new registration techniques might be defined and will give rise to a new version ofthe toolbox.

The first criteria which will be implemented are the following:

- ***Accuracy***

  Several methods can be thought of to quantify the *accuracy* of a given registration method.

  •• a first method consists ofregistering the same set of data manually and automatically Then, considering the manual registration as our "ground truth", the error between manual and automatic registration characterizes the accuracy of the automatic registration.

  •• another method requires a processing which corrects for the illumination variations from two scenes. This correction would be applied after transforming back the sensed image by the computed deformation model. Then, a Mean Square Error (MSE) would be computed between transformed sensed image and reference image.

  •• if the registration is performed between a remotely sensed image and a map, the MSE can be computed on selected ground features such as coastlines.

  •• another way to quantify the accuracy of an automatic method is described in [31]; it utilizes high-resolution data such as Landsat-TM or SPOT ("Satellite Pour l'Observation de la Terre") data which are degraded to lower spatial resolution. Then the lower resolution data are registered and *accuracy* can be measured at a subpixel level using the full high-resolution data

- ***Computational Requirements***

  The computational requirements of each method will be computed from two means:

  1 the computational complexity of each algorithm will be evaluated

  2. each method will be implemented and timed on various architectures.

- ***Level of Automatization***

Given the future large amounts of data to process, automatic techniques should be as free as possible of parameters to tune. Whenever possible, thresholds or other such parameters will be computed adaptively from within the programs. If necessary, training on large numbers of data will be perbrmed and parameters will be chosen from this training.

- ***Applicability***
  This last criterion intuitively corresponds to qualitative judgments, such as "if the scene includes a city grid, a comer-based method will work faster than a region-based method." A quantitative way to evaluate the "Applicability" criterion might be statistical; a large amount of sensor data over a large variety of scenes will be gathered and the results ofthe three previous criteria will be combined to compute a probability of the applicability of an automatic registration technique given a particular dataset and particular scene contents.

In future work, we could also consider (as was proposed by Rignot [9] and Manjunath [10]) to couple the registration toolbox with a planning system which would use the above criteria to decide which algorithm to use depending on the application, the type of data, the requested *accuracy,* and the time and computational constraints. Such a planning-scheduling system has been developed at NASA/Goddard for the RVC's and is based on the work of N.M. Short and A. Lansky [32,33].

5.2. Test Datasets

The toolbox algorithms will be first evaluated on a large variety of NASA datasets. They will represent at least three main types of applications:

- ***Multi-temporal studies*** with multi-temporal datasets of one sensor over the same areas collected at different times (various times of the day, various seasons, multiple *years,*), such as AVHRR/LAC and GAC, Landsat/TM and MSS, GOES,
- ***Multi-Instrument data fusion*** with multisensor datasets representing multiple spatial, temporal and radiometric resolutions, such as AVHRR/LAC versus Landsat/TM or MSS, GOES versus Landsat/TM or AVHRR/LAC, Landsat/TM versus **SPOT,** and MAS versus Landsat/TM.
- ***Channel-to-channel co-registration*** with multiple radiometric and spatial resolutions of the different channels of one given sensor, such as GOES, MAS, and an hyperspectral instrument such as ASAS.

6. Conclusion and Future Work

A first version of the toolbox of image registration techniques has been developed. This collection of algorithms will be evaluated by utilizing datasets representative from many current and future Earth science applications. Future versions of the toolbox will include a larger variety of algorithms and the quantitative evaluation of these algorithms will be extended to other criteria, using a larger number of datasets, such as space science data, medical imagery or military applications data. Future work will also include the study of computational issues and will focus on the computational aspects and speed of processing of the proposed techniques. in specific, we will focus on the algorithm enhancement, performance evaluations, and parallel implementations of the proposed methods.

Bibliography

[1] J. Townshend et al, "The Impact of Misregistration on Change Detection," ***IEEE Tr Geosc.Rem.Sens.,*** Vol.30-5, 1992.

[2] W.J Campbell et al, "Distributed Earth Science Validation Centers for Mission to Planet **Earth,"** ***8-th Intern. Symp. Methodologies for Intelligent Systems,*** ISMIS'94, Charlotte, NC, Oct. 1994.

[3] P Van Wie, M Stein, "A Landsat Digital Image Rectification System," ***IEEE Trans. on Geosc. Electr*** , GE-15, ***3,*** 130-137, 1977

[4] B.J Devereux et al, "Geometric Correction of Airborne Scanner Imagery by Matching DelaunayTriangles," ***Int. J. Remote Sensing,*** Vol. 11, 12,2237-2251, 1990.

[5] L.Brown, "A Survey of Image Registration Techniques," ***ACM Comp.Surv.,*** Vol 24, 4, 1992.

[6] W.K. Pratt, "Correlation Techniques of Image Registration," ***IEEE Trans. Aerosp. and Electr Systems,*** Vol.AES-10, ***3,*** 1974.

[7] D.I. Barnea, and H.F Silverman, "A Class of Algorithms for Fast Digital Registration," ***IEEE Trans. on Computers,*** Vol. C-21, 179-186, 1972.

[8] Y.C Hsieh et al, "Performance Evaluation of Scene Registration and Stereo Matching for Cartographic Feature Extraction," ***IEEE Trans. on PAMI,*** Vol. 14, no. 2, 1992.

[9] E.J Rignot et al, "Automated Multisensor Registration. Requirements & Techniques," ***Photogr Engin. & Remote Sensing,*** Vol. 57, no. 8, pp. 1029-1038, 1991

[10] B.S. Manjunath, "Registration Techniques for Multisensor Sensed Imagery," ***Photog. Engin. and Remote Sensing Journal,*** 1996.

[11] J Le Moigne et al, "An Automated Parallel Image Registration Technique of Multiple Source Remote Sensing Data," CESDIS TR-96-182 - submitted to ***IEEE Trans.Geosc. and Rem. Sens.***

[12] G. Stockman et al, "Matching Images to Models for Registration and Object Detection via Clustering," ***IEEE Trans. PAMI-4, 3,*** 1982

[13] J Le Moigne, "Towards a Parallel Registration of Multiple Resolution Remote Sensing Data," *IGARSS'95,* Firenze, Italy, July 1995.

[14] C.Kuglin et al, "Map-Matching Techniques for Terminal Guidance Using Fourier Phase

Information," ***SPIE Conf. Digital Proc. of Aerial Images,*** May 1979.
[15] R. Gonzalez, P Wintz, ***Digital Image Processing,*** Addison-Wesley Co., 1987
[16] J.LeMoigne, "Parallel Registration of Multisensor Remotely Sensed Images Using Wavelet Coefficients," ***SPIE Aerosense,*** Apr. 1994.
[17] E. Kaymaz et al, "Multi-Resolution Geo-Registration of Remote Sensing Imagery Employing the Wavelet Transform," ***SPIE 25th AIPR Workshop,*** October 1996.
[18] J.C. Tilton, "Hybrid Image Segmentation for Earth Sensing Remote Sensing Data Analysis," *IGARSS* Nebraska, May 1996.
[19] S. Raghavan et al, "Real-Time Extraction of Meta-Data from Multi-Source Imagery," ***Phase II SBIR Report,*** 1996.
[20] M McGuire, H. Stone, "Techniques for Multiresolution Image Registration in the Presence of Occlusions," ***IRW97,*** NASA-GSFC, Nov 97
[21] B. Blancke et al, "The Aerospatiale Meteosat Image Processing System (AMPS)," I-st ***Intern. Symp. Scientific Imagery and Im. Proc.,*** Cannes, France, Apr 1995,
[22] J.C. Tilton, "Comparison of Registration Techniques for GOES Visible Imagery Data," *IRW97,* NASA-GSFC, Nov 97
[23] J LeMoigne, N. El-Saleous, E. Vermote, "Iterative Edge- and Wavelet-Based Image Registration of AVHRR and GOES Satellite Imagery," *IRW97,* NASA-GSFC, Nov 97
[24] J C. Tilton, "Image Segmentation by Iterative Parallel Region Growing and Splitting," ***IGARSS*** Vancouver, July 1989.
[25] R.F Cromp, and W J. Campbell, "Data Mining of Multidimensional Remotely Sensed Images," ***2nd Int. Conf. Information and Knowledge Management,*** Nov 1993.
[26] B.Lerner, "Gray-Level Images. *An* Algebraic Approach through Compact Polynomials," ***IEEE-SMC Conf.,*** 1988.
[27] M. Mareboyana and J. Le Moigne, "Finding Corner-Point Correspondence from Wavelet Decomposition of Image Data," *IRW97,* NASA-GSFC, Nov 97
[28] A. Goshtasby et al, "A Region-Based Approach to Digital Image Registration with Subpixel Accuracy," ***IEEE Trans. Geosc. and Remote Sensing,*** Vol. GE-24, 3, May 1986.
[29] A. Goshtasby, "Registration of Images with Geometric Distortions," ***IEEE Trans. Geosc. and Rem. Sens.,*** Vol. 26, 1, pp. 60-64, 1988.
[30] D. Mount, N Netanyahu, J. LeMoigne, "An Efficient Algorithm for Robust Feature Matching", *IRW97,* NASA-GSFC, Nov 97
[31] A.P.Cracknell, and K.Paithoonwattanakij, "Pixel and Sub-pixel Accuracy in Geometrical Correction of AVHRR Imagery," ***Intern. Journ. Rem. Sensing,*** Vol. 10, nos. 4,5, 661-667, 1989.
[32] M. Boddy et al, "Planning for Image Processing," NASA/*Goddard Conf. Space Appl. Artificial Intelligence,* May 1994.
[33] A.L. Lansky et al, "The COLLAGE-KHOROS **Link:** Planning for Image Processing Tasks," ***AAAI Spring Symposium on Integrated Planning Applications,*** Stanford Univ., March 1995.

Aerial Image Registration by PFANN (Point Feature & Artificial Neural Network) Matching

Jiegu Li, **Dan Kong,** Zhebin **Qian**
Institute **of** Image Processing & Pattern Recognition
Shanghai Jiao **Tong** University
Shanghai 200030, **China**

Abstract — This paper presents **an** image registration method based on a two-dimensional Hopfield Neural Network, Where the problem of image matching is treated with the minimization of the energy function of the Hopfield neural network. The method was originally proposed by N **M** Nasrabadi (R.1), but, in order to make it more practical, four items are studied in this paper, they are: the modification of the compability measure in the energy function for obtaining a more robust result, the evaluation of the point feature extraction methods for selecting a more effective one, the "blocking" algorithm of expediating the matching procedure and the subpixel matching for obtaining a more precise object location Theoretical explanations & laboratory results of these four parts are presented

1 Introduction

In many visual intelligence systems, such **as** robotics, crimi**nal** recognition & autonomous navigation, etc, Image registration techniques **are** used to recognize and locate the object by matching the object **(O)** image acquired at realtime **with** the fiducial (F) image stored in the system originally. Usually, the F image occupies a greater pixel area than the **O** image. In order to perform this kind of work, these have been employed different methods based on statistical studies, such **as** calculating the sum of differences of the pixel gray values of two images or the maximum correlation of their pixel gray values, But in **some** actual cases, the **O** image acquired at realtime may be distorted by many factors, such **as** resolution difference, geometric error, different emission angles of the light source, strong noise, etc, which may exert tremendous influence on the gray value distribution of the **O** image, and thus may lead to mismatching.

In overcoming the disturbances stated above, the registration system constructed on the basis of image feature matching is more stable than statistical coniparison of the gray value distributions. Here, corner points are taken **as** the image features to be matched. Hence, the first problem (Section **II** of the paper) *is* to find an effective method to extract the corner point feature. The third section **gives an** experimental evaluation of the methods selected, the evaluation is based on applying the selected methods to specially designed image patterns, where corner with different angles are known previously.

Although after evaluation we can find a best way for corner point extraction, yet actually none of the feature extraction algorithms now available is capable of avoiding data missing or false alarming, especially under the influences of the distortion factors stated before. Therefore matching work has to be done with two gaphs where the point distributions are not exactly the same, that means, an "elastic" matching method is needed, for which a **2D** Hopfield N **N** (R.1) is **an** ideal one Section **III** describes the basic theory of this N **N** , but here comparisons are made not only on the geomatric locations between pomt features of the two images (as proposed in R.1) but on the gray gradient surrounding the points to be matched, as is suggested in our modification The factor added shows a more robust character on overcoming the distortion caused by the resolution differences & gtometric errors.

Section IV gives a fast matching algorithm of the **2D** Hopfield N N which is based **on** blocking the F image into subimages and doing matching work hierarchically. For the fast algorithm, the matching time is reduced to one–fifth of the original one.

Section V studies the possibility of giving the matching with a sub-pixel accuracy which is accomplished by establishing a continous model for the discrete **F** image edge features and calculating the high order moments to relocate the corner points precisely Some synthesized images are used to check the accuracy of the method proposed.

2 An evalution on the methods used for image corner detection (R.2)

Considering from the view-point *of* simplicity applicability, seven methods are selected, they are:

1 IO method (R.1), whch detects the corner points with a statistical study on the directions of horizontal, vertical, 135° and 45° within a certain size of image window

2. Median filter mehtod (R.3), whch finds the corner points by counting the difference between the filtered original images.

3. Distance evaluation method (R.4), which comes from the definition "corner points are the intersection points of two or more edges"

4. KR method (R.5), which is based on the calculation of the second order derivatives of the gray pictures.

5. DET method (R.6), which is based on the calculation of the Gaussian curvature of the gray pictures.

6. DN method (R.7), which is a modification of DET method.

7 A combination of the **10** & Median Filter methods

Specially designed images (sized **320** x **240)** with geometrical patterns (polygons & circle sectors) possessing different corner angles (varied for **15° ~ 345°) &** gray levels are used to evaluate the seven methods. **Also,** gaussion noises with different strengths are added to the testing images.

In the evaluation procedure, six parameters are used, **they** are:

1 Sensitivity S,
S = RP/AP, where
RP — number of right corners detected by the methods,
AP — corner numbers in the images
2. Efficiency E,
E = RP/SP, where
SP — totalnumber of the detected corner
3. Robustness **R** (against the noise),
4. Location errors **e** of the detected corners, where
e_{max} — maximum error of the detected corner,
Σe^2 — sum *of* the error square of all the detected corners
5. Rightness **A, which is** defined as
A = S x E
6 Execution time T

To save space, only one example of the laboratory results is shown in table 1 The final grades, marked as G (good), NB (not bad) and NR (not recommended) are shown in table **2**

3 The 2-D Hopfield NN Method (R.1)(R.9)

The point correspondence problem can be formulated as an optimization task where a cost function (energy equation) representing the constraints on the solution is minimized And, the minimization problem can be mapped onto a **2D-**Hopfield N **N** such that the **cost** energy function is the same as the Lyapunov function of the network, with the synaptic interconnection weights between the neurons representing the constraints imposed by the correspondence problem The Lyapunov function represents the collective behavior of the network When the network is at its stable state the energy function is said to be at its local minimum The basic principle of the neural network is to make a cooperative decision based on the simultaneous input *of* **a** whole community of neurons in which each neuron receives information through giving information to every other neuron. This information is used by each neuron to force the network **to** converge to a stable state in order to **make** a decision.

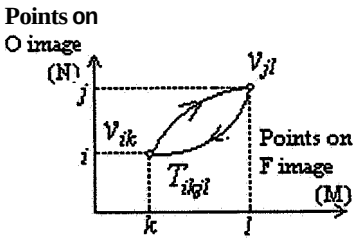


Figure 1 **2D** Hopfield N.N

A schematic diagram of a **2D** Hopfild N N is shown in Fig.1, where *ij* and *kl* are pairs of points **to** be matched on **O** (object) image & F (fiducial) image Neurons are located on the joint points of the corner points of two Images, such as V_{ik} and V_{jl} and T_{ikjl} represents the weight between V_{ik} & V_{jl} The matchmg problem is to minimize a Hopfield style energy function given by

$$E = -\frac{1}{2}\sum_{i=1}^{M}\sum_{k=1}^{N}\sum_{j=1}^{M}\sum_{l=1}^{N} T_{ikjl}V_{ik}V_{jl} + \sum_{i=1}^{M}\sum_{k=1}^{N} I_{ik}V_{ik} \quad (1)$$

where ***M*** x N is the network size,

$$T_{ikjl} = AC_{ikjl} - B\delta_{ij} - \quad (2)$$

and

$$I_{ik} = 2B \quad (3)$$

The expression for the compatibility measure C_{ikjl} is given **by**

$$C_{ikjl} = \frac{2}{1 + e^{\lambda(X-\theta)}} - 1 \quad (4)$$

Where θ is **a** threshold value, X in R.1 is only a distance measure countmg the geometric differences between the pomts (k,l) and the points (j,i). The X used now is a mixed measure of distance & gradient differences (R.9), where

$$X = w_1\Delta D + w_2\Delta I \quad (5)$$

where w_1, w_2 are the weights, AD the distance difference and AI the gradient difference

The AI factor makes the matching result more robust than when it is absent, as in R.1 In our experiment, when the O image has a rotation angle of 30° and a scale amplification of **1.2,** a correct matching **result** can still be obtained.

4 A Fast Block-Based Algorithm (R.8)

The F image is much lager in size than the **O** image with the same definition, that means the diagonal of the **F** image is much longer than that of the **O** image. From the distance measure ***AD*** in expression ***X***, it can be seen that the matching process for the pair of points (k,l) departing farther than the diagonal of the **O** image is redundant, it not only increases the processing time but also produces mismatching. Thus, a fast improved algorithm based on blocking is put forward here The algorithm can be described as follows:

a) Accortling to the size of the **O** image, the F image is divided into several consecutive blocks. Suppose the size of the F image is X_n x Y_n, and that of the **O** iamge is X_m x Y_m, then a wmdow sized $(2X_m - 1) \times (2Y_m - 1)$ sliding on the F image with the step width $(X_m \times Y_m)$ will contain the complete contents of the **O** image at a certain step (this is a correspondence position). In practice, the window or block size is designed as

$$\{2 \times [X_n/(X_n/X_m)] - 1\} \times \{2 \times [Y_n/(Y_n/Y_m)] - 1\} \quad (6)$$

and the consecutive blocks are overlapped nearly one half of each other

b) Each block of the F image is matched with the **O** image by the **2D** Hopfield NN

c) **For** the **2D** Hopfield NN, under any circumstance, the system has a minimum energy stable state, thus decision should be made to find the matched block from the whole F image The criterion consists of two factors, the number of the matched points between two images & the memberships of the nearest-neighbor clusters, for which distance deviations are used **to** cluster the matched pomts.

Although adjacent windows are overlapped nearly one half of each other, the computation time reduces dramatically, the reason is that the comptitational complexity of

Table 1: Recognition possibility of different corner angles

Y - can be recognized;
N - can't be recognized.

Methods used	Angle values									
	15°	30°	45°	60°–120°	135°	150°	165°	195°	210°	225°
1	Y	Y	Y		Y	Y	N	N	Y	Y
2	Y	Y	Y		Y	N	N	N	N	Y
3	Y	Y	Y	Y	N	N	N	N	N	N
4	Y	Y	Y		Y	Y	N	N	Y	Y
5	N	N	Y		Y	Y	N	N	Y	Y
6	N	N	N		Y	Y	N	N	Y	Y
7	N	Y	Y		Y	Y	N	N	Y	Y

Table 2: Evalutation Results

Methods used	Parameters					
	S	R	e	A	T	overall
1	G	NB	G	NB	G	G
2	G	NR	NB	NR	G	Nb
3	NR	NB	NR	NR	NR	NR
4	G	G	G	NG	NB	G
5	NB	G	NR	G	NB	NB
6	NR	G	G	G	NB	G
7	NB	NB	NB	G	NB	NB

the neural network is nonlinear, when the number of neurons decreases, the number of iterations decreases with a nonlinear relationship. In our experiment, with the sizes of the O images & F images as *36 x 36* and *160 x* 160 respectively (one hundred pairs of O images & F images are taken as test *images),* after blocking the F images into *3 x 3* blocks, the average (of the 100 pairs) iteration loops decrease from 11 to 1.65 and the average processing time decreases from 261 sec to 57 sec The result shows the effectiveness of the blocking algorithm.

5 Sup-pixel localization

The sub-pixel localization is performed through using a sub-pixel edge detection method to relocate the feature points in the F image with sub-pixel precision For a particular shape (a projection of some objects, for instance, a building) in the F image, the sub-pixel relocalization procedure consists of three consecutive steps, they are edge detecion, corner point joining & then shape center finding. The sub-pixel edge detection, as the unportant theoretical background of this section, is based on an ideal continuous edge model (R.10) in a unit circle (Fig 2), where four paremeters I, θ, h, lt determine the unique straight *edge*, in which h & k are background gray-level & gray level contrast respectively, and l is the vertical distance from the circle center to the straight edge, I $\in [-1, 1]$, θ is the angle between the normal of edge & X axis, $\theta \in [-\pi/2, \pi/2]$

Let $f(x, y)$ represent image data, its 2D spatial moments are as follows:

$$M_{pq} = \int\int x^p y^q f(x, y)\, dy dx \qquad (7)$$

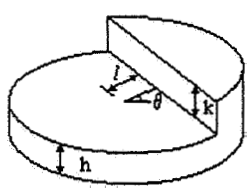


Figure 2: 2D continuous edge model

In order to reduce the dimensionality of the edge problem, rotate the window by $-\theta$, and the solution of I, h, lt will be independent of θ Then rotated moments

$$M'_{pq} = \sum_{r=0}^{p}\sum_{s=0}^{q} \binom{p}{r}\binom{q}{s} (-1)^{q-s} (\cos\theta)^{p-r+s}$$

$$(-\sin\theta)^{q+r-s} M_{p+q-r-s,r+s} \qquad (8)$$

Because the rotated edge is perpendicular to X axis, that is

$$M_{01} = 0 \qquad (9)$$

So,

$$\theta = \tan^{-1}\frac{M_{01}}{M_{10}} \qquad (10)$$

Then, the other parameters can be worked out from M'_{pq}

$$l = \frac{4M_{20} - M_{00}}{3M_{10}} \qquad (11)$$

$$k = \frac{3M_{10}}{2\sqrt{(1-l^2)^3}} \quad (12)$$

$$h = \frac{1}{2\pi}[2M_{00} - k(\pi - 2\sin^{-1} l - 2l\sqrt{1-l^2})] \quad (13)$$

It must be pointed out that if pixel sampling makes a gray-level of $h + \delta k$ on the edge pixel in the image, two level continuous model will definitely introduce errors. To improve precision, a bias table is used to correct the error since we have previously synthesized edges of all kinds of parameters, and recorded the biases between accurate values and estimated values, we can correct each calculated result through checking with the table and through interpolation.

An integrated test platform with carefully designed synthesize images is constructed to evalute the method. Synthesized images are of size 75 x 75, background gray-level 100, gray-level contrast 100. The objects are equal-sided heptagons whose edge location is known at the scale of high resolution. The tests consist of three parts.

(1) Measurements on the location of detected features. Statistic results are shown in Table 3 (X direction error measured in pixel)

(2) Measurements on the object outlines. Image center is taken as the origin of coordinate system, edges are represented by polar coordinates. The unit of ρ is pixel and the unit of θ is degree. Fitting results of seven edges are shown in Table 4.

(3) Measurements on object comer locations. Locating results of seven corners in pixel units are shown in Table 5.

We also synthesized many equal-sided polygons of arbitrary sire, initial angle and number of sides. All results are feasible, even with the adding of gaussion noise

After evaluation, the method is applied to a pair of real images (Fig 3). The corner points extracted & matched result are shown in Fig 4

The method is used to relocate the initial coordinates of the points in the window (Fig.4). After estimating the biases between precise locations and initial locations of all comers, doing weighted average to biases in reverse proportion to distance from every corner to object center, average results are used as the center bias to correct matched result. (shown as Table 6)

6 Acknowledgement

The authors would like to thank Dr Yu-ming Zhao & Mr Xu-tong Zhou for providing us some of their working results. Thanks also to the Chinese National Science Foundation for their support.

References

[1] N M Nasrabadi & C. Y Choo Hopfield Network for stereo vision correspondence, IEEE Transaction on Neural Networks Jan 1992, pp 5-13

[2] Xu-Tong Zhou A study on the methods of corner recognition, Technical paper of the Image Processing & Pattern Recognition Institute, Shanghai Jiao-tong University 1994

[3] E. R. Davies: Machine Vision - Theory, Algorithm & Practicalities, Academic Press, 1990.

[4] Z. H. Huang & Q. Yu A Direct Corner Detecting Algorithm, Int con on Pattern Recognition, 1986, pp853.

[5] L. Kitchen & A Rosenfeld Gray-level Corner Detection, Pattern Recognition Letters, Dec 1982pp95-102.

[6] P R. Beaudet: Rotationally Invariant Image Operators, Int. J Conf. on P.R. 1987, pp579.

[7] L Dreshler & H H Nagel: Volumetric Model & 3D-Trajectory of a Moving Car Derived from Monocular TV-Frame Sequences of a street Scene, Proc. Int J. Conf. on AI 1981, pp692.

[8] Zhe-bin Qian, Jie-gu Li: Use of Hopfield Neural Network for complex image Registration, Proceedings of ICTAI'97 (to appear).

[9] Y M Zhao, J G Li: Neural Network Matching Algorithm Using Interesting Points (in Chinese), Pattern Recognition & Artificial Intelligence Dec 1995pp 313-319

[10] E P Lyvers, O R. Mitchell, M L. Akey & A P Reeves Subpixel measurement Using a Moment based Edge Operator, IEEE Trans. on PAMI No. 12 1989, pp 1293-1309.

Table 3: Results of feature detection test

Max Abs. Err	Min Abs. Err	Mean Err	Err Stand. Vari.
0.3763	0.0035	0.0324	0.1922

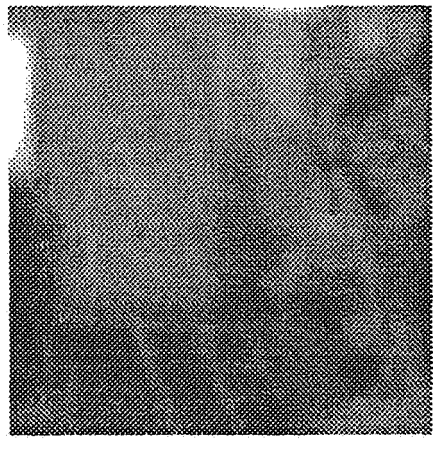
Table 4: Results of object outline fitting test

	Origin ρ	Origin θ	Estimated ρ	Estimated θ	Error of ρ	Error of θ
1	24.3262	10.0000	24,4030	10.0694	-0.0768	-0.0694
2	24.3262	61.4286	24.4171	61.7147	-0.0909	-0.2861
3	24.3262	-67 1429	24.4083	-67.0526	-0.0821	-0.0903
4	24.3262	-15.7143	24.3999	-16.0445	-0.0737	0.3302
5	24.3262	35.7143	24.4176	35.2036	-0.0914	0.5107
6	24.3262	87 1429	24.3840	86.7213	-0.0578	0.4236
7	24.3262	-41.4286	24.3778	-41.3627	-0.0516	-0.0659

Table 5: Results of object corner locations

	Origin coordinate	Estimated coordinate	Coordinate variance	Variance of distance
1	(21.2389, 58.9223)	(21.1986, 58.8993)	(0.0403, 0.0230)	0.0464
2	(10.0336, 38.3458)	(9.9955, 38.4064)	(0.0381, -0.0606)	0.0716
3	(19.1345, 16.7559)	(19.0580, 16.7676)	(0.0565, -0.0117)	0.0577
4	(41.6885, 10.4102)	(41.6644, 10.4166)	(0.0241, -0.0064)	0.0249
5	(60.7120, 24.0871)	(60.6730, 24.0274)	(0.0390, 0.0597)	0.0713
6	(61.8799, 47.4877)	(61.8359, 47.6132)	(0.0440, -0.1255)	0.1330
7	(44.3127, 62.9909)	(44.3714, 62.9901)	(-0.0587, 0.0008)	0.0587

O image (36x36)
origin at (460,794)



F image (160x160)
origin at (347,728)

Figure 3: Natural images

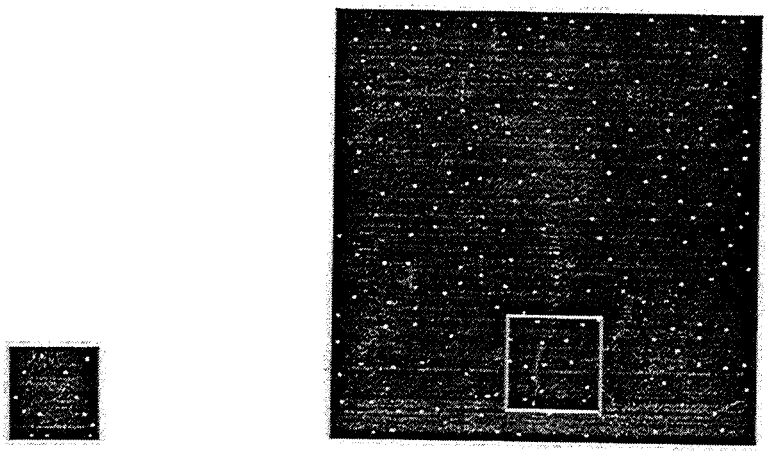


Figure 4 Corner points and matched result (the square window)

Table 6: Quantitative results of relocation

Feature points	O image coord	F image init. coord.	F image sub-pixel coord.	Bias	Weight values
1	(4, 13)	(115, 77)	(115.31, 77.33)	(0.31,0.33)	0.0607
2	(5, 30)	(116, 94)	(114.57, 91.83)	(-1.43, -2.17)	0.0510
3	(10, 6)	(122, 67)	(121.77, 67,93)	(-0.23, -0.93)	0.0595
4	(10, 22)	(122, 88)	(121.68, 88.06)	(-0.32, 0.06)	0.0894
5	(19, 3)	(130, 67)	(134.63, 68.66)	(4.63, 1.66)	0.0700
6	(19, 26)	(130, 90)	(130.27, 89.44)	(0.27, -0.56)	0.1137
7	(24, 6)	(132, 73)	(133.34, 74,79)	(1 34, 1 79)	0.1202
8	(25, 12)	(136, 76)	(135.62, 76.14)	(-0.38, 0.14)	0.1368
9	(25, 29)	(139, 96)	(138.12,95.10)	(-0.88, -0.90)	0.0595
10	(33, 9)	(144, 73)	(144.37, 72.80)	(0.37,-0.20)	0.0647
11	(33, 15)	(141, 79)	(143.44, 80.77)	(2.44, 1 77)	0.0968
12	(33, 31)	(144, 95)	(144.57, 95.60)	(0.57, 0.60)	0.0527
Center of the shape	(20.00, 16.83)	(130.92, 81.25)	(131.54, 81.65)		

Clinical Relevance of Fully Automated Multimodality Image Registration by Maximization of Mutual Information

Frederik Maes*, Dirk Vandermeulen, Guy Marchal, Paul Suetens,
Laboratory for Medical Imaging Research†
Department of Electrical Engineering, ESAT‡ and
Department of Radiology, University Hospital Gasthuisberg§
Katholieke Universiteit Leuven, Belgium.

Address for correspondence:
Laboratory for Medical Imaging Research (ESAT/Radiologie)
Departement Radiologie, UZ Gasthuisberg
Herestraat 49, B-3000 Leuven, Belgium
Email Frederik.Maes@uz.kuleuven.ac.be

Abstract

We have developed a very general and powerful algorithm for fully automated affine geometric registration of three-dimensional (3D) multimodality images based on maximization of mutual information of the gray-values of corresponding voxels. The robustness and subvoxel accuracy of the method has been validated extensively for matching of CT, MR and PET brain images. In this paper, we illustrate the clinical usefulness of the approach by showing results for various clinical applications, such as matching of CT and PET images of the thorax and of CT and MR images of the prostate. Our results demonstate that the generality and robustness of the method and the fact that it does not require any pre-processing of the images or user interaction make it a useful tool in routine clinical practice.

1 Introduction

In medical diagnosis, treatment planning and therapy follow-up, clinical decisions are often based on information obtained from various imaging modalities, such as computed tomography (CT), magnetic resonance (MR) and positron emission tomography (PET) Images of different modalities often contain complementary information and in many applications clinicians benefit from joint visualization and manipulation of multimodal images Combined analysis of multiple images requires these images to be geometrically aligned. Multimodality image registration is therefore a fundamental task in medical image processing.

Recently, we have proposed to use the concept of mutual information (MI), as defined in information theory, as a new measure for automated image matching [1, 4]. The MI registration criterion states that the mutual information between the image intensity values of corresponding voxels in the images to be registered is maximal at registration. Maximization of mutual information (MMI) is a very general and powerful approach that can be applied automatically without image preprocessing or user interaction on a large variety of applications.

The MMI method has been evaluated extensively for matching of CT, MR and PET images of the brain. The subvoxel accuracy of the criterion was validated by comparison with the external marker-based registration result and a comparitive validation study found the performance of the method to be superior than that of most other state-of-the-art registration algorithms [8]. The robustness of the approach has been evaluated with respect to image degradations, such as noise, intensity inhomogeneity and geometric distortion, and with respect to implementation related issues, such as optimization, interpolation and sampling [4].

In this paper, we demonstrate the clinical relevance and usefulness of the MMI method by showing results for various clinical applications, involving CT, MR and PET data of the brain, the thorax and the abdomen. In all cases, the same algorithm was used without tuning it to any specific application. All results were visually inspected by radiologists and it was found that the method routinely produces clinically acceptable results on clinical patient data.

In section 2 we shortly describe the MMI algorithm. A more extensive descnption can be found in [4]. In sections 3, 4, 5 and 6 we show results for matching of CT, MR and PET brain images, matching of CT and MR prostate images, matching of CT and PET thorax images and matching of CT pre- and post-operative images of the chin. Section 7 summarizes our current findings and gives some directions for further work.

2 The mutual information criterion

2 1 Theory

The mutual information $I(A, B)$ of two random variables A and B measures the degree of dependence of A and B by means of the Kullback-Leibler distance [2] between their joint distribution $p_{AB}(a,b)$ and the product $p_A(a).p_B(b)$ of

*Frederik Maes is Aspirant of the Belgian National Fund for Scientific Research (NFWO).
†Directors: André Oosterlinck & Albert L Baert
‡Kardinaal Mercierlaan 94, B-3001 Heverlee, Belgium.
§Herestraat 49, B-3000 Leuven, Belgium.

their marginal distributions:

$$I(A,B) = \sum_{a,b} p_{AB}(a,b) \log \frac{p_{AB}(a,b)}{p_A(a).p_B(b)} \quad (1)$$

Mutual information is a measure of statistical dependence between two random variables. If A and B are related by a one-to-one mapping T such that $p_A(a) = p_B(T(a)) = p_{AB}(a,T(a))$, $I(A,B)$ is maximal and equal to the entropy $H(A)$ of A, while $I(A,B)$ is zero if A and B are independent (i.e. $p_{AB}(a,b) = p_A(a).p_B(b)$).

The image intensity values a and b of a pair of corresponding voxels in the two images that are to be registered can be considered to be random variables A and B respectively and estimates for their joint and marginal distributions can be obtained from the joint histogram of the overlapping volume of both images. a and b are related through the geometric transformation $\mathbf{T}_\alpha$ defmed by the registration parameters α. The MI registration criterion states that the images are geometrically aligned by the transformation $\mathbf{T}_{\alpha^*}$ for which $I(A,B)$ is maximal. This criterion assumes that the intensities of two images viewing the same scene are statistically related and that this relationship is strongest when the images are properly aligned The nature of the relationship between the image intensities of corresponding voxels is of course highly data dependent. The mutual information criterion, however, does not make limiting assumptions regarding this relationship and does not impose constraints on the image content of the modalities involved.

2.2 Implementation

Let $(f(s), r(T_\alpha s))$ be a pair of image intensity values in the images F and R to be registered at corresponding sites s in F and $\mathbf{T}_\alpha s$ in R with $\mathbf{T}_\alpha$ a geometric transformation with parameters α. Let $p_F(f)$, $p_R(r)$ and $p_{FR}(f,r)$ be the marginal and joint probability distributions of $f(s)$ and $r(\mathbf{T}_\alpha s)$. The mutual information registration criterion can then be summarized by the following equations:

$$I(\alpha) = \sum_{f,r} p_{FR}(f,r) \log \frac{p_{FR}(f,r)}{p_F(f)p_R(r)} \quad (2)$$

$$\alpha^* = \arg \max_{\alpha} I(\alpha) \quad (3)$$

with $I(\alpha)$ the mutual information of F and R at registration position α and α^* the position of optimal alignment at which $I(\alpha)$ is maximal.

Implementation of this criterion requires estimations of the intensity distributions to compute $I(\alpha)$ One of the images is selected to be the *floating* image F from which samples $s \in S$ with intensity $f(s)$ are taken and transformed into the *reference* image R by the transformation $\mathbf{T}_\alpha$ The set of sample points S usually is the set of grid points of F, although subsampling of the floating image can be used to increase speed performance. For each value of the registration parameter α only those samples $s \in S_\alpha \subset S$ are retained for which $\mathbf{T}_\alpha s$ falls inside the volume of R.

The joint image intensity histogram $h(f,r)$ is constructed by binning the image intensity pairs $(f(s), r(\mathbf{T}_\alpha s))$. In order to do this efficiently, the floating and the reference image intensities are fist linearly rescaled to the range $[1, n_F - 1]$ and $[1, n_R - 1]$ respectively, $n_F \times n_R$ being the total number of bins in the joint histogram. Typically, we use $n_F = n_R = 256$. Normalization of $h(f,r)$ yields estimations for the marginal and joint image intensity distributions $p_F(f)$, $p_R(r)$ and $p_{FR}(f,r)$ of the overlapping volume of the images:

$$p_{FR}(f,r) = \frac{h(f,r)}{\sum_{f,r} h(f,r)} \quad (4)$$

$$p_F(f) = \sum_r p_{FR}(f,r) \quad (5)$$

$$p_R(r) = \sum_f p_{FR}(f,r) \quad (6)$$

In general, $\mathbf{T}_\alpha s$ will not coincide with a grid point of R and interpolation of the reference image is needed to obtain the image intensity value $r(\mathbf{T}_\alpha s)$ at the position $\mathbf{T}_\alpha s$ in the reference image. *Nearest nezghbor* (NN) interpolation is generally insufficient to guarantee subvoxel accuracy, as it is insensitive to translations up to 1 voxel Other interpolation methods, such as *trilinear* (TRI) interpolation, may introduce new intensity values which are originally not present in the reference image, leading to unpredictable changes in the marginal distribution $p_R(r)$ for small variations of α. To avoid this problem, we have proposed to use *trilinear partial volume distribution* (PV) interpolation to update the joint histogram for each voxel pair $(s, \mathbf{T}_\alpha s)$ [4] Instead of interpolating new intensity values in R, the contribution of the image intensity $f(s)$ of the sample s of F to the joint histogram is distributed over the intensity values of all 8 nearest neighbors of $\mathbf{T}_\alpha s$ on the grid of R, using the same weights as for trilinear interpolation. Each entry in the joint histogram is then the sum of smoothly varying fractions of 1, such that the histogram changes smoothly as α is varied

The optimal registration parameters α^* are found by maximization of $I(\alpha)$ using Powell's direction set method [5]. The images are positioned initially such that their centers coincide and that the floating and reference image coordinate axes corresponding to the same patient axis (i.e. right to left, anterior to posterior, and inferior to superior) are properly aligned, which assumes that the orientation of each of the images with respect to the patient is known.

The algorithm has been implemented on an IBM RS6000 workstation (58 MHz, 185 SPECfp92, AIX 4.1.3). The complexity of one evaluation of the MI registration criterion was found to vary linearly with the number of samples taken from the floating image. While trilinear and partial volume interpolation have about the same complexity, nearest neighbor interpolation is about three times as efficient (1.4 vs. 0.5 CPU seconds per million samples for one evaluation of the criterion). The number of evaluations performed during optimization depends on the initial orientation of the images and on the convergence parameters specified for the Powell algorithm and typically varies between 300 and 600 for the experiments described in this paper

3 Matching of CT, MR and PET images of the brain

The MMI method has been extensively validated for rigid body and affine registration matching of CT, MR and PET images of the brain. We have applied the method successfully on a large number of cases (about 500) involving low and high resolution MR images (5 to 1 mm slice thickness) of multiple orientations (axial, coronal and sagittal) and acquired with various imaging protocols (T1 and T2 weighted, proton density, gadolinium enhanced), high resolution MRA images, conventional and spiral CT and CTA images, PET FDG and H20 images and SPECT images. We have also applied the method to register MR time series images to monitor the variation of MS lesions over time and to register MR images of different individuals to correlate fMRI

brain activity measurements. In all cases the method is applied completely automatically without user-intervention and does not require any pre-processing of the images. The method was found to be largely insensitive to local patient-related deformations, such as induced by surgery or lesion evolution over time.

The accuracy of the method has been evaluated by comparison with the external marker-based reference solution as part of the Retrospective Registration Evaluation Project conducted by J.M Fitzpatrick of Vanderbilt University [3, 8]. The MI criterion was found to be able to consistently and robustly match CT to MR and PET to MR images with subvoxel accuracy and the performance of the method was superior to that of most other state-of-the-art registration methods that were evaluated in this study. These findings have been confirmed by other groups [6, 7]

Some results are shown in figures 1 and 2 for CT/MR and PET/MR matching. Figure 1 demonstrates the joint display of the registered and reformatted images to visually inspect registration correctness using a linked cursor In figure 2 orthogonal slices through the registered and reformatted volumes are used to inspect image alignment along all three axes.

4 Matching of CT and MR images of the prostate

Planning of radiotherapy treatment planning of prostate cancer is usually performed on CT images. However, there is some indication that the prostate, and especially its apex, can be more reliably and accurately delineated on MR images An initial study was set up by the Radiotherapy Department of the UZ Gasthuisberg, Leuven, Belgium, in collaboration with the Radiology Department of the same hospital, to establish the usefulness of additional MRI for radiotherapy planning

Spiral CT images (512 x 512 matrix, 50 slices, 0.7 mm pixel size, 5 mm slice thickness) are acquired in radiation position using routine imaging protocols on a GE scanner MR images (512 x 512, 22 slices, 0.7 mm pixel size, 5 mm slice thickness) are acquired on a Siemens scanner using a standard prostate imaging protocol (interleaved turbo spin-echo sequence, TE = 120 ms, TR = 6 s, total acquisition time = 1 min 45 s) The time between both acquisitions is kept at a minimum, such that intestinal and bladder filling differences are small. The images are aligned with a rigid-body geometrical transformation using the mutual information matching criterion After registration, the images are reformatted and visualized together and the registration result is inspected by a radiologist.

The procedure has been applied to 15 cases so far After having estimated an appropriate initialization for the vertical translational parameter (which was the same in all cases), the registration was performed completely automatically in all cases. Average registration time was about 5 minutes on an IBM RS6000 workstation. In all cases, the result was judged to be excellent by radiologists. A typical result is shown in figure 3. The quality of the registration result can be evaluated by inspecting the alignment of Corresponding rigid structures in both images, such as the pelvic bone. Differences in the positioning of the patient in both scanners may induce large non-rigid deformations of soft tissues such as skin, fat and muscles. However, the MMI method seems to cope well with such local distortions and to align the corresponding rigid structures without being confused by the presence of structures that can not be aligned by a rigid-body transformation.

Once the images have been registered and reformatted, delineations of the prostate in both images separately can be easily compared. As is clear from figure 3d, prostate contours defined by the same clinician in either image may show significant differences. This is due to the different appearance of the prostate in both images. Especially on CT, it is difficult to distinguish the prostate from surrounding tissues such as blood vessels. Quantitiative evaluation of these differences and the evaluation of the inpact on radiation therapy are currently being performed.

5 Matching of PET and CT images of the thorax

At the Nuclear Medicine Department of the UZ Gasthuisberg, Leuven, Belgium, and in collaboration with the Radiology Department of the same hospital, a study was set up to investigate the benefit of using FDG-PET images in conjunction with spiral CT images for detection and staging of metastatic mediastinal lymph nodes (MLN). FDG-PET is more accurate for detecting MLN than CT thorax, but localisation of lesions, essential for adequate staging, is difficult due to the little anatomic references on PET images.

Registration of the PET and CT images allows to combine the information contained in both modalities, i.e. to detect the metastatic nodes in PET and locate them anatomically in CT

Registration of PET and CT images of the thorax was performed for 50 patients with non-small cell lung cancer Spiral CT images were acquired in a single spiral on a GE high speed scanner (512 x 512 matrix, 50 slices, 5 mm slice thickness, 50 s acquisition time). PET transmission images were acquired prior to injection of FDG on an ECAT 931 scanner (2 bedpositions, 128 x 128 matrix, 30 slices, 6.5 mm slice thickness) Patient positioning was similar for both acquisitions with the arms beside the body and the patient normally breathing during acquisition The MMI registration algorithm was applied to the PET transmission and CT images without any preprocessing of the images. The registration procedure is fully automated, except for a rough initial estimation of the axial translation, and takes less than 3 minutes CPU time on a SUN Sparc 20 workstation After registration, the PET transmission images were reformatted to the CT image grid and the quality of the registration was visually inspected by radiologists using the carina and the trachea as anatomical landmarks Registration accuracy was considered very good in all cases. The same reformatting was applied to the PET emission images, assuming that the emission and transmission images are aligned by acquisition. The registered CT and PET emission images were used for localisation of MLN (figure 4).

Preliminary evaluation of the results confirms the clinical relevance of combining PET and CT images for lung cancer staging. However, the automated and highly reliable MMI algorithm has greatly improved the efficiency of the registration procedure and has changed the whole diagnostic procedure. Currently, PET-CT registration is used in clinical routine to decide on lung cancer treatment for 2 patients a week. Weekly clinical sessions are organized in which the available image information, after registration and after post-processing to indicate the relevant structures, is studied jointly by pneumologists, thorax surgeons, nuclear medicine experts and radiologists.

6 Matching of pre- and post-operative CT images of the chin

In this study conducted at the Department of Orthodontics of the UZ Gasthuisberg, Leuven, Belgium, in collaboration with the Radiology Department of the same hospital, the

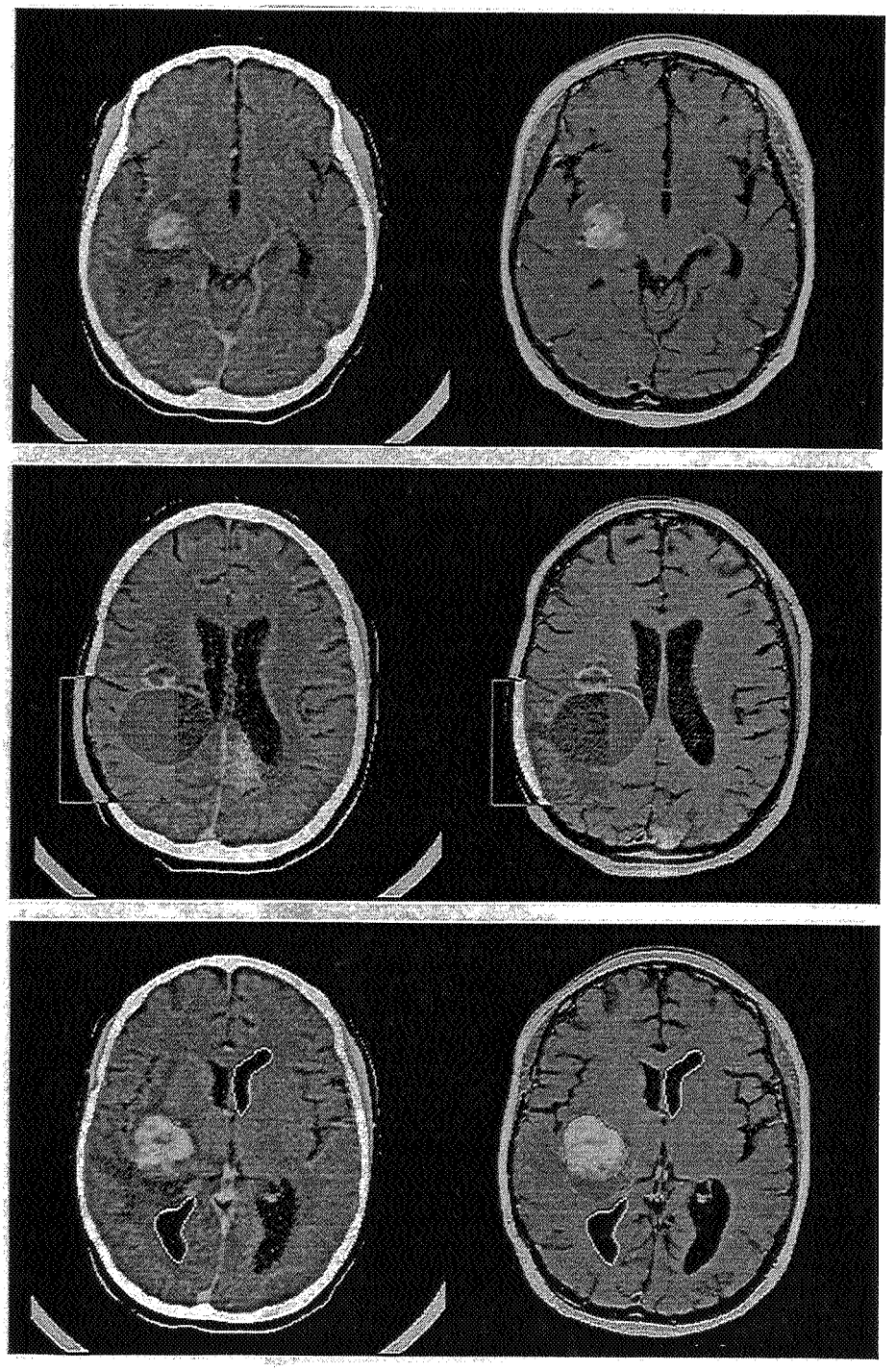


Figure 1 Registered and reformatted axial CT and MR images of the brain. After registration and reformatting, the images can be jointly visualized (top) and the registration accuracy can be evaluated using a moving window overlaying part of one image over the other (mddle). Use of a linked cursor (bottom) allows to delineate contours in one image and transfer these to the other for comparision.

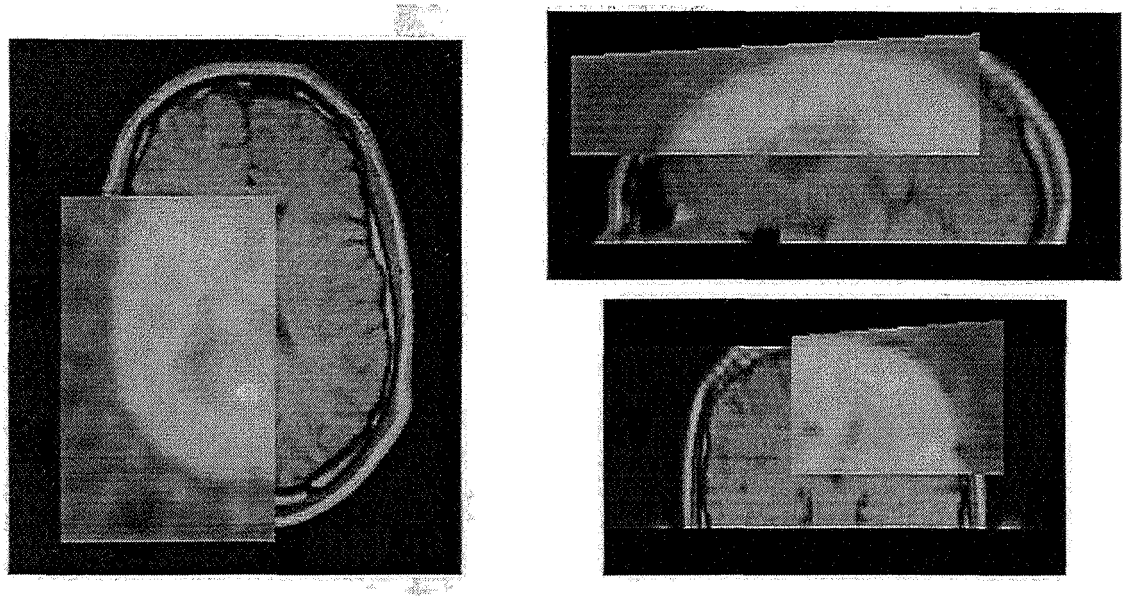


Figure 2: Registration of PET and MR images of the brain. Reformatted axial, sagittal and coronal slices through the PET volume overlayed on the corresponding slice of the MR volume. The mutual information criterion has been validated to be subvoxel accurate compared to the stereotactic reference solution on a large number of cases within the RREP project [8].

value of spiral CT imaging was investigated for the planning and evaluation of chin bone graft implants in alveolar cleft patients. Pre- and post-operative CT images were acquired two weeks before and 9 months after the operation. Registration of pre- and post-operative images allows to define the cleft region in the pre-operative image and localize the same area in the post-operative image, such that the condition of the transplant can be followed over time.

The procedure was applied to 8 patients, all children between 8 and 10 years old. Spiral CT images (512 x 512, 40 to 70 slices, 0.2 mm pixel size, 1 mm slice thickness) were acquired using standard imaging protocols pre- and postoperatively. The images were registered completely automatically and without pre-processing using the MMI registration algorithm. After registration, the pre-operative image was reformatted to match the post-operative image and the result was visually inspected by a radiologist. In all cases, registration accuracy was considered to be very good

A typical result is shown in figure 5. Contours delineated in one image and transfered to the other allow to inspect the registration accuracy. Note that in the 9 month period between both acquisitions morphological changes have occurred due to bone growth, teeth movement and surgery However, these non-rigid movements do not impair the registration algorithm to find the rigid transform that globally very well aligns both images.

7 Conclusion

The mutual information registration method allows for accurate, robust and completely automated registration of multimodality medical images. The method is highly data independent and requires no user interaction, segmentation or preprocessing of the images The same method can therefore be applied succesfully to a variety of different clinical applications. In this paper we demonstrated the usefulness of the method for matching CT, MR and PET images of the head, the thorax and the prostate. In all experiments a rigid body transformation was used, but the method is largely insensitive to local non-rigid deformations, such as induced by positioning differences or surgery. The generality and high reliability of the method makes it a very useful tool in clinical practice. Integration of this tool in clinical practice will stimulate to exploite the benefits of combining information from different modalities in image based decision making.

8 Acknowledgements

This research was supported by IBM Belgium (Academic Joint Study) and by the National Fund for Scientific Research, Belgium, under grant numbers FGWO 3.0115.92, 9.0033.93 and G.3115.92 The authors would like to acknowledge the contributionsto this paper of Patrick Dupont and Sigrid Stroobants to section 5, of Raymond Oyen to section 4, and of Emile Alberga and Johan Van Cleynenbreugel to section 6.

References

[1] A. Collignon, D Vandermeulen, P Suetens, G. Marchal, "Automated Multimodality Medical Image Registration Using Information Theory", *Proc. of the XIV'th Int'l Conf on Informataon Processing in Medical Imaging, IPMI'95*, Brest, France, June 1995, in *Computatzonal Imaging and Vision*, Vol. 3, pp 263-274, Kluwer Academic Plublishers, 1995.

[2] TM. Cover, J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., New York, 1991

[3] J.M Fitzpatrick, Principal Investigator, "Evaluation of Retrospective Image Registration", National Institutes of Health, Project Number 1 R01 NS33926-01

[4] F Maes, A. Collignon, D. Vandermeulen, G Marchal, and P Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Bans. Medical Imaging*, Vol. 16, no. 2, pp. 187-198, April 1997

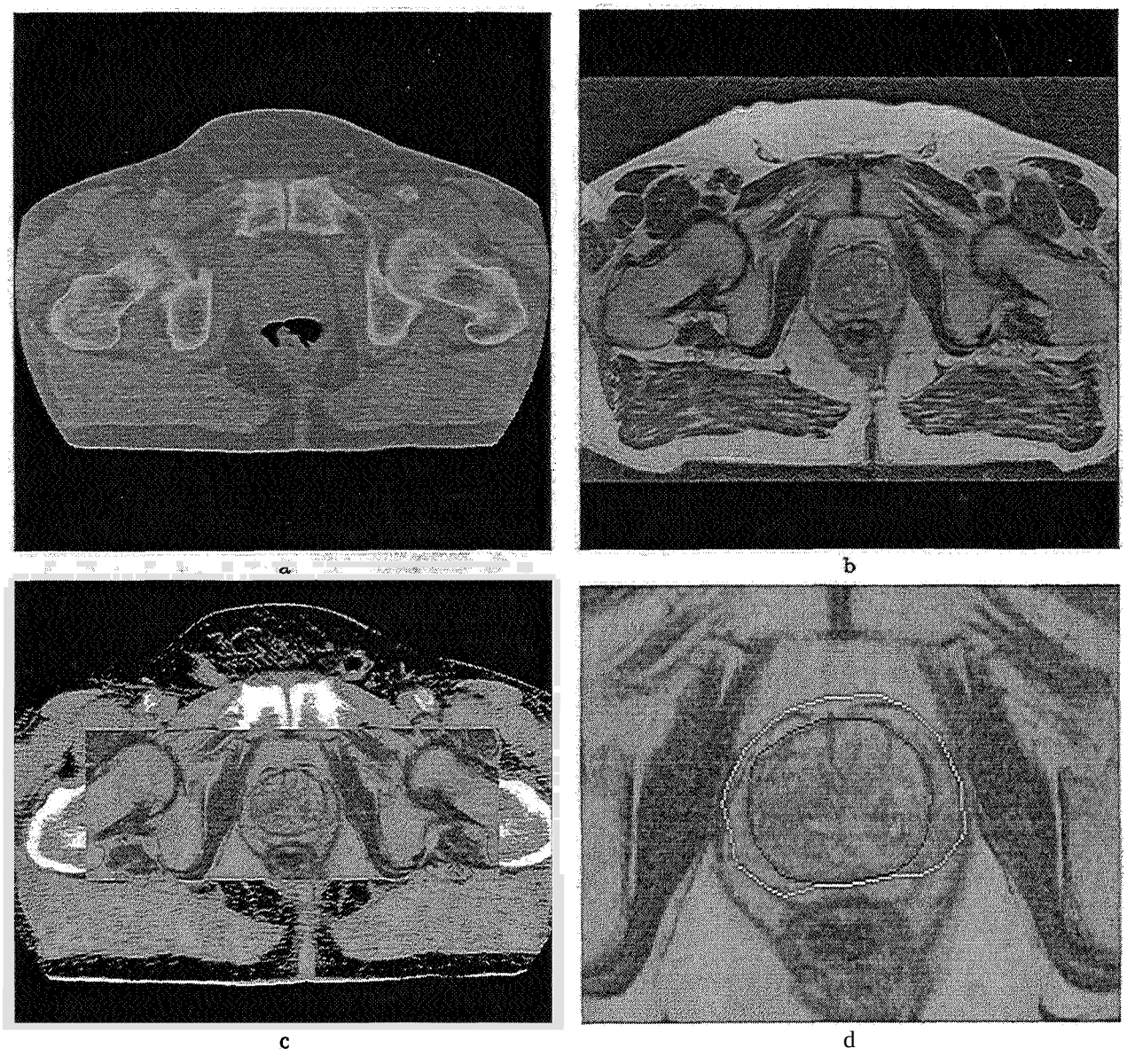


Figure 3 Registration of CT and MR images of the prostate. a. Original spiral CT image. b. Original MR image. c. Registered and reformatted MR image (middle window) displayed on the CT image. This example shows the robustness of the mutual information registration criterion to local rigid deformations. The method correctly registers the rigid bony structures in both images, although large positioning errors and patient related deformations of soft tissues are present. d. Manual segmentation of the prostate in the MR image and the corresponding contour delineated by the same clinician on the CT image and transformed to the MR image.

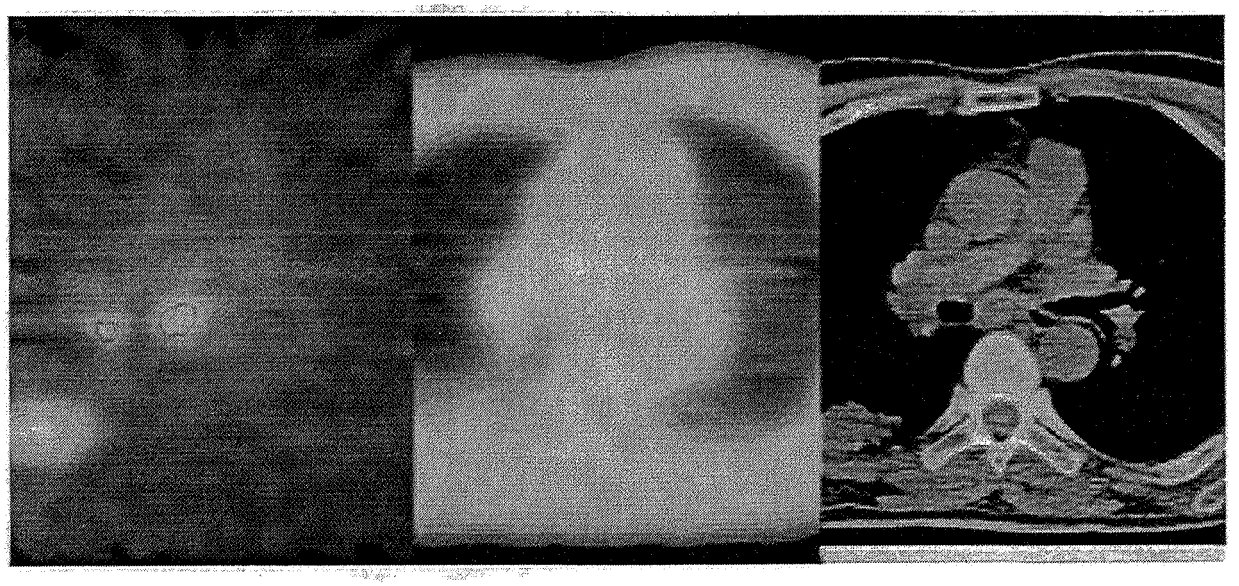


Figure 4 Matching PET and CT images of the thorax for mediastinum cancer staging. The PET transmission image (middle) was matched to the CT image (right), assuming alignment at acquisition of the PET transmission image with the corresponding emission image (left). The PET images have been reformatted to the CT grid to make use of the higher resolution of the CT image. Hot spots detected in the PET emission image (left) can easily be located and anatomically identified in the CT image (right) using a linked cursor

[5] W.H Press, B.P Flannery, S.A Teukolsky, and W T Vetterling, "Numerical Recipes in C," Second Edition, Cambridge, England Cambridge University Press, 1992.

[6] C. Studholme, D.L.G. Hill, D.J Hawkes, "Automated 3-D Registration of MR and CT Images of the Head," *Medical Image Analysis,* Vol. 1, no. 2, pp. 163-175, 1997

[7] W M Wells III, P Viola, H Atsumi, S. Nakajima, and R Kikinis, "Multi-modal volume registration by maximization of mutual information," *Medical Image Analysis,* vol 1, no. 1, pp. 35-51, Mar 1996.

[8] J West, J.M Fitzpatrick, M Y Wang, B.M Dawant, C.R. Maurer,Jr., R.M Kessler, R.J Maciunas, C. Barillot, D. Lemoine, A Collignon, F Maes, P Suetens, D. Vandermeulen, P.A van den Elsen, S. Napel, T.S. Sumanaweera, B. Harkness, P F Hemler, D.L.G. Hill, D.J Hawkes, C. Studholme, J.B.A Maintz, M.A Viergever, G. Malandain, X. Pennec, M.E. Noz. G.Q. Maguire,Jr., M. Pollack, C.A Pellizari, R.A. Rob, D. Hanson, R.P Woods, "Comparison and evaluation of retrospective intermodality brain image registration techniques," *Journal of Computer Assisted Tomography,* Vol. 21, no. 4, pp. 554566, 1997

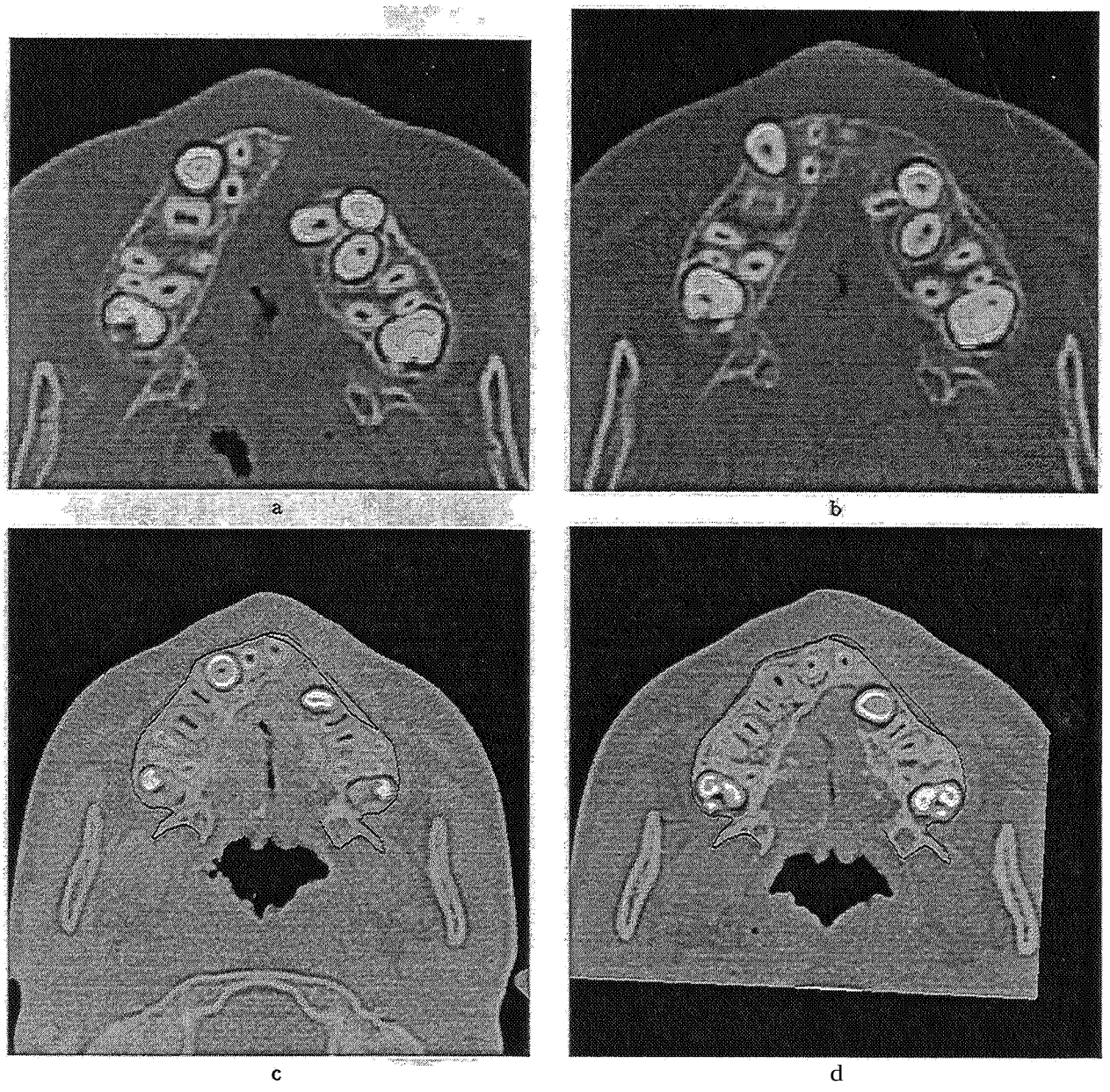


Figure *5:* Matching of pre- and post-operative CT images of the chin of alveolar cleft patients having undergone a bone graft transplantation. Top: Original pre-operative (a) and post-operative (b) images. Bottom. registered Preoperative (c) and post-operative (d) images. The pre-operative image **was** reformatted to match the post-operative image. Contours delineated on the reformatted images using a linked cursor **allow** to evaluate registration accuracy Non-rigid morphological changes due to bone growth, teeth movement and surgery do not impair the algorithm from finding the globally best rigid transform between both images.

A ROBUST GENERALIZED REGISTRATION TECHNIQUE FOR MULTI-SENSOR AND WARPED IMAGES

Molly Dickens, Jagadeesh Neeruganti, Sunanda Mitra, and Edgar O'Hair
Department of Electrical Engineering, Texas Tech University, Lubbock, Texas 79409-3102

ABSTRACT

Registration of multi-sensor and warped images poses difficult problems when control point pairs are not precisely known. The cepstrum registration technique has been used successfully in combination with polynomial warping for registering and finding the disparity between sequential images. Recently, synthetic aperture radar *(SAR)* as well as infra-red (IR) images have been registered by cepstrum techniques. The advantage of this robust area-based technique is that it is more tolerant to noise and rotational shift as well as being less computationally intensive than the traditional correlation measure techniques including the sequenhal search algorithms. We are currently developing a unified and automated registration algorithm for multi-sensor *(SAR,* IR, and Electro-optic) image registration. A comparative review of the current algorithms for registration of multi-sensor and sequential images corrupted by warping and other artifacts such as embedded noise patterns and random noise is presented.

INTRODUCTION

Automatic detection of changes in a scene or recognition of specific objects from a sequence of images acquired under varying conditions and from different modalities often require registration prior to fusion. Without the knowledge of the onginal control points, development of site models for appropnate positioning of the data has been suggested [I-21. However, site models may not be available in many situations. A combination of image analysis and signal processing techniques has been found to be useful in correcting rotational as well as translational shifts [3-6]. Such techniques improve the accuracy of registration yet reduce the computation time compared to statistical correlation or sequential similarity detection techniques [7-8].

The power of these combined techniques lies in the removal of false registration parameters and embedded noise by extracting the dominant features by preprocessing and correction of rotational and translational shifts by power spectrum and power cepstrum respectively [9-10]. Although the translational correction can be done by either the power cepstrum or the conventional phase-correlation techniques [11-12], the power cepstrum technique provides more accurate registration over the phase correlation technique for noisy images with changing dominant features. Finally, a polynomial interpolation can be used to unwarp the images if required.

In ***summary,*** we present a generalized and robust registration technique that is applicable to a broad class of images acquired by different sensors under varying conditions.

METHOD

Application of Power Spectrum for Rotation Correction

Since the Fourier power spectrum has the shift-invariance property, the rotational difference between two simplified images in a sequence can be obtained from minimization of the spectral-intensity difference between two images. The total spectral-intensity difference, $D(\theta)$, can be expressed as a function of the rotational angle between the two images as follows:

$$D(\theta) = \sum_{u}\sum_{v}\left|R(u,v) - S_{\theta}(u,v)\right| \qquad (1)$$

where $R(u,v)$ and $S_\theta(u,v)$ are the Fourier power spectra of the reference and rotated-shifted images, respectively, and u, v are spatial frequencies. Because the original rotation center is not known a priori, the rotation through the angle θ_{min} for minimum $D(\theta)$ makes the dominant features of the two images parallel to each other. Thus only a translational correction must be performed.

Power Cepstrum for Translational Correction

The power cepstrum [13] is usually used to determine the arrival times of a signal and its echoes and their relative amplitudes. Here, the cepstrum technique is extended to two-dimensional signals. In this application, the only problem to be solved by the cepstrum technique is to determine the translahonal difference or disparity between two sub-areas, which are analogous to echo arrival times in one-dimensional signals. In the one-dimensional problem (spatial domain), the function is changed into an image intensity function (a two-dimensional function of position). The power cepstrum is expressed as, (for an image function $I(x,y) = r(x,y) + l(x,y)$),

$$\text{Power Cepstrum} = \left|F\left\{\ln\left[\left|F(I(x,y))\right|^2\right]\right\}\right|^2 \qquad (2)$$

where F is the notation for the Fourier transform of a function and $r(x,y)$ and $l(x,y)$ are the right and left images or the sub-images chosen from the reference image and the shifted image, respectively.

The result of the power cepstrum of $I(x,y)$ is

$$\left|F\{\ln[|F(r(x,y))|^2]\}\right|^2$$
$$+A\delta(x,y)+B\delta(x\pm xo,y\pm yo) \quad (3)$$
$$+C\delta(x\pm 2x0,y\pm 2y_0)+.$$

= Power cepstrum of the right image r(x,y) plus a train of impulses occurring at integer multiples of the translation or disparity (x_0, y_0)

After subtracting $\left|F\{\ln[|F(r(x,y))|^2]\}\right|^2$, the power cepstrum of the right image, from Equation (3), the shift between the two sub-areas can be detected by inspecting the remaining impulse train. In other words, the disparity between is equivalent to the distance between the origin and where the first peak $\delta(x\pm x_0, y\pm yo)$ occurs in the cepstrum plane. This standard power cepstrum needs four Fourier transform computations. A modified power cepstrum technique using three Fourier transform operations has been developed where the impulse train is expressed as:

$$A\delta(x,y)+B\delta(x\pm x_0,y\pm y_0)$$
$$+C\delta(x\pm 2x_0,y\pm 2y_0)+....$$
$$=\left|F\{\ln[1+2a_0\cos(2\pi(ux_0+vy_0))]\}\right|^2 \quad (4)$$
$$=\left|F\{\ln[|R(u,v)+L(u,v)|^2/|R(u,v)|^2-1]\}\right|^2$$

where R(u,v) and L(u,v) are Fourier transforms of r(x,y) and l(x,y), respectively

Finally in the following section, a geometrical transformation using a set of polynomial functions is introduced.

Polynomial Interpolation

One of the most commonly used geometrical transformation techniques for registering warped images is polynomial interpolation using a set of second-degree or higher order functions to warp images and register them [14]. In image registration, two variables x and y are necessary and sufficient for interpolating the surfaces. A set of polynomial functions of second degree in both x and y are expressed as:

$$x'=a_1x^2+a_2xy+a_3y^2+a_4x+asy+a_6 \quad (5)$$
$$y'=b_1x^2+b_2xy+b_3y^2+b_4x+b_5y+b_6 \quad (6)$$

Giving enough information and solving for the coefficients ai through a_6 and b_1 through b_6, a set of polynomial functions can be used to interpolate the warped surfaces and register them. The point (x',y') are the control points on the reference image and the point (x,y) are the control points on the warped image. To solve for these twelve coefficients in Equations (5) and (6), a minimum of six pairs of (x,y) and (x',y') points is needed for a reliable result

After the coefficients have been found, the warped image can be transformed into an image registered to the reference image by converting the old coordinates of the warped image into the new coordinates (i.e., the output (x',y') of the Equations (5) and (6)). The more sets of control points that are chosen, the more accurately the coefficients can be obtained.

RESULTS

Figure 1(b) shows the first cepstral peaks of the image in figure 1(a) of two Ts shifted from each other and fused into one image. Figures 2(a) and 2@) show two infrared images of the same scene shifted from each other, Figure 2(c) shows the cepstral peaks from which the translational shift between the images can be found.

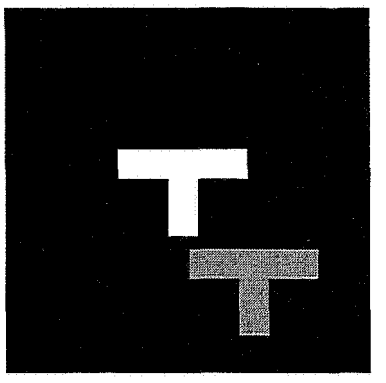


Figure 1(a).

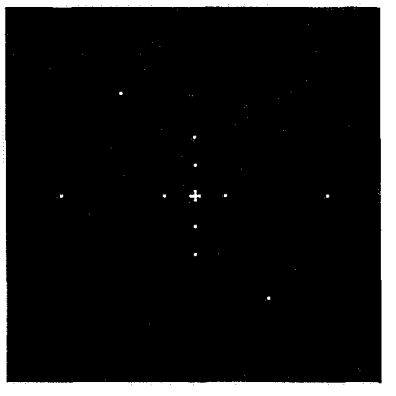


Figure 1(b).

Figure 2(a).

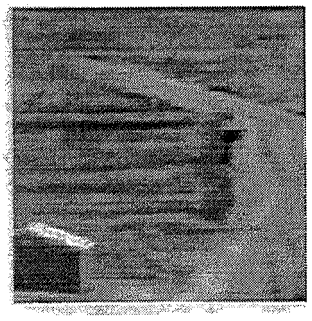


Figure 2(b).

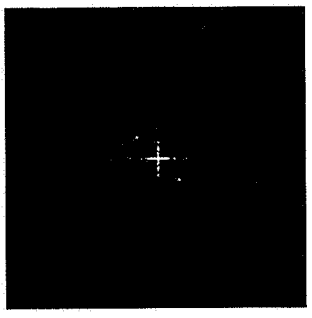


Figure 2(c).

CONCLUSIONS

For noisy images, the power cepstrum registration technique is a robust estimator of translational shifts and tolerates scaling differences up to four percent and rotation up to three degrees. The more traditional phase correlation technique is quite sensitive to noise. In addition, registration by the power cepstral technique does not require the knowledge of control points. Using the shift invariance property of the Fourier transform, the rotational shift can be corrected prior to correcting the translational shift. For removal of warping, however, selection of control points plays a significantrole.

When sequential frames of images are available, a motion stereo model [4] can be used for 3-D visualization of the scene. Application of the power cepstral registration technique to multi-sensor images prior to fusion holds potential advantage over site models due to the unavailability of such models in many situations.

ACKNOWLEDGEMENTS

The current work is supported by contract #S770 1E from Modem Technologies Corporation. The authors would also like to thank Bob Frink of Raytheon-TI Systems for providing us with the IR images.

REFERENCES

1 R. Chellapa, Q. Zheng, P Burlina, C. Shekhar, and K.B. Eom, "On the positioning of multisensor imagery for exploitation and target recognition," Proc, IEEE, Vol. 85, No.1, pp 120-138, 1997

2. T Strat, and W.D. Climenson, "Site model content," Proc. Darpa Image Understanding Workshop, Monterey, California, pp143-147, 1994.

3. D.J. Lee, T.F Krile, and S. Mitra, "Power Cepstrum and Spectrum Techniques Applied to Image Resistration," Applied Optics, 27, 1099-1106, 1988.

4. S. Mitra, D.J. Lee, and T.F Krile, "3-D Representation from Time-sequenced Biomedical Images Using 2-D Cepstrum," paper presented at the First Conference on Visualization in Biomedical Computing, Atlanta, GA, May 22-25,1990.

5. D.J. Lee, S. Mitra, and T.F Krile, "Accuracy of Depth Information from Cepstrum Disparities of a Sequence of 2-D Projections," paper presented at the Advances in Intelligent Robotics Systems Symposium," Philadelphia, PA, November 5-10,1989

6. A Kher and S. Mitra, "Registration of Noisy SAR Imagery Using Morphological Feature Extractor and 2-D Cepstrum," SPIE Proc. on Applications of Digital Image Processing IV, Vol. 1771, 1992.

7 S. Mitra, B.S. Nutter and T.F Krile, "Automated Method for Fundus Image Registration and Analysis," Applied Optics, 27,1107-1112,1988.

8. D.I Barnea, and Silverman, "A Class of Algorithms for Fast Image Registration," IEEE Trans. Computers, C-21, 2, 179-186,1972

9. D.J. Lee, S. Mitra, and T.F Krile, "Analysis of Sequential Complex Images Using Feature Extraction and 2-D Cepstrum Technique Learning Rule," JOSA A, Vol. 6, Feature Issue on Pattern Recognition and Image Understanding, pp 863-870,1989.

10. D.J. Lee, S. Mitra, and T.F Krile, "Hybrid Registration and Analysis Technique for Sequential Complex Images," SPIE Proc. 1153 on Applications of Digital Image Processing, pp 193-202, 1989.

11 C.D. Kuglin and D.C Hines, "The phase correlation image alignment method," IEEE Proc. Conference on Cybernetics Society, pp163-165, 1975.

12. C. Morandi, F Piazza, and R Capancioni, "Digital image registration by phase correlation between boundary maps," IEEE Proc. 134, Part E, No. 2, pp 101-104, March, 1987

13. R.C. Kemerait and D.G. Childers, "Signal detection and extraction by cepstrum techniques," IEEE Transaction on Information Theory, Vol. IT-I8, No. 6, pp 745-759, 1972.

14. W.K. Pratt, "Digital Image Processing," John Wiley & Sons, 1978.

Fusing stereo images for photogrammetric analysis: a guided approach

George G. Moore, Philip J. Morrow and Donal Roantree
School of Information and Software Engineering
University of Ulster, Coleraine,
Northern Ireland BT52 1SA

email: g.moore1@ulst.ac.uk

Abstract

The photogrammetric analysis of urban scene using stereo aerial photographs is a growing industry Traditionally is has relied on manual analysis techniques. However the trend is now towards automation of this process although to date no automated approach has been able to fully satisfy industry's needs. This paper suggests a rationalisation of this trend by integrating the best aspects of both manual and automated systems to form a guided approach. The focus of this paper is on the first major operation of this ***Guided Aerial Image Analysis*** (GAIA) approach, that of ***Guided Stereo Pair Fusion*** (GSPF).

1 introduction

Photo-interpretation based on stereoscopic aerial photogrammetry has grown rapidly in recent years with applications in areas as diverse as digital mapping, Geographic Information Systems, cartography, coal heap analysis and land survey In particular, the mapping of urban scenes is becoming increasingly important with communications companies requiring accurate data for optimal placement of transmitters and receivers.

The technologies currently employed by companies involved in this form of photo-interpretation are primarily manual and rely heavily on the human analyst's powers of cognition. This results in a manpower intensive procedure that is slow, expensive and sensitive to the characteristics of the workforce.

Research performed by other individuals and groups active in the field of stereoscopic aerial photogrammetry has shown significant progress towards the automatic analysis of stereoscopic aerial photographs, notably work carried out at Carnegie Mellon University (McKeown et al. [1994] and Zitnick and Webb [1995]) and the Institut National de Recherche en Informatique et en Automatique (Zhang and Faugeras [1991] and Cohen et al. [1992]).

However, this research has also highlighted the difficulties present in implementing a fully automatic approach and how this may not be the most desirable path to take at this time.

With the above in mind the primary goal of this research has been defined as developing partially automatic techniques to aid the process of interpreting stereo aerial photographs of urban scenes for photogrammetric purposes. This should allow the provision of a more robust and realisable system that will compliment the skills and current procedures of the photogrammetry industry while, at the same time, rationalising and extending research in the field of photogrammetric Image Processing.

To this end an approach referred to as ***Guided Aerial Image Analysis*** (GAIA) has been developed. GAIA rationalises research in automated photogrammetry by accepting that a fully automatic solution may not be feasible and that current technology could be better used in a guided manner. While rationalising research in its field GAIA also contains several novel aspects which in turn extends research in the automatic analysis of aerial photographs for photogrammetric purposes.

Two high-level operations have been identified for detailed consideration. These operations are not only significant in an industrial context but are also representative of *two* of the mqor problem areas encountered in automated photogrammetry In particular, the operations outlined are ***Guided Stereo Pair Fusion*** (GSPF) and ***Guided Artefact Segmentation using Stereo Intelligent Deformable Templates*** (SIDT's).

For reasons detailed later in this paper initial research into GAIA is targeted at developing a prototype that proves the usefulness of the approach. It is hoped that once GAIA has been proven useful further research could be focused on implementing it as an industrial system.

This paper will provide an overview of the main problems within current manual and automated approaches to photogrammetric aerial photo-interpretation. It will then introduce GAIA ***(Guided Aerial Image Analysis)*** at a high level before focusing on the first area considered during research, and the main topic of this paper, namely, ***GSPF*** Finally, by way of closing, an overview of fiture work will be outlined.

2. Problems with current approaches

GAIA is being developed in an attempt to address the problems that exist within current manual and automated approaches to the photogrammetric analysis of stereo aerial photographs of urban scenes. The photogrammetric analysis of such photographs can be split into several key tasks: fusion of the stereo pair to form a registered and calibrated stereo model, segmentation or plotting of the photographed artefacts and classification of the extracted artefacts. For the purposes of this

paper the stereo pair fusion process is of primary importance, although the other operations will be introduced later in the overview of GAIA and future work.

2.1. Manual stereo pair fusion

In the manual case fusion is an iterative process based on the removal of parallax – differences in the two photographs introduced by camera movement – at a number of control points in the area common to both photographs, i.e. the stereo model. There are two types of control point considered: surveyed control points (obtained form ground survey with known position and elevation attributes) and pass points (points of reference selected by a human analyst) which are of unknown position and elevation. The use of pass points is required due to the prohibitive expense of ground survey appropriate to provide surveyed control points for every stereo model in a project. Using pass point and trigonometrical techniques to perform aerial-triangulation the number of surveyed control points can be reduced to one per third model in a strip of stereo models (BKS Surveys [1996]). *Figure 2 1* shows a stereo pair and the stereo model they constitute along with the pair's pass points.

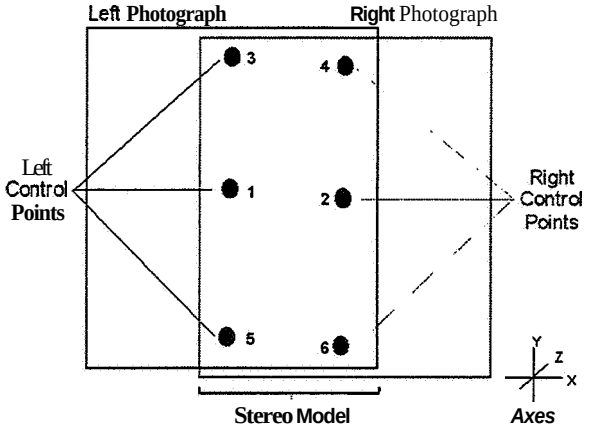


Figure 2 1 Stereo pair showing stereo model and pass points

Typically six pass points are used per model, although five would suffice (Jain et al. [1995]). Each photograph contributes three points to the model – positioned, roughly, along a line running perpendicular to the camera's direction of travel, with one on or near the photographs centre of perspective, as shown in *figure 2 1* Through a structured sequence of adjustments to the position of the photographs in a stereoplotter the projected images of the photographed area are fused, i.e. mapped onto each other, to produce a stereo model which *can* be perceived in three dimensions by a human operator. This will be considered in more detail later in the paper

The main problems in this approach are: the disjoint nature of the fusion process from the later plotting and classification phases, the lack of feedback during analysis and the limitations of using a human to perform the highly iterative and tedious fusion process.

At present older analogue stereoplotting equipment is still widely used in the photogrammetry industry As a result it is common practice to mark pass points and other control information physically on the photographs under analysis and to use these physical marks for reference when moving photographs between analysis devices. This results in wasted time as operators at each device, essentially, repeat some of the work already performed. *An* example of this is where an operator uses a device to physically mark the control points on a photograph. Then at a later stage a stereoplotter operator has to relocate and record the position of the points for reference on that device to facilitate pair fusion and other tasks. The process would benefit from a more cohesive method of marking and recording control information, especially if this could be used to provide a degree of automated assistance to the operator While analytical stereoplotters provide many of these features the trend is towards automation and therefore it is preferable to build these facilities into any computerised solution for increased cohesion and fluency of approach.

As mentioned, another factor is the lack of feedback received by the operators of stereoplotters during the analysis of photographs. While newer stereoplotter technologies have provided improvements in this area more could be done. It is possible that through the provision of more substantive feedback the analysis process could be speeded up and accuracy improved.

This leads on to the final major problem in manual analysis, the limiting factor of a human operator. While humans outperform automated approaches, they are not ideally suited to the efficient photogrammetric analysis of stereo models. Tedium is a major factor here, and is addressed by other aspects of GAIA as outlined later The highly iterative process of manual pair fusion requires repetitive checking for parallax at specified points and the rechecking over and over again lends itself to error

2.2. Automated stereo pair fusion

The automatic fusion of image pairs has received a great deal of attention from the image processing and pattern recognition world for many different reasons. As a result of interest in this area many varied approaches exist. These approaches range from dealing with global rigid and affine transformations between images using cross-correlation and sequential methods such as Svedlow *et* al. [1976] and Barnea and Silverman [1972], to more involved problems dealing with localised deformations of unknown transformation type such as those dealt with by Flusser [1992] using piecewise interpolation methods and Bajscy and Kovacic [1889] using multiresolutional elastic matching.

Brown [1992] defines image registration methods as a combination of feature space, search space, search strategy and similarity metric. Despite the extensive work carried out in the field of image registration a technique that can combine the above elements, to process stereo aerial photographs and produce output

that would be satisfactory to the requirements of the photogrammetry industry, in a useful way has not been achieved.

Typically, problems exist with finding suitable points or features to use for image alignment. Valumetric points are not appropriate for such uncontrolled conditions as those found in aerial photography, while in cluttered urban scenes, higher level features such as corners and edges lead to ambiguity. This exaggerates the ever present correspondence problem and leads to results too uncertain to leave unchecked. Other problems result from the unknown nature of the transformation required for successful matching, given the possible variations of aircraft movement and factors such as perspective. Where results are satisfactory the computational costs are usually prohibitive. This is especially the case where the process would be expected to be undertaken in-line with other manual, or guided, plotting and classification operations.

2.3. Rationalisation of approach

Given that the manual approaches are not satisfactory, being slow and open to error, and that automatic approaches are not suited to be left unsupervised or cannot be applied alongside current manual procedures, even where performance is satisfactory, perhaps a new approach is required.

It may be appropriate to step back from attempting to create a completely automatic solution and use the best features of both manual and automated approaches. These could be combined in a way that allows the automated tasks to be guided by manual supervision and in turn allow manual tasks to be guided by automated assistance. This guided approach to the analysis of stereo aerial photographs is the basis of GAIA and is outlined in the following section of this paper

3. GAIA - Guided Aerial Image Analysis

As introduced above it has been decided that this research should focus on a guided approach to aerial image analysis. GAIA is designed to remove many of the more tedious low-level tasks from the human operator allowing them to oversee the photogrammetric analysis process at a level more suited to a human's abilities. This would provide a guided approach to image analysis free of many of the limitations inherent in the current industrial system.

Guided Aerial Image Analysis (GAIA) rationalises research into automated photogrammetry by accepting that a fully automatic solution may not be feasible and that current technology could be better used in a guided manner GAIA also contains several novel aspects that further research into the automatic analysis of aerial photographs for photogrammetric purposes.

Restrictions imposed upon this research which are outside its control makes it necessary to see this project as proving an hypothesis that would merit the remedying of these external factors. To this end, it has been necessary to 'scope' the research to allow a prototype to be created that would prove this hypothesis. The restrictions relate to the capabilities of stereoplotter equipment, currently employed in the photogrammetric analysis process, and difficulties in achieving the desired level of integration between these devices and the GAIA system.

Ideally, a stereoplotter modified for use with GAIA would be capable of displaying graphical information overlaid on the three dimensional projection viewed in the stereoplotter eyepieces. This would provide a degree of feedback not currently available and would allow output from automated activity to be viewed in a natural way that would allow more accurate evaluation of results. In addition the stereoplotter would be capable of sending details of photograph alignment and other parameters to GAIA and in turn GAIA would be capable of automatically adjusting the settings of the stereoplotter The solution to the problems of implementing the above is an engineering task and it is hoped that the GAIA prototype would prove these tasks worthwhile. It is believed that development of the prototype will not be restricted by the absence of the required degree of integration between GAIA and the stereoplotter and that the required events can be simulated and performance evaluated using prototype interfaces and current stereoplotter technology

Two main high-level operations, introduced earlier, were targeted for consideration in the prototyping of GAIA These operations are not only significant in an industrial context but are also representative of two of the major problems encountered in the area of automated photogrammetry In particular, the operations outlined are ***Guided Stereo Pair Fusion*** and ***Guided Artefact Segmentation using Stereo Intelligent Deformable Templates*** (SIDT's).

3.1 Guided Stereo Pair Fusion

This operation uses control information, supplied by an overseer using a stereoplotter, to automatically perform aerial-triangulation on a stereoscopic pair of aerial photographs. This will generate the necessary adjustments to the stereoplotter settings to fuse the projected images of the photographs and facilitate the viewing of a three dimensional projection using the stereoplotter. The control information would be marked virtually on the separate photographs using a stereoplotter modified for use with GAIA, or in the case of the prototype on images displayed on a computer monitor.

Once control information has been recorded GAIA would attempt to fuse the separate photographs using a set of steps similar to those used in the manual aerial-triangulation process. The overseer could then view the resulting stereoscopic model using a stereoplotter to verify the quality of the automatic fusion, once again this would have to be simulated for prototype purposes. If errors exist in the automatically fused pair the

overseer could guide the pair into correct alignment. Having achieved fusion the parameters of the aerial-triangulation process could be recorded and retained for re-alignment at a later stage. Once this operation is complete both the stereoplotter and GAIA would have access to the registration and calibration parameters of the pair and the stereoplotter would be displaying the projected three dimensional model.

3.2. Guided Artefact Segmentation

Artefact segmentation under GAIA proposes the use of ***Stereoscopic Intelligent Deformable Templates*** (SIDT's) to segment artefacts of interest from the projected stereoscopic model, as viewed by a human overseer using a stereoplotter modified for use with GAIA. It is desirable not only to segment the artefacts but to also classify them. GAIA will allow the overseer to select a type of SIDT to use in the segmentation process. This will have the effect of classifying the artefact prior to segmentation and allows certain characteristics of the artefact to be implied thus focusing the segmentation process more precisely than would be possible with a more generic approach.

As shown in ***figure 3.1*** a human overseer would seed an artefact in the projected stereoscopic model with an appropriate SIDT class This would be done using a stereoplotter modified for use under GAIA or a display on a computer monitor for prototype purposes. This template will then separate into its component parts – a left template, a right template and an output template – and attempt to segment the host artefact. The left and right templates segment the artefact from their corresponding photographs and combine their output in the output template. All three templates will communicate with each other in an attempt to segment the artefact as precisely as possible. The overseer can then monitor the results and correct as necessary This correction process can then be used to provide the templates with heuristic qualities to improve future performance.

The use of a modified stereoplotter to seed the artefact with a SIDT allows the overseer to view the scene in three dimensions and thus is conducive to the correct interpretation of the stereoscopic model. In addition, the stereoplotter can provide GAIA with information relating to the seed position. This reduces the search space for matches between the individual photographs of the stereoscopic pair to a degree that circumvents the correspondence problem.

As already mentioned, the focus of this paper is on the ***Guided Stereo pair Fusion (GSPF)*** phase of this research. It is this topic that is considered next in greater detail.

4. Guided Stereo Pair Fusion

The purpose of this operation is to allow scanned images of the photographs in a stereoplotter to be analysed by GAIA to determine the necessary adjustments to each photograph's orientation within the stereoplotter This would allow GAIA to adjust the settings of the stereoplotter, in which the photographs relating to the scanned images have been placed, in a manner that will present the stereoplotter operator with a fused stereo pair The setting can then be recorded to facilitate future realignment and other stereoplotter and image processing operations.

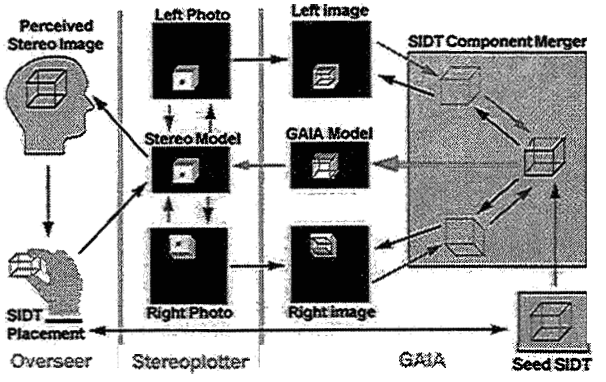


Figure 3.1 Guided Artefact Segmentation using Stereo Intelligent Deformable Templates

The ***GSPF*** process intended for implementation within GAIA is based closely on the manual process used in fusing a stereo pair in a stereoplotter. It was thought important to use the same approach in both the manual and automated components of GAIA to ensure reproducibility of results by both and to maintain consistency Given that the manual system produces acceptable results this appeared to be a logical basis for the approach. The manual process used as a basis is that typically used within industry and is in line with procedures detailed in Spurr [1960] and by Jain et al. [1995].

4.1. Aerial-triangulation

The process of fusing a stereo pair using pass points and trigonometrical techniques is referred to as aerial-triangulation. This process has four distinct stages: control information preparation, interior orientation, relative orientation, absolute orientation, This paper will not attempt to define the aerial-triangulation process in full but comment on it where relevant in describing the ***GSPF*** approach adopted within GAIA.

Aerial-triangulation allows the location and elevation of a point in a stereo pair to be determined through the use of trigonometrical techniques. In practice, the measurements made during initial aerial-triangulation are relative to the stereo pair under consideration. This acts as a guide for building the relative model and the later fitting to ground survey points. It is at this stage that true, real-world, location and elevation information is gained.

Without aerial-triangulation truly accurate photogrammetric observations would require extensive ground survey Such ground surveys may not always be possible due to restrictions imposed by terrain. Where extensive ground survey is possible the effect would be to greatly increase the cost, in both time and money, of photogrammetry projects. Aerial-triangulation

allows the number of survey ground controls to be reduced to as little as one per third stereo model (BKS **Surveys** [1996]), a stereo model being the area of overlap between the photographs in a stereo pair

The basic alignment of the photographs in a stereo pair using aerial-triangulation requires that five pass points be selected in each stereo model (Jain et al. [1995]) although six are used for the purposes of this paper. Each photograph contributes three pass points which are marked across the original direction of aircraft travel, the y-axis, with one on or near the photograph's principle point. By considering each of these points in turn and removing parallax with the corresponding point in the other photograph fusion within a stereoplotter can be achieved. This provides the relative model which is later fitted to surveyed control points. The resulting absolute model is capable of yielding accurate measurements in real world metics.

The process of removing parallax between corresponding pass points in the two photographs is a highly iterative process and has a set of steps that must be performed in strict order

4.2. Removing parallax

To remove parallax the photographs in the stereoplotter must be positioned so as their projected images align in a way that removes parallax introduced by movements of the aircraft during the photographic run. The possible movements of the stereoplotter plate holding the photographs and the effect of the movements on points in the projected images of the photographs are given below.

x-axis: Translation along the original direction of aircraft travel causing all points to be translated in the x direction. There is no movement of points in any other direction.

y-axis: Translation across the original direction of aircraft travel causing all points to move only in the y direction and no other

z-axis: Scaling, causing all points to move in or out from the centre. Clears y-parallax from any point except those along the x-axis.

Kappa: Rotation about the z-axis causing movement of all points in the y direction, except those along the y-axis and movement of all points in the x direction, except those along the x-axis.

Phi: Rotation about the y-axis causing x movement for all points and y movement for points not on the x or y axes.

Omega: Rotation about the x-axis causing y movement to all points and x movement to all points not on the x or y axes.

Given the above it is possible to exploit the properties of the effects of the possible movements to realign the projected images of the photographs and achieve fusion. This is the heart of the fusion process and the process is detailed in the following set of steps *(figure2.*I shows the numbering scheme for pass points).

(i) Clear y-parallax at pass point 1 with Kappa movement of the second photograph,

(ii) Clear y-parallax at pass point 2 with Kappa movement of the first photograph,

(iii) Clear y-parallax at pass point 3 with Phi movement of the second photograph,

(iv) Clear y-parallax at pass point 4 with Phi movement of the first photograph,

(v) Observe y-parallax at pass point 5.
If pass point 5 exhibits y-parallax remove the parallax and overcorrect with the introduction of one and a half times the observed y-parallax in the opposite direction. This is done using Omega movement of either photograph.
Repeat steps (i) to (v) until no y-parallax is observed at pass point 5 in step (v),

(vi) Check pass point 6 for y-parallax.
If y-parallax exists at pass point 6 the entire procedure must be repeated as unobserved y-parallax must still be present at one or more of the other pass points.

The above steps remove y-parallax from the entire stereo model. This effect can be explained by considering the effect of each movement type detailed earlier and noting the following. Step (ii) does not introduce y-parallax to pass point 1, therefore both pass points 1 and 2 are clear of y-parallax at this stage. Step (iii) does not introduce y-parallax at either pass points 1 or 2, therefore pass points 1, 2 and 3 are now clear of y-parallax. Step (iv) does not introduce y-parallax at pass points 1, 2 or 3, therefore at this stage pass points 1, 2, 3 and 4 are clear of y-parallax. Step (v) introduces y-parallax at all other points, repeating steps (i) to (v) will remove this again but several repetitions of steps (i) to (v) may be necessary before no y-parallax is observed at pass point 5.

Retention of x-parallax is desired for the computation of point elevations. At this stage measurements would be in relative terms due to incorrect amounts of x-parallax and other unknowns. However, once fitting of the relative model to surveyed ground control points and the camera model are considered real-world measurements will be available, **Spurr** [1960] and Jain et al. [1995].

4.3. Automating the pair fusion process

Modern analytical stereoplotters will, to some degree, allow the automatic fusion of stereo pairs. However as stressed earlier, there is a trend towards automation of the photogrammetic analysis process. By way of preparing the approach that drives GAIA for this ideal goal it was decided that a core operation such as pair fusion should be implemented. This process having been implemented will also assist in the next phase of **GAIA** development, ***Guided Artefact Segmentation.*** The

implementation of ***GSPF*** has two areas of interest: setting pass points and finding the necessary transformations to align the projected images.

Selection of pass points will be left under operator control. This removes the problems of finding correspondence between the control points in each photograph, a typical cause for fully automated approaches to fail or to be too unreliable to leave their output unchecked. If the operator is to check the point selection and matching processes then the points must be of a type that a human *can* accurately perceive. To this end it is preferable to allow the operator to select the points thus improving overall performance and ensuring that human verification of the process is possible. A degree of assistance is provided however by allowing **GAIA** to highlight matches in each image for points selected in the other Once the pass points have been selected the necessary transformations required to remove parallax can be computed.

Rather than attempting to optimise the transformation required to bring the photographs into alignment (Brown [1992]) the transformations used in the six step manual process above will be applied in the order used in that process. As each of the transformations are reasonably simple rotations and perspective projections and will be operating on a small data set (the pass points) processing is expected to be fast enough to be implemented in line with other photogrammetric processes. It is appropriate to only compute the transformations at the pass points as the goal is to determine parameters for the stereoplotter, not to rectify the scanned images.

This approach has novelty in that it does not attempt to optimise the transformation process by computing the best transformation to align the images but attempts to simplify the process to a point where optimisation is not an issue. This is the same as the approach taken towards the correspondence problem, which uses human guidance to remove the need to search for matches. All this hinges on the assertion that it is necessary to retain human intervention if output is to be of an acceptable level for industry requirements.

A prototype of the complete system is under development using Khoros – a UNIX based toolset for manipulating data sets that specialises in image processing – to prove this approach. A screen shot showing pass points being marked is shown in ***figure 4.1*** The **GSPF** process has already been implemented using two images diplayed one beside the other The operator interacts with these images to select pass points which are then visibly marked and recorded. To assist the operator zoom facilities allow precise positioning of the points. Testing is still underway but initial results show that the **GSPF** process can accurately fuse an image pair at a speed suitable for an in line process.

5. Current and future work

Testing and refinement of the ***GSPF*** process is still in progress. Once ***GSPF*** has been implemented and tested to gauge its performance, work will progress on the other main **GAIA** operation, ***Guided Artefact Segmentation using SIDT's***. When both main operations have been implemented they will be integrated into a single system. It is hoped that results from this final prototype would justify further research into this approach.

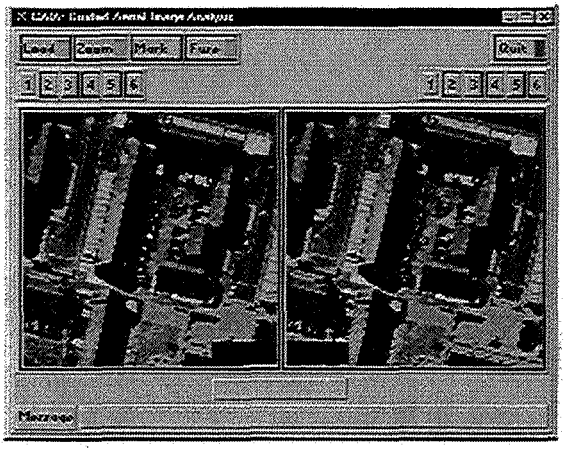


Figure 4.1 Screen shot of GAIA being used to mark pass points on aerial images

6. Summary

It has been indicated that current manual and automated approaches to the photogrammetric analysis of stereo aerial photographs of urban scenes are not satisfactory Manual approaches are slow and open to error, while current technology and techniques can not offer an automated solution that meets the needs of the photogrammetry industry It has been suggested that a rationalisation of approaches should take place and that the best aspects of both be integrated to form a guided, rather than automated, approach. ***Guided Aerial Image Analysis*** **(GAIA)** has been proposed as such an approach.

An introduction to GAIA's main operations – ***Guided Stereo Pair Fusion*** and ***Guided Artefact Segmentation using*** *SIDT's* – was given before focusing on the former, which was the main subject of this paper It has been shown how many of the major problems of automatic stereo pair fusing can be avoided by the inclusion of human intervention at an appropriate level. Further, it has been shown that by simply adapting current manual processes a workable guided approach *can* be provided which would improve upon existing manual procedures.

It is believed that the philosophy behind **GAIA** provides a workable solution to the needs of the photogrammetric industry It is hoped that through the production of a prototype this hypothesis will be proved.

References

Bajscy, R., and Kovacic, S. (1889). Multiresolution elastic matching. CVGIP, Vol. 46, 1-21

Barnea, D.I., and Silverman, H.F. (1972). A class of algorithms for fast digital registration. IEEE Transactions on Computing C-21, pp. 179-186.

BKS Surveys (1996). Personal communications.

Brown, L.G. (1992). A survey of image registration techniques. ACM Computing Surveys, Vol. 24, No. 4, pp. 325-376.

Cohen, I., Cohen, L.D., and Ayache, N. (1992). Using deformable surfaces to segment 3-D images and infer differential structures. CVGIP Image Understanding, Vol. 56, No. 2, pp. 242-263.

Flusser, J. (1992). *An* adaptive method for image registration. Pattern Recognition, Vol. 25, No. **1,** pp. 45-54.

Jain, R., Kasturi, R., and Schunck, B.G. (1995). Machine Vision. MIT Press and McGraw-Hill.

McKeown, D.M., Cochran, S.D., Ford, S.J., Gifford, S.J., Harvey, W.A., Hsieh, Y.C., McGlone, J.C., Polis, M.F., Roux, M., and Shufelt, J.A. (1994). Research in the Automated Analysis of Remotely Sensed Imagery 1993-1994. Technical report, School of Computer Science, Carnegie Mellon University

Spurr, S.H. (1960). Photogrammetry and Photo-Interpretation. The Ronald Press Company

Svedlow, M.R., McGillem, C.D., and Anuta, P.E. (1976). Experimental examination of similarity measures and preprocessing methods used for image registration. In symposium on machine processing of remotely sensed data. pp. 4A-9.

Zhang, Z., and Faugeras, O.D. (1991). Estimation of displacements from two 3D frames obtained from stereo. Institut National de Recherche en Informatique et en Automatique, Programme 4 Robatique, Images et Vision, Report No. 1440.

Zitnick, L.C., and Webb, J.A. (1995). The Reformation of Stereo Vision. Technical report, School of Computer Science, Carnegie Mellon University

Acknowledgements

BKS Surveys, Coleraine for details of the current situation within the photogrammetric industry

The Department of Education for Northern Ireland are finding this project.

An Efficient Registration and Recognition Algorithm via Sieve Processes

P Jonathon Phillips
U S. Army Research Laboratory
AMSRL SS SE
2800 Powder Mill Rd
Adelphi, Md 20783
e-mail:jonathon@arl.mil

Junqing Huang
Stanley M. Dunn
Department of Biomedical Engineering
Rutgers University
New Brunswick, NJ 08903
e-mail:junqing@occlusal rutgers .edu
e-mail:smd@occlusal rutgers edu

Abstract

A fundamental problem in computer vision is establishing correspondence between features in two images of the same scene. The computational burden in this problem is solving for the optimal mapping and transformation between the two scenes. In this paper we present a *sieve* algorithm for efficiently estimating the transformation and correspondence. A sieve algorithm uses approximations to generate a sequence of increasingly accurate estimates of the correspondence. Initially, the approximations are computationally inexpensive and are designed to quickly sieve through the space of possible solutions. As the space of possible solutions shrinks, greater accuracy is required and the complexity of the approximations increases. The features in the image are modeled as points in the plane, and the structure in the image is a planar graph between the features. By modeling the object in the image as a planar graph we allow the approximations to be designed with point-set matching algorithms, geometric invariants, and graph-processing algorithms. The sieve algorithm is demonstrated on three problems. The first is registering images of muscles taken with an electron microscope. The second is aligning images of geometric patterns taken with a charged-couple device (CCD) camera. The third is recognizing objects taken with a CCD camera.

1 Introduction

Two fundamental problems in computer vision are (1) registration of two images of the same scene and (2) object recognition or identification. In some instances both problems can be solved by variations of the same technique. In this paper we solve these problems via a sieve process that combines graph matching and point-set matching. The sieve process is a general framework for efficiently finding the correspondence between images, or images and models.

The underlying formulation is a point-set matching problem that is the study of algorithms for measuring the similarities between two sets of points or features. Examples of features are corners, intersections of lines, segments, and points of high curvature on a boundary. The output of a point-set matching algorithm is a mapping ℓ between features and a transformation T that aligns the images. The computationally expensive portion of point-set matching algorithm is searching for the correct mapping. This *can* be done by exhaustively searching a set of mappings [6] or by pruning methods [2, 7, 16].

In a sieve process, a sequence of stages gradually reduce the number of mapping functions that need to examined, i.e., the analogy of sieving through the set of mappings with gradually finer meshes until the answer is found. The set of sieves is derived from an image model, which is a planar graph that encodes the spatial organization of the features. In the sieve algorithm, the graph represents the image and objects in the image. The planar graph need not literally represent the object, but can represent the geometric structure of the features being used by the correspondence algorithm

The sieves reduce the set of mappings that need to be examined to less than 10. The reduction is accomplished by searching the graphs for local structures that most likely correspond to each other The searching is done with geometric hashing and geometric invariants. By utilizing the graph structure, the number of geometric invariants and hash values calculated is sublinear in the number of features. By comparison, in a straightforward application of these techniques the number of values computed is cubic in the size of the feature set.

There is a long history of graph-matching algorithms in

the literature [5, 15], where graphs are globally matched to each other In real-world applications the graph extracted from two images of the same object are not isomorphic (figure 1) and handle differences between the two graphs by introducing a set of parsing rules that model changes in the graphs. The sieve algorithm is robust to global deformations in the graph because only local features are considered. The algorithm can register images as long as a single vertex neighborhood exists in both images that has not been corrupted by noise.

Point-set matching using the vertices of the graphs is robust to global variations in the graph. Usually, deformations in the graph are generated by the addition or deletion of vertices. A robust point-set distance is not sensitive to the addition or deletion of a reasonable number of vertices. In section 3 we describe a robust point-set distance function. In addition, point-set matching is not sensitive to topological changes in the graph because the vertices, not the edges, are used.

The technique above is demonstrated on three problems. The first is from biomedical imaging: finding the correspondence for a stereo image from a pair of images taken with an electron microscope [12, 13]. In this problem, the equipolar line is not known and must be recovered by the sieve algorithm. This is different from algorithms in the literature where the equipolar line is assumed to have been given [1, 3, 9]. Currently, the system of Peachy and Heath [11] performs this reconstruction; however, it requires a human operator. In this system, a stereo pair of images is displayed side by side for the operator to view stereoscopically, with the operator performing three dimensional (3-D) reconstruction The major drawback of this system is that it is time-consuming for the operator to align the images. Approximately one week is needed to process a series of 12 images. The automatic algorithm of Huang et al [4] finds the alignment between micrographs using geometric invariants and geometric hashing. In the algorithm, all possible geometric invariants are computed, whereas in the sieve algorithm we restrict attention to the most likely correspondence between features. Thii produces an algorithm that is computationally efficient. The second problem is from computer vision: aligning two images of the same geometric pattern taken with a charged coupled device (CCD) camera. The third problem is modelled after the research of Mokhtarian and Murse [10], where a library of models is given and the algorithm locates and identifies these objects in a set of images.

Performing these three experiments demonstrates the versatility of the sieve algorithm. Aligning the micrographs is a real-world problem with applications in biomedical research. Registering the geometric patterns demonstrates the algorithm in a different imaging domain. Identifying objects has important applications in manufacturing and parts inspection

2 Sieve Algorithms

The correspondence between images A and B is found by a five-stage sieve process. The input to the first stage is a planar graph that represents the structure of the objects in the image. The graph is extracted by a preprocessing algorithm. As the preprocessing algorithm depends on the class of images being processed, it is not discussed here. The output of the final stage of the sieve algorithm is the correspondence between the features and the transformation between the images. Between the initial stage and the final stage, a sequence of approximation functions generates, evaluates, and prunes hypothesized matching functions. Initially, computationally simple methods are used. As the algorithm proceeds, the hypothesizing and pruning functions grow in complexity The design of each of the functions is based on the geometry and structure of the problem. A description of the five-stage sieve algorithm follows.

1) Screen features locally — The first stage screens the set of features (vertices) to select a subset of features that are most likely to occur in both images. The screening is accomplished via local graph structure. Because of noise introduced by image formation, artifacts introduced in the preprocessing step, and occlusion, the graphs produced are not topologically identical. Overall, the graphs have the same basic structure, but locally there are differences, such as missing or additional vertices, and missing edges. The graph processing step's goal is to identify edges and vertices most likely to occur in both images using the assumption that the greater the edge length, the greater chance it is in both images. The rationale is that the preprocessing step is more likely to create a false vertex that splits an edge, or fails to extract an edge, than to concatenate two edges into a single edge.

The graph-processing module is executed independently on each image of the pair The first step is to compute the length of each edge in the graph Each edge of the graph is then marked GREEN if its length is above a threshold and RED otherwise. A vertex is then marked GREEN if there is a sufficient number of GREEN edges adjacent to it, otherwise the vertex is marked RED. The lit of GREEN vertices from both images is passed to the next stage.

2) **Assign indices to features —** In the second stage, a set of indices is assigned to the GREEN features. By comparing two indices, one can estimate the likelihood that two features correspond to each other The indices are a pair of geometric invariants, the angle between edges adjacent to a vertex and ratio of the length of the edges.

Given three points, a, b, and c, one can compute the above invariants. One point, b, is selected as the center point. The first invariant is the interior angle between segments $\overline{ab}$ and $\overline{bc}$ and the second is the ratio of lengths of these segments. If there are n vertices in an image, then there are $O(n^3)$ different invariant values for that image. For computational efficiency we want to compute only a small fraction of the set of all invariants. This restriction is accomplished by computing an invariant pair $\theta(i,j,k)$ and $r(i,j,k)$ only if v_j is a GREEN vertex and if the vertices v_i and v_k are neighbors of v_j Since each GREEN vertex has degree 3, each GREEN vertex is the center of three invariant pairs. Let m_A (m_B) be the number of GREEN vertices in image A (image B), then $3m_A + 3m_B$ invariant pairs are computed.

3) Prioritize matches — In this stage, the invariants estimate the likelihood that a GREEN vertex from image A maps to a GREEN vertex from image B The GREEN vertices were selected because they are most likely to represent features that are common to both images. To proceed we must determine which GREEN vertices in image A correspond to which GREEN vertices in image B We do this by comparing each invariant pair from image A with each invariant pair from image B. Each of these pairings is given a score that assesses the likelihood that the vertices correspond to the same feature. Each of the pairiigs between the invariants from the different images is placed in a priority queue Q The most likely pairings are placed at the front of the queue. The order is determined by a function $h(k,l)$, where k is the index of an invariant for image A and l is the corresponding index from image B Each

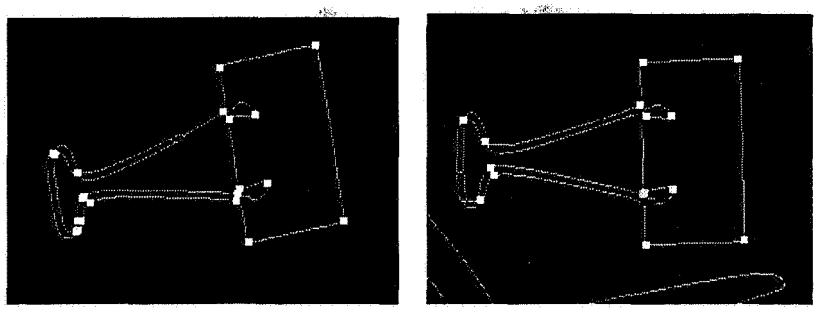


Figure 1- The graphs extracted for different images of the same object, illustrating differences in the graphs. The images have been processed by the sieve algorithm and the vertices that were matched are marked as white squares. (a) Graph from image *A* and (b) Graph from image *B*.

invariant pair, $\theta_A(k)$ and $r_A(k)$, is referenced by a single index *k*. Since each GREEN vertex generates three pairs of invariants, a set of index functions refers back to the vertices that generated the invariants. These functions are $a(\ )$, $c(\ )$, and $\beta(\ )$. The values of the invariants are referenced as $\theta_A(k) = \theta(\alpha(k), c(k), \beta(k))$ and $r(k) = r(\alpha(k), c(k), \beta(k))$. The functions $\theta_B(k)$ and $r_B(k)$ are defined in a similar manner Let $h(k,l)$ be the measure of similarity between two invariant pairs, where $h(k,l)$ is a function of $\theta_A(k)$, $r_A(k)$, $\theta_B(l)$, and $r_B(l)$. We choose $h(k,l) = \lambda_1|\theta_A(k) - \theta_B(l)| + \lambda_2|r_A(k) - r_B(l)|$, where λ_1 and λ_2 are constants that determine the relative importance of the angle invariant versus the ratio invariant.

The comparison of invariant *k* from image *A* and *l* from image B generates a unique matching ℓ_{loc} between vertex $v^A_{c(k)}$ and its three neighboring vertices, and between $v^B_{c(l)}$ and its three neighboring vertices (figure 2). The vertex $v^A_{c(k)}$ is mapped to $v^B_{c(l)}$. The vertex $v^A_{\alpha(k)}$ is mapped to either $v^B_{\alpha(l)}$ or $v^B_{\beta(l)}$. The vertex to which $v^A_{\alpha(l)}$ is mapped is determined by the topology of the graphs G_A and G_B The vertex $v^A_{\beta(k)}$ is mapped to the other vertex, the one not mapped to $v^A_{\alpha(k)}$ The vertices $v^A_{c(k)}$ and $v^B_{c(l)}$ both have one neighboring vertex that was not used in computing the geometric invariant. These vertices are mapped to each other Once the local matching ℓ_{loc} is generated, the optimal transformation T_{loc} between the two sets of vertices is computed.

Figure 2 illustrates the above situation The vertex v^A_c (v^B_c) is GREEN in image *A* (image *B*). The vertices v^A_α and v^A_β (v^B_α and v^B_β) are the neighboring vertices in G_A (G_B) used in computing the invariants. The vertices v^A_δ and v^B_δ are neighboring vertices not involved in computing the invariants.

4) Extend local matches to global matches — Through the third stage, all decisions are based on local information. In the fourth stage candidate local matchings are extended to global mappings. Local matchings are extended until a suitable global mapping is found. The order in which they are examined is determined by the priority queue. The first step of the examination extends matches previously generated into an initial global mapping and measures the goodness of this global mapping. The goodness measure determines if the initial global mapping is sufficiently close to the optimal mapping that it can be passed to the final stage of the sieve algorithm for refinement.

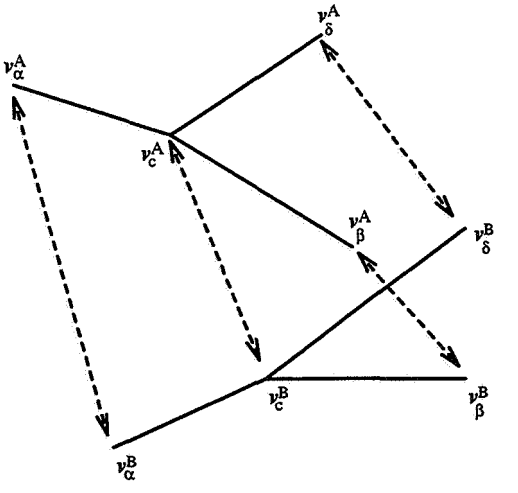


Figure 2: Matching between two GREEN vertices and their neighborhood. v^A_c and v^B_c are the two GREEN vertices. The vertices used in computing the invariants are v^A_α, v^A_β, v^B_α and v^B_β The vertices v^A_δ and v^B_δ are the vertices not involved in computing the invariants, and are mapped to each other

To fully evaluate a hypothesized mapping and transformation, the algorithm uses global information about the entire point set and its geometric structure. The hypothesized mappings are examined according to their order in the priority queue, For this purpose, the local mapping ℓ_{loc} is extended to the entire point set, and a goodness measure evaluates the mapping. The transformation T_{loc} generates an initial global mapping ℓ_{init}. The initial mapping ℓ_{init} is constructed as follows:

$$\ell_{init}(a_i) = \arg\min_{b_j \in B} d_2(T_{loc}(a_i), b_j).$$

The goodness measure that evaluates the initial mapping is the number of pairs of mapped points that are greater than δ_{init} from each other The measure d_{init} is

$$d_{init} = \#\{a_i : d_2(T_{loc}(a_i), \ell_{init}(a_i)) \geq \delta_{init}\},$$

where $\#\{\ \}$ is the size of the set. If $d_{init} \leq \Delta_{init}$, then the initial mapping is sufficiently accurate, and ℓ_{init} and T_{loc} are passed to the fixed-point algorithm. If $d_{init} > \Delta_{init}$, then the initial mapping is dismissed, and the next pair of invariants in the priority queue is examined. Experience has shown that the current method of evaluating an initial mapping is very robust. The score for a correct initial mapping is several factors smaller than a score for an incorrect mapping.

5) Use fixed-point algorithm — The last step in the algorithm is to take the initial match generated by step four and use a fixed-point procedure to generate the final estimate of the matching function $\hat{\ell}$ and the transformation $\hat{T}$ All the points from images A and B are used in this step. The fixed-point algorithm is the following two-step iterative algorithm. Given a match, the optimal transformation is computed This is followed by a computation of the optimal matching function given the estimated transformation. This procedure is repeated until the matching function and, therefore, the transformation do not change. The point at which this occurs is a fixed point.

3 Robust Point-Set Distance

In this section we describe the robust point-set distance function used by the sieve algorithm. The distance allows for missing points in both point sets.

Formally, The point-set matching problem is expressed as follows: Given two sets of points A and B, a family of transformations $\mathcal{T}$, and a family of mappings $\mathcal{F}$ between the sets A and B, find $T \in \mathcal{T}$ and $\ell \in \mathcal{F}$ that minimize the distance between $T(A)$ and B The mapping ℓ tells which points in A correspond to which points in B

The family of allowable transformations $\mathcal{T}$ is the set of rigid transformations in the plane. The family of allowed mappings $\mathcal{F}$ is the set of functions from A to B. This family of mappings was chosen over the set of injections or bijections because it better reflects situations that arise in practice, such as when the preprocessing module splits a single feature into two features. Preprocessing modules are also notorious for introducing extraneous features. This problem is handled by a robust interpoint distance function An example is the interpoint distance function where a penalty of c is incurred if the distance between the points is greater than δ Formally, the distance function is

$$dist_p(x,y) = \begin{cases} d_2^2(x,y) & \text{if } d_2(x,y) \leq \delta \\ c & \text{otherwise} \end{cases},$$

where $c \geq \delta^2 > 0$. The point-set distance function used in this paper is a robust version of the Procrustean metric, where $dist_p(\ ,\ )$ is used instead of the Euclidean metric. The resulting point-set distance function, given T and ℓ, is

$$\begin{aligned} D_p(T,\ell) &= f(dist_p(T(a_1),\ell(a_1)), \quad , \\ &\qquad dist_p(T(a_m),\ell(a_m))) \\ &= \sum_{i=1}^{m} dist_p(T(a_i),\ell(a_i)) \end{aligned}$$

To avoid the computational expense of searching for the global minimum, we use approximation methods. One approximation method to compute the metric is the fixed-point algorithm The fixed-point algorithm computes the optimal transformation given a matching, then given that transformation, it computes the optimal matching with respect to that transformation. The algorithm continues in this manner until a fixed point is encountered. The algorithm reaches a fixed point when the transformation and matching do not change. This is summarized in the following theorem, which is proven in the appendix.

Theorem 1 If $f(x_1, \quad , x_m)$ and dist(·,) are defined us above and $\mathcal{F}$ is the set of all functions *from* A to B, then given any $T \in \mathcal{T}$ or $\ell \in \mathcal{F}$, the fixed-point algorithm *will* converge to a *fixed* point.

4 Experiment: Registering Micrographs

We demonstrate the efficacy of the sieve algorithm by applying it to the problem of registering thick section micrographs. A micrograph is an image taken with an electron microscope; one of its uses is to image biological specimens. The specimen being imaged is referred to as a section. The physical depth of sections varies. If the section is sufficiently thick, one can determine from stereo micrographs the 3-D structure of an anatomical system Reconstructing the depth information from the stereo pair requires the correspondence and transformation between the images. One system of interest is the Tubule-system (T-system) in muscle tissue.

In the T-system the network structure can be modeled as a planar graph With a pair of images of the T-system, the sieve algorithm generates a mapping for construction of a stereo image.

The T-system network can be seen when the skeletal muscle tissue samples are treated with Golgi stain and imaged in an intermediate voltage electron microscope (IVEM) The resulting images are micrographs, which are 2-D projections through the 3-D specimen The resulting image of the 3-D network is a 2-D network with dead ends. The specimens are relatively thick, on the order of several micrometers, and contain important 3-D information that is lost by the imaging process.

As the image formation process is orthographic, it is not possible to reconstruct the 3-D structure of the sample from a single 2-D image. To recover the 3-D structure requires a second image of the specimen at a different viewing angle. Using the two images and triangularization based on the difference in viewing angle, the 3-D structure of the network in the specimen can be recovered. To perform the triangularization requires knowledge of the transformation between the two images of the specimen and the correspondence between the features in the two images.

The two images of a specimen are taken separately by the same electron microscope. For the first micrograph, the specimen is placed in the microscope, rotated about the y-axis by $-\beta$, and imaged. Because of the mechanics of taking an image, when the second image is taken, the specimen is repositioned in the microscope. This results in unknown translation parameters and an unknown rotation parameter about the z-axis. The optical axis is parallel to the z-axis. To allow for recovery of depth information, the specimen is rotated about the y-axis by a known angle β. Figure 3 contains a stereo pair of micrographs.

In the preprocessing module a planar graph is extracted from the micrograph. The planar graph represents the T-system structure. The first step in the preprocessing module is to apply a median filter to the micrograph. The module creates a binary image by thresholding this image, which separates the T-system from the background The threshold is automatically estimated from the image histogram. The module then extracts the planar graph by thinning the segmented image to its skeletal frame. Vertices (features) are extracted from the skeletal frame. Thinning is accomplished with a combination of iterative and morphological methods. The iterative method removes pixels from the boundary of the segmented object, layer by layer, until all that remains is the skeleton The morphological method uses a combination of erosions and dilations to achieve the thinning.

The planar graph from the micrographs is the input to the sieve algorithm. Figure 4 shows the graph produced by the preprocessing module, as well as the GREEN vertices and adjacent edges extracted by the graph-processing stage of the sieve algorithm Figure 3 shows the results of the sieve algorithm applied. The vertices that were successfully matched in each image are marked as white squares. The stereo pair in figures 3 and 4 is image set 5 in table 1.

The sieve algorithm was run on five pairs of micrographs. The correspondence mapping for each pair of images was compared with the ground truth for that pair, the ground truth was generated by visual inspection, according to the accepted practice in this field. We report the number of incorrect matches for each pair along with the number of multiple matches (table 1). A multiple match occurs when a feature in image *A* is matched with two or more features in image *B* Because the preprocessing module does not extract isomorphic graphs, there are vertices in one image that do not have a match in the other image.

This experiment demonstrates that the sieve algorithm is an efficient and accurate method for registering micrographs. The system of Peachy and Heath [11] takes one week to register and align a set of 12 images. It is not possible to directly compare the performance of our work with that of Peachy and Heath, because their algorithm does not extract point features. Our results are comparable to those of Huang et al [4], an algorithm based on geometric hashing [8] and has a running time of $O(n^3)$. The key algorithmic components of our approach are graph-search algorithms and sorting. The complexity of these algorithms is $O(n)$ and $O(n \log n)$, respectively.

5 Experiment: Geometric Patterns

In a second experiment, we demonstrate the sieve algorithm by registering two images of a geometric pattern This is an example of registration for constructing stereo images or object recognition in computer vision.

In our experiment, we took two images of a geometric pattern with a CCD camera (figure 5). Image *A* was captured, then the camera was rotated and translated before image *B* was taken. Before being processed by the sieve algorithm, the images were preprocessed and a graph was extracted from the images. The first step of the preprocessing extracts the edges from the image (figure 6). After the edges were extracted, vertices were identified. Vertices are placed at T intersections in the edge image and at places of high-edge curvature along the edges. Two vertices are adjacent if they are connected by an edge in the edge image.

The extracted graphs were the input to the sieve algorithm; the results are displayed in Figure 5. The white dots represent vertices that are correctly matched in both images. In evaluating the results, we look at the vertices that belong to the geometric pattern and the vertices in the background If the transformation and correspondence are to be correctly recovered, then the vertices in the background should remain unmatched. This hypothesis is confirmed by this experiment. In image *A*, 51 vertices were extracted by the preprocessing algorithm. Of these vertices, 44 were matched to the correct vertex in image *B*, with no false matches. Thus, enough vertices were matched to recover the transformation and align the geometric pattern.

6 Experiment: Object Recognition

In a third experiment, we demonstrate the sieve algorithm by recognizing objects. The objects were placed on a light table and the images acquired with a CCD camera. The objects—fork, knife, and cardboard cutout—in figures 7b, 7c, and 7d served as the models in the library. The sieve algorithm was then given a fourth image (figure 7a) with the three objects in it and asked to locate and identify the three objects.

Before being processed by the sieve algorithm, the images were preprocessed and a graph was extracted from the images. The first step of the preprocessing extracts the edges from the image. After the edges were extracted, vertices were identified. Vertices are placed at T intersections in the edge image and at places of high-edge curvature along the edges. Two vertices are adjacent if they are connected by an edge in the edge image.

The extracted graphs were the input to the sieve algorithm. The algorithm successfully located and identified the three objects; the results are displayed in figure 7 The white dots represent vertices that are correctly matched with the corresponding object, e.g., the white dots on the fork in figure 7a match the image in figure 7d Figure 1 shows two-edge images of a paper clip that were successfully matched. The graph in figure 1 was run against the image in figure 7a. The sieve algorithm did not report any matches, indicating that the paper clip was not in the image. This experiment shows that the sieve algorithm is capable of locating and recognizing objects in an image, and reporting when an object is not in the image.

7 Summary

In this paper, we have presented a new approach to solving the correspondence problem. The approach combines point-set and graph matching, and uses a sieve process to efficiently sieve through the space of all possible mappings between the features in each image. We successfully demonstrated the algorithm on aligning micrographs and images of geometric patterns, and locating and identifying objects.

There are a number of future research directions to pursue in this area. One is to explicitly incorporate the prepro-

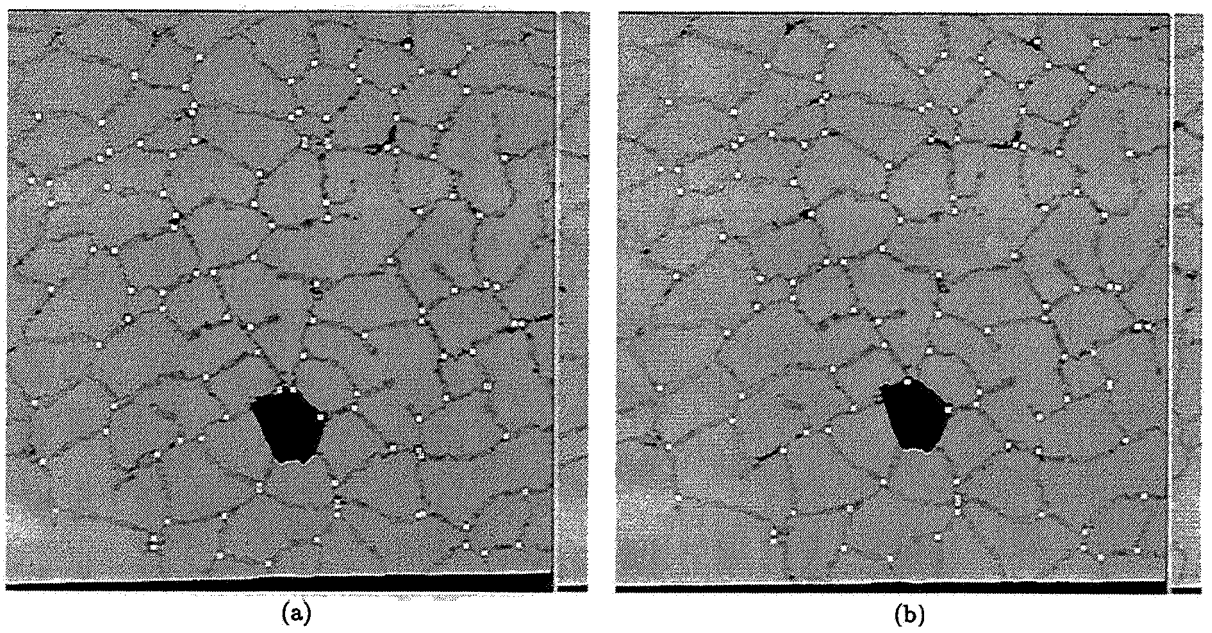


Figure 3: The results of matching two stereo micrographs. The vertices that were matched are marked as white squares. (a) Image A and (b) Image *B*

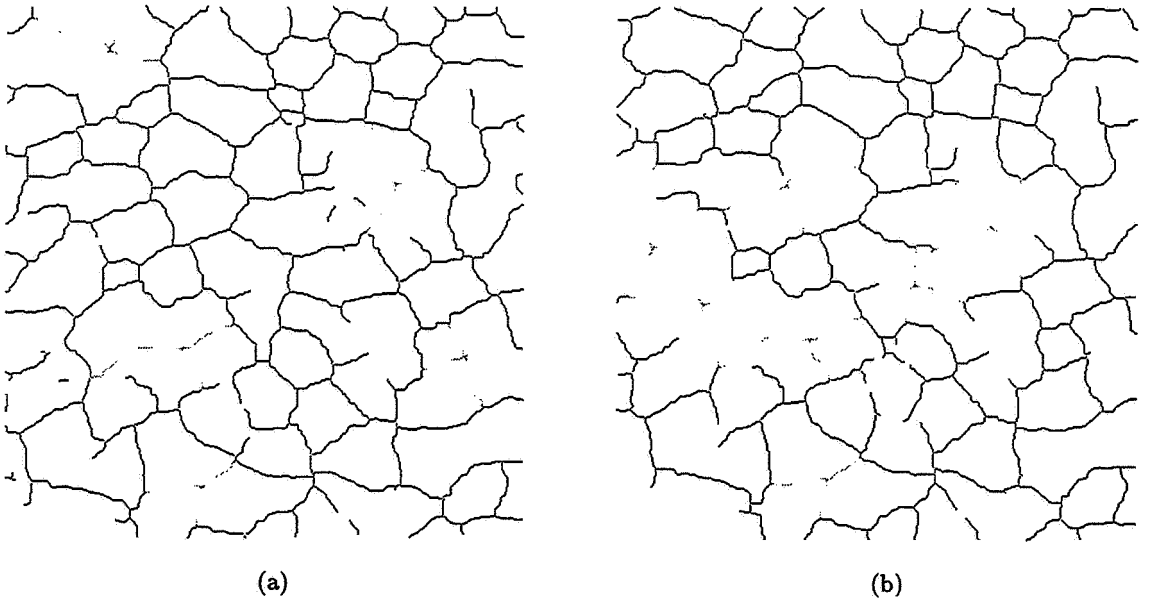


Figure 4 The graph structure in (a) Image A and (b) Image B. Edges adjacent to GREEN vertices are gray

Table 1 Performance on muscle T-system micrographs

Image set	Number of nodes in first image	Number of nodes in second image	Number of matches	Number of correct matches	Multiple matches	Number of incorrect matches
1	150	146	126	115	7	4
2	148	161	127	116	6	5
3	118	129	97	88	1	8
4	145	142	104	87	6	11
5	133	131	123	115	2	6

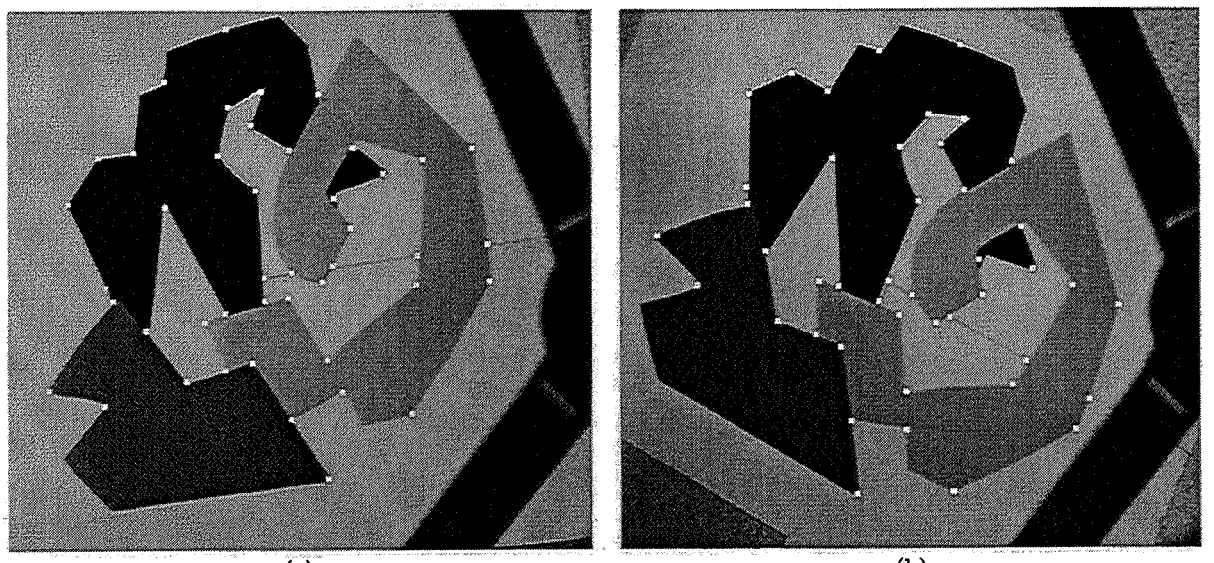


Figure 5: Geometric patterns: (a) Image *A* and (b) Image *B*

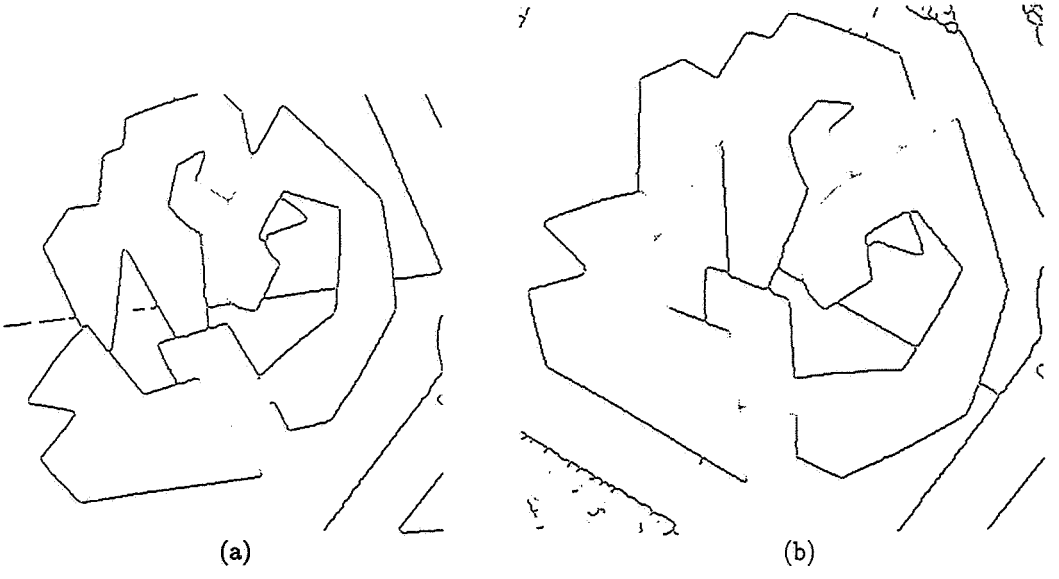


Figure 6: Edge image extracted **from** (a) Image *A* and (b) Image *B* Edges adjacent to GREEN vertices are gray.

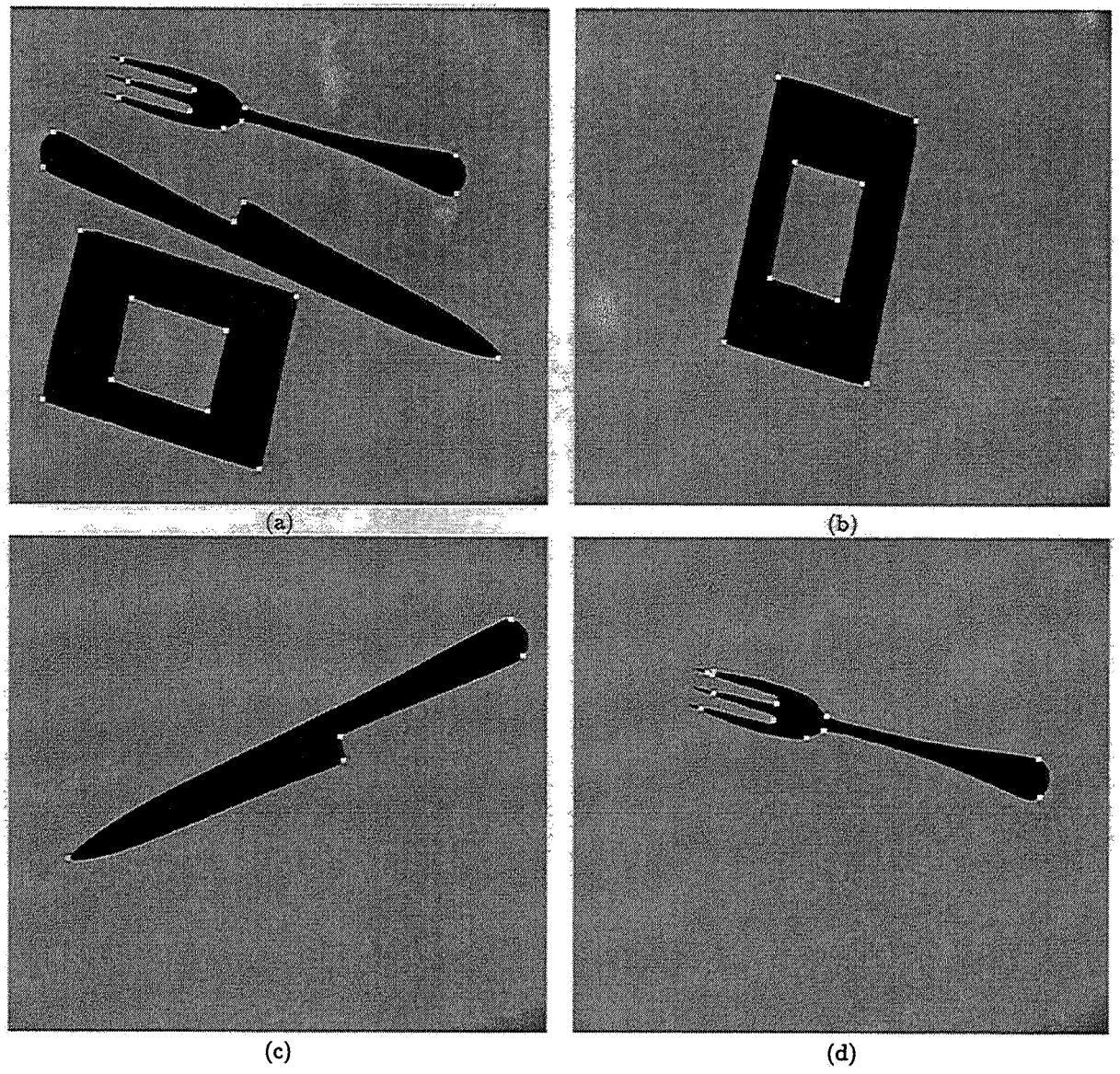


Figure 7 (a) Image with three objects that are identified. (b), (c), and (d) Images of the objects in the library. The white dots represent vertices that are correctly matched with the corresponding object.

cessing routine that extracts the graph into the algorithm. This would increase the robustness of the algorithm to noise. Another is to find correspondences and register images of the same scene taken with different sensors. This problem is of particular interest to the medical community.

8 Appendix

Finding the global minimum for the point-set distance function is a computationally expensive procedure. To circumvent this difficulty we use a fixed-point algorithm, and in this appendix we prove that the fixed-point algorithm converges. The proof proceeds in a number of steps. The first step shows that only a small fraction of the set of functions from A to B needs to be examined. This is the case whether one is searching for the global or a local minimum. In the process of showing that only a limited number of functions need to be examined, we show how all such functions can be generated. The allowable set of transformations is a function of three parameters. We determine the set of interesting functions by partitioning the parameter space. Each cell in the partition corresponds to an interesting mapping. The second step characterizes the fixed-point algorithm as a walk through the cells in the partition of parameter space. This characterization is used to prove that the fixed-point algorithm will converge to a fixed point from any initial set of parameters.

As before, let $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_n\}$ be two point sets in the plane, where at most three points are co-circular. Let $dist(\cdot,\cdot)$ be the distance between two points and $d_2(\cdot,\cdot)$ be the Euclidean distance between two points. The development of the partition of parameter space makes use of the *Voronoi diagram* of a set of points, where $Vor(A)$ denotes the Voronoi diagram of the point set A. The Voronoi diagram is a partition of the plane into a set of cells with a cell C_i for each point a_i in A. The cell C_i consists of the region of the plane that is closer to a_i than any other point in A, e.g., $x \in C_i$ if and only if $d_2(x, a_i) < d_2(x, a_j)$ for all $j \neq i$. The set of boundary points between the cells is a planar graph. A point is a vertex in the graph if the point is a boundary point between three or more cells. The arcs in the graph are line segments between vertices. The number of vertices and arcs is linear in the number of points in the point set [14]. In addition, each cell is convex.

Initially the set of allowable transformations $\mathcal{T}$ will be restricted to translation in the plane. (The extension to rotation is straightforward.) The first step is to characterize the optimal mapping given a translation $t \in \mathcal{T}$. Consider $Vor(-B \oplus a_i)$, where $-B \oplus a_i = \{-b_1 + a_i, \ldots, -b_n + a_i\}$. Each cell C_{ij} in $Vor(-B \oplus a_i)$ is generated by $a_i - b_j$, and for all translations t in the cell C_{ij}, a_i is mapped to b_j. Let D be the decomposition of the plane created by overlaying all m Voronoi diagrams of the form $Vor(-B \oplus a_i)$.

Each cell in D corresponds to a different mapping between points in A and B. The mapping is determined by looking at the cells C_{ij} that contain a cell in D. If a cell in C_{ij} contains cell D_k from D, then a_i is mapped to b_j. By the construction of D, each cell in D is only contained in one cell from each $Vor(-B \oplus a_i)$. This decomposition also generates all mappings between A and B that need to be examined. Suppose that this is not the case. Let $\ell \in \mathcal{L}$ be a mapping that does not correspond to a cell in D. Compute the transformation T that minimizes $D_p^M(\ell)$, where $D_p^w(\ell)$ is the point-set distance given the matching ℓ. The transformation T corresponds to a cell in D, and let ℓ_T be the mapping generated by this cell. Now, $dist(T(a_i), \ell_T(a_i)) \leq dist(T(a_i), \ell(a_i))$ because $d_2(T(a_i), \ell_T(a_i)) \leq d_2(T(a_i), \ell(a_i))$. Hence, $D_p^*(T, \ell_T) \leq D_p^*(T, \ell)$ because of the definition of the function $f(dist(\cdot,\cdot), \ldots, dist(\cdot,\cdot))$.

The image-formation model has three parameters: two of the parameters are translation and the third is for rotation about the z-axis. In the model, there is also rotation about the y-axis by a known angle β. Let $R_\gamma p$ denote the transform of rotating the point a by γ about the z-axis and by β about the y-axis, where β is known and fixed. The Voronoi diagrams $Vor(-B \oplus R_\gamma a_i)$ are the diagrams that result from transforming A by R_γ. Let D_γ be the decomposition induced by the Voronoi diagrams $Vor(-B \oplus R_\gamma a_i)$. Using D_γ the point-set matching problem can be solved for a fixed γ. The approach to partitioning the parameter space for the full model is to transform A, which will produce movement in $Vor(-B \oplus R_\gamma a_i)$ as γ. As the γ varies there will topological changes in D_γ. These changes will occur as vertices and segments from different $Vor(-B \oplus R_\gamma a_i)$ intersect and form new cells in D_γ. Each of these new cells corresponds to a new mapping between A and B. As γ sweeps between $[0, 2\pi)$, a partition of the parameter space is created. Let D_R be this decomposition. As in the translation only case, the optimal mapping corresponds to one of the cells in the decomposition.

The final step is to use the decompositions D and D_R to prove that the fixed-point algorithm converges to a fixed-point. To show this we need to define a fixed-point graph. The fixed-point graph is a directed graph with loops allowed, where a loop is an arc from a vertex to itself. The graph for the translation case is defined first. As before, let D be the decomposition of the plane induced by the Voronoi diagrams $Vor(-B \oplus a_i)$. Let G denote the fixed-point graph. Each vertex in G corresponds to a cell in D. If C_i is a cell in D, then v_i is the corresponding vertex in G. A cell C_i determines a matching ℓ_i and this matching produces a minimum and a corresponding translation t_i. The translation t_i is contained in some cell C_j (note: j may or may not be equal to i). The arc from v_i to v_j is an arc in G if and only if $t_i \in C_j$. So that, the fixed-point procedure is well defined, if the translation t is contained in a boundary of D, then t is assigned to the adjacent cell with the lexicographically smallest matching.

The method for defining the fixed-point graph for the full transformation is the same as the translation only case, except that the decomposition of the parameter space needs to be appropriately defined. In the translation and rotation case there are two translation parameters and one rotation parameter $\gamma \in [0, 2\pi)$. This decomposition, D_R, is constructed by starting with D_0 and sweeping the Voronoi diagrams $Vor(-B \oplus R_\gamma a_i)$ through the parameter θ. The decomposition D_γ is the slice of D_R when the rotation parameter equals θ.

Let the graph G^* be the same as G with the loops removed. Call a vertex v_i of G^* absorbing if it has out degree zero, where the out degree of a vertex vi is the number of arcs of the form (v_i, v_j). The relation between G^* and the fixed-point method is contained in the following theorem.

Theorem *2 Let G^* be a fixed-point graph with the loops removed. Then the following holds:*

1. *The graph G^* is acyclic.*
2. *There is a one-to-one correspondence between fixed points and absorbing vertices.*

Proof: Assume that G^* is not acyclic. Let the cycle have length k, and without loss of generality assume that the

cycle consists of vertices of $v_1, \ldots, v_k$ in that order Let $D_P^M(i)$ denote the minimum point- set distance given the matching corresponding to cell C_i and T_i, the transformation that achieves that matching. Because $T_i \in C_{i+1}$, the value of $D_P^M(i) > D_p(T_i) \geq D_p^M(i+1)$ Taking the string of inequalities around the cycle implies that $D_p(1) > D_p(1)$, a contradiction. The one-to-one correspondence between fixed points and absorbing vertices is a direct consequence of the definition of the arcs in G^* and the acyclic nature of G' ∎

Because G' is acyclic, the number of fixed points is equal to the number of connected components in G^* Another consequence of G^* being acyclic is that starting with any t or γ, one will always converge to a fixed point.

Acknowledgements

Portions of the work in this paper are taken from Jonathon Phillips' PhD dissertation. Dr Phillips thanks the Operations Research program at RUTCOR, Rutgers University, for the program's support of a nontraditional operations research dissertation. Furthermore, he gratefully thanks his thesis adviser, Yehuda Vardi, for his time, guidance, support, and many helpful conversations. Prof. Vardi's National Science Foundation Grant is acknowledged. Stan Dunn and Junqing Huang acknowledge the support of National Institute of Health through the Mid Atlantic IVEM resource.

References

[1] N Ayache and B. Faverjon. Efficient registration of stereo images by matching graph description of edge segments. ***International Journal of Computer Vision,*** 1(2):107–131, 1987

[2] H S. Baird. ***Model-Based Image Matching Using Location.*** MIT Press, Cambridge, Massachusetts, 1985.

[3] R. Horaud and T Skordas. Stereo correspondence through feature grouping and maximal cliques. ***IEEE Trans PAMI,*** 11(11):1168–1180, 1989.

[4] J. Huang, S. Dunn, S. Wiener, and P DeCosta. A method for detecting correspondences in stereo pairs of electron micrographs of networks. ***J. of Computer-Assisted Microscopy,*** 6(2), 1994.

[5] R. A. Hummel and S. W Zucker On the foundations of relaxation processes. ***IEEE Trans. PAMI,*** 5:267–287, 1983.

[6] D. P Huttenlocher, K. Kedem, and J M Kleinberg. On dynamic voronoi diagrams and the minimum hausdorff distance for sets of point sets under euclidean motion in the plane. In ***Eighth ACM Symposium on Computational Geometry,*** pages 110–119, 1992.

[7] D P Huttenlocher, G. A Klanderman, and W J Rucklidge. Comparing images using the Hausdorff ***distance.*** ***IEEE Trans. PAMI,*** 15(9):850–863, 1993.

[8] Y Lamdan and H Wolfson. Geometric hashing: a general and efficient model-based recognition scheme. In ***Second International Conference on Computer Vision,*** 1988.

[9] G. Medioni and R. Nevetia. Segment-based stereo matching. ***Computer Vision, Graphics, and Image Processing,*** 31:2–18, 1985.

[10] F Mokhtarian and H. Murse. Shilouette-based object matching through curvature scale space. In ***Fourth International Conference on Computer Vision,*** pages 269–274, 1993.

[11] L. Peachy and J. Heath. Reconstruction from stereo and multiple tilt electron microscope images of thick sections of embedded biological specimens using computer graphics methods. ***J. Microsc.,*** 153, 1989,

[12] P J. Phillips. ***Representation and Registration in Face Recognition and Medical Imaging.*** PhD thesis, RUTCOR, Rutgers University, 1996.

[13] P J. Phillips, J Huang, and S. M Dunn. An efficient micrograph registration algorithm via sieve processes. ***J of Computer-Assisted Microscopy,*** 8(1), 1996.

[14] F P Preparata and M. I. Shamos. ***Computational Geometry.*** Springer-Verlag, New York, 1985.

[15] R. C. Wilson and E. R. Hancock. Relational matching with dynamic graph structure. In ***Fifth International Conference on Computer Vision,*** pages 450-456, 1995.

[16] K Zikan. ***The Theory and Applications of Algebraic Metric Spaces.*** PhD thesis, Department of Operations Research, Stanford University, 1989. (Research Report 579, Artificial Intelligence Series, No. 9, Hughes Aircraft Company Research Laboratories, 1989.).

Alignment of functional and anatomical tomograms based on automated and real-time interactive procedures

Uwe Pietrzyk, Alexander Thiel, Helmut Lucht and Alexander Schuster
Max-Planck-Institut Für Neurologische Forschung, Lindenthal, Germany

INTRODUCTION

Image registration is an inevitable prerequisite for modern brain mapping techniques like the analysis of brain activation studies either with PET or fMRI. It is essentially the correlation of functional images with the underlying anatomy where image registration is required. Different cases can be distinguished, like:

- multiple acquisitions of a single modality
  ----> intramodality matching
- multiple acquisitions of a single subject
  ----> intraindividual matching
- acquisition with different modalities
  ----> intermodality matching
- acquisition of different subjects
  ----> intersubject matching but also
- reorientation of multiple subjects to a standard reference system
  ----> interindivdual intersubject matching

Differentiation is required whether multiple modalities are employed and whether a single subject or different subjects are studied. Usually, multiple separate procedures are to be employed in order to

- register multiple acquisitions of a single modality (PET, fMRI); (itra-individual matching)
- reorient *MR* images to an anatomical standard position (i.e., Talairach coordinate system)
- register individual PET measurements of different subjects to reoriented MR images in an anatomical standard position; (interindividual registration).

This provides the opportunity to compare activation studies in different subjects.

Since especially PET activation studies are acquired as a series of separate measurements, each measurement has to be aligned first to a common reference system prior to calculation of activation images. Frequently, the first PET study servers as reference-In a second step, after averaging across all registered PET studies this mean image is aligned to the anatomical image (MRI). The realignment of each PET study separately to the MRI study is not practicable because a single O-15-water PET study is too noisy to yield save, accurate alignment.

Hence, two interpolation steps occur within this analysis: firstly, following intramodality, intrasubject registration (PET vs. PET), and secondly, following intermodality, intrasubject registration (PET vs. MRI in standard position). Repeated interpolations, however, have a disadvantage as they act like a smoothing filter To avoid any additional smoothing and filtering of the data a procedure has been developed to work with a single interpolation. Also, it was attempted to combine the various types of registration as defined above in a single analysis step avoiding any additional data handling.

MATERIALS and METHODS

DATA

Tomographic data usually come as image matrices (128x128 or 256x256 pixels) stacked on top of each other fully covering the brain with up to 160 slices. The relative position of the brain within this 3D volumetric data set might differ even for measurements performed within a short time. For this study, PET data obtained with an ECAT EXACT HR (CTI-Siemens, Knoxville TN, USA) were used. The images were reconstructed to 47 slices with 2 mm pixel size in a 128 x 128 pixel matrix and 3 125 mm interslice distance. O-15-water served as tracer to measure the pattern of cerebral blood flow. Up to 12 consecutive measurements were performed to study effect of changes in cerebral blood flow during functional activation, like repeating words, recognizing pictures or finger tipping. Acquisitions lasted approximately 10 min. each, yielding a period of two hours in total.

Anatomical reference studies were obtained with MRI of the same patients. A T1 weighted 3D-FLASH sequence was used on an Siemens Impact (Siemens Medical Systems, Erlangen, Germany). The studies comprised of 64 slices with 25 mm interslice distance and 1 mm pixel size on a 256 x 256 pixel matrix. All data were available in digital format.

PREPROCESSING and FILTERING

Before starting the process of realignment, i.e. finding the optimal transformation to be applied to one image to match the reference image, images must be preprocessed in order to eliminate noise

and to define the parts of the image, relevant for coregistration. This step is necessary for later calculation of similarity measures. In case of brain PET-images, this is done by applying a 2D smoothing filter with kernel size of 5 pixels and afterwards choosing a threshold which defines the brain-contour and eliminates extracranial structures. Usually a threshold of 30% image maximum is used for definition of the brain contour. The automated step of intraindividual realignment, described below, only uses supra-threshold pixels for calculation of similarity measures. It should be noted that this filtering is effective only during the realignment step and not during reslicing to obtain matched studies which are then stored onto disk.

REGISTRATION TECHNIQUES

While robust automated procedures exist for single modality alignment, registration of functional (PET) and anatomical (MRI) images cannot be performed completely automated without user interaction, like skull editing, i.e. removing non-brain structures, or adjusting the proper threshold to eliminate unwanted portions within the images. Also, the final results of any registration procedure are bound to be checked visually. Apart from automated procedures, interactive user driven procedures have been accepted by the scientific community

In order to combine the advantages of automated and interactive registration techniques, a hybrid technique is proposed here. Owing to the computing power available with desktop systems today, it was tempting to design a hybrid technique combining a previously established real-time interactive registration technique [1] with options of automated procedures. Some computational intense parts can be run automatically like the alignment of multiple individual PET scans exploring one of the established similarity measures (ratio image uniformity [2], sum of absolute differences [3], correlation coefficient [4]). The results can be immediately checked visually, The procedure is split up into 7 steps, from which 6 are handled by one program. Fig. 1 show the general data flow and the separate processes labeled by number 1 through 7

Step 1 represents the intra-modality, intra-subject alignment of multiple PET acquisitions (in this case 6 acquisitions). This part is run automated fashion.
During step 2 a mean PET is calculated by averaging across all individual PET acquisitions applying the realignment transformation obtained during step 1
Step 3 stands for reorienting the MR images to comply with a standard anatomical reference system like that of Talairach.
Step 4 represents the inter-modality, intra-subject registration of PET vs. MRI.
For the PET studies, transformation parameters from step 1 and step 4 will be combined and applied on each PET measurement in step 5.
This step, too, is performed automatically but subject to visual control.
Step 6 is performed in a separate procedure. So far there has been no difference in handling the PET data, whether the measurement was performed during an activation task or in a resting state. The activation images are calculated by subtracting 'Activation' and 'Rest' after the images have been filtered. Details of which can be found in [5].
Step 7 is performed again in the initial program. Here PET and MRI images are fused into a single image.
Following step 3, 5 and 6 the data usually is written onto a hard disk for further processing like brain surfacing rendering techniques.

SLICE SELECTION and OPTIMIZATION

The reference study has to be selected prior to starting the automated alignment process. During the automated alignment the selected similarity measure is calculated, however, not for all slices. Instead, this calculation is limited to preselected orthogonal cuts, transaxial, coronal and sagittal to constraint the registration procedure in three dimensions. The selection of these cuts may be critical as pointed out below Usually, a transaxial slice through the thalamus, an anterior or posterior coronal slice and a sagittal plan parallel to the hemispherical plane is used. The optimization consists of a simple search. It starts by varying the current transformation parameters within a preset intervals, reducing the step size with every repeated search through the 6 parameters. The search continues in the direction of optimal similarity at this point and stops at a step size of 0,1 mm and 0.1 degree, respectively. Gross misplacements (> 1 cm) may and should be handled interactively before the start of the automated optimization.

RESULTS

PROTOCOL

The following protocol proved practicable for the analysis of multiple scans from 0-15-water activation studies with PET and subsequent alignment with MR images:

1) Definition of filter size, threshold and choice of orthogonal cuts to be used for the calculation of similarity measures.

2) Run automated alignment of individual PET scans; storetransformationparameters

3) calculation of an average image across all individual realigned scans with improved contrast and reduced noise

4) reorientation of MRI images into a defined standard position(parallel to AC-PC plane) interactive realignment of mean PET images to reoriented MR images

5) combine transformation parameter fiom 1) (automated alignment of individual PET study) with those from step *3)* (interactive alignment to the MR images) in order to align each PET study to the reoriented MR images.

It should be noted, that the combined transformation parameters are applied to the original PET studies directly without intermediate interpolation or reslicing. Thus, the final PET data are in the same orientation as the reoriented MR images.

CHOICE of PARAMETERS

Standard values for filter parameters and threshold levels can be used for one image modality For O-15-water PET images these are: smoothing filter with a 5 pixel kernel and threshold of 25 - 30% of maximum. Higher thresholds include to few pixels thus the estimation of similarity measures becomes insecure and small deviations induce large effects in similarity measures. Furthermore, we experiencedthat an increase in filter size beyond 6 pixels did not yield improved results and comparatively large rotations show only small changes in similarity measures. The above mentioned parameters yielded good results for O-15-water images, but will need adjustment for different image modalities with different image statistics.

TIMING

The time needed for the alignment process of a typical 15-0-water(12PET images aligned to one MRI study) takes about 10 minutes, including visual control and manual correction if necessary This is fasterthan automated algorithms because the time consuming segmentation of the MRI scan is not necessary due to the interactive real-time registration ofthe mean PET image to the MRI scan.

MEDICAL ISSUES

The use of automated procedures is limited if the information content of the images changes rapidly between images, especially if many supra-threshold pixels appear in one image in a certain location and disappear in the following image. A patient who suffered fiom an attack of hayfever while being examined had a severe blood flow increase in the nasal mucosa. The mucosa was not visible in the first scans. These image pixels of high intensity could not be eliminated by filtering or thresholding and might prevent the automated algorithm from finding the correct position. Hence, it was necessary to interact manually, for instance, by selecting a different set of orthogonal slices, where the nasal mucosa is not present.

DISCUSSION and CONCLUSION

A hybrid technique for medical image registration was presented combining automated and interactive procedures. Subsequent transformation steps were combined into a single transformation during the alignment of several individual PET scans, whose average image was then aligned with respect to a *MR* image.

Thus, only a single reslicing of all PET and MR data is required avoiding multiple interpolations and averaging of the data. Visual control of all steps is guaranteed throughout the entire procedure, which can be performed within less then 10 min. including the final visual control.

REFERENCES

[1] Pietrzyk U, Herholz K, Fink G, et al (1994). An Interactive Technique for Three-Dimensional Image Registration: Validation for PET, SPECT, MRI and CTBrain Studies. J Nucl Med 1994;35:2011-2018.

[2] Woods RP, Cherry SR, Mazziotta JC (1992). Rapid automated algorithm for aligning and reslicing PET images. J Compute Assist Timor 1992;16:620-633.

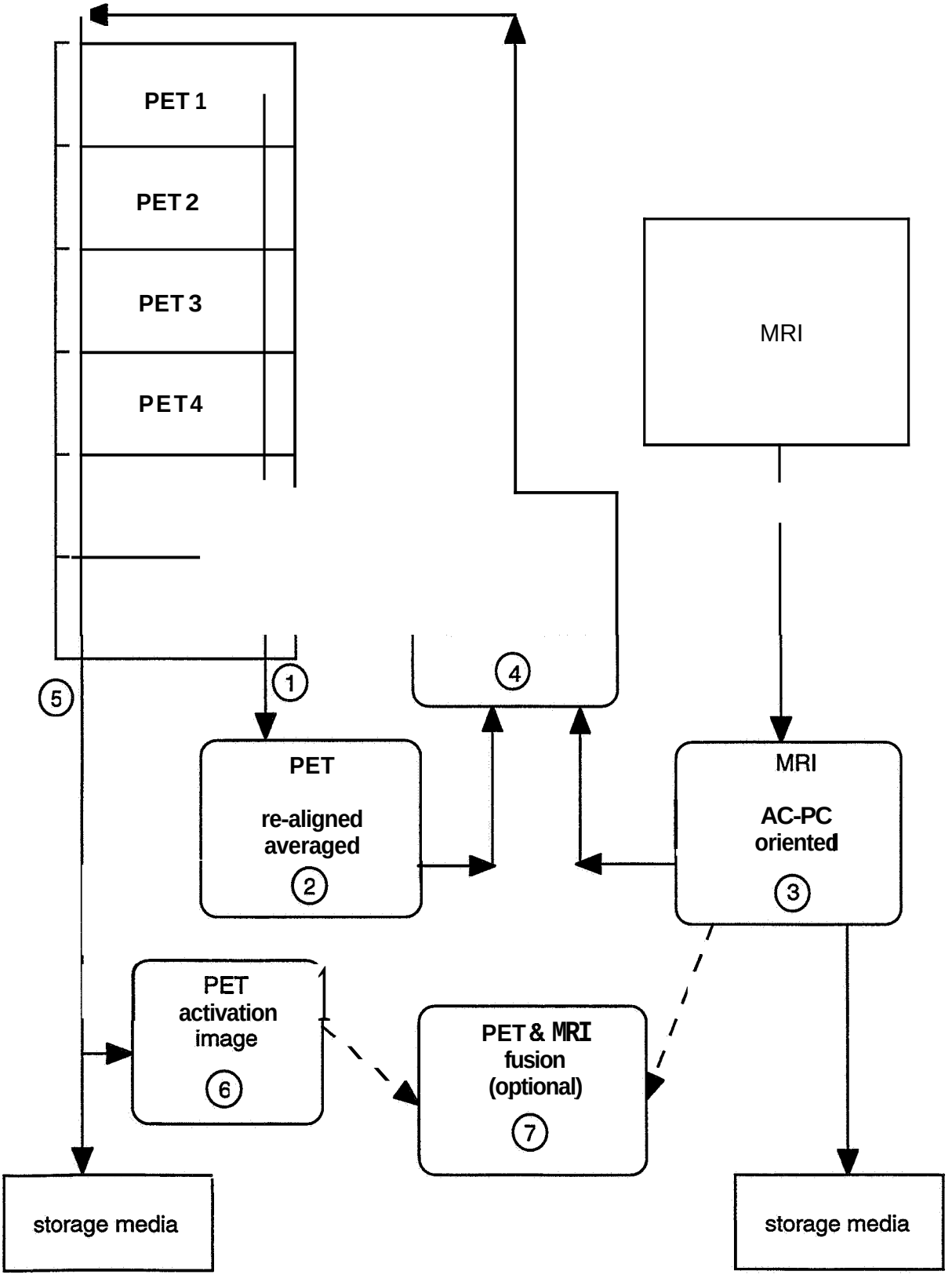
[3] Svedlow M, McGillem CD, Anuta PE (1978). Image registration. Similarity Measure and Preprocessing Method Comparisons. IEEE Trans Aeros Elect Syst 1978;14:141-149.

[4] Andersson JLR (1995). A rapid and accurate method to realign PET scans utilizing image edge information. J Nucl Med 1995; 36:657-669.

[5] Herholz K, Thiel A, Wienhard K et al (1996). Individual Functional Anatomy of Verb Generation. Neuroimage 1996; 3:185-194.

LEGENDS

Figure 1. Flow chart of the proposed hybrid image registration procedure using a combination of automated and interactive registration techniques. See text for explanation of individual steps.



Tracking Hurricane Paths*

Nagarajan Prabhakaran, Naphtali Rishe, and Rukshan Athauda

High Performance Database Research Center
School of Computer Science
Florida International University
University Park, Miami, FL 33199

Abstract

The South East coastal region experiences hurricane threat for almost six months in every year To improve the accuracy of hurricane forecasts, meteorologists would need the storm paths of both the present and the past. A hurricane path can be established if we could identify the correct position of the storm at different times right from its birth to the end.

We propose a method based on both spatial and temporal image correlations to locate the position of a storm from satellite images. During the hurricane season, the satellite images of the Atlantic ocean near the equator are examined for the hurricane presence. This is accomplished in two steps. In the first step, only segments with more than a particular value of cloud cover are selected for analysis. Next, we apply image-processing algorithms to test the presence of a hurricane eye in the segment. If the eye is found, the coordinate of the eye is recorded along with the time stamp of the segment. If the eye is not found, we examine adjacent segments for the existence of hurricane eye.

It is probable that more than one hurricane eye could be found fiom different segments of the same period. Hence, the above process is repeated till the entire potential area for hurricane birth is exhausted. The subsequent/previous position of each hurricane eye will be searched in the appropriate adjacent segments of the next/previous period to mark the hurricane path.

The temporal coherence and spatial coherence of the images are taken into account by our scheme in determining the segments and the associated periods required for analysis.

1. Introduction

A significant improvement in the prediction of natural calamities, have a profound social and economical impact on people. Among other natural events, hurricanes take longer time to shape up and move compared to earthquakes, tornadoes, etc. However, as hurricanes cover vast areas, a large number of people and pets would need to be relocated.

A correct prediction can save lives and helps to minimize property damages. In contrast, an incorrect prediction would cost a lot of effort and financial resources for the government and individuals. This demands an increased accuracy of hurricane predictions. Tools and methodologies used by meteorologists, play a vital role in the accuracy of weather predictions.

Remote sensing satellite images and aerial photography [2] provide immense amount of data for timely forecasting. Efficient processing [5] of these vast amounts of data is crucial to extract essential features for hurricane predictions.

In *this* paper, we attempt to provide an automated method to track hurricanes from their births through all stages. The next section presents the analysis of satellite images for hurricanes. Section 3 describes the method for determining hurricane paths. The last section outlines conclusive remarks and future work.

2. Analysis of satellite images

Weather satellite images are obtained in the form of area files (Geostationary Operational Environmental Satellite images). As these images

* This research was supported in part by NASA (under grants NAGW-4080, NAG5-5095 and NRA-97-MTPE-05), ARO (under grants DAH04-96-1-0049 and DAAH04-96-1-0278), NSF (under CREST, CDA, and RIA), and State of Florida.

cover wide areas of earth, they require large amount of memory for processing. For this reason, we decompose satellite images into segments of small size. The segment where there is a high probability of hurricane development is examined first. This will speed up the hurricane search process.

The structure of the process of tracking hurricane paths is shown in Figure 1 The process begins by selecting an image segment that is highly probable for hurricane birth. The segment is evaluated to assess the cloud coverage. This is carried out with image processing techniques (conversion to gray scale image, reduction in brightness and increase in contrast with no gamma correction or techniques similar to those outlined in [1]) as in Figure 2 and Figure 3 Subsequently the cloud coverage area and the distribution of clouds are computed.

If the cloud cover parameters of the segment are greater than a threshold value (that is dependent on the type of images and the segment size), the segment may be associated with a hurricane. If not, the segment is discarded and another segment is selected for hurricane presence test.

If a segment passes the cloud cover analysis, next we examine if a hurricane eye is present in the segment. We perform this by applying image smoothening techniques such as water color method to remove extraneous noise from the image (see Figure 4). Then we check if the segment has a closed region of black pixels present inside the segment. This is ensured with the application of a flood filled algorithm from all boundary pixels of the rectangular segment (replacing black pixels by another color) and then we search for the presence of black pixels inside the segment.

There could be multiple black regions that may be potential hurricane eyes (four regions in Figure 4). To identify if any of them qualify as the hurricane eye, the following tests are performed, First, the bounding box of each cloud region as well as each black region is computed. The height and width of these boxes are evaluated for the presence of the cloud size indicative of a hurricane and the black region indicative of its eye. If multiple black regions within a cloud region pass the above test, we calculate the relative position of these regions with respect to the cloud region and select the black region that is closest to the center of the cloud, as the hurricane eye. The coordinates of the hurricane eye along with the time stamp of the segment are stored in a database.

If the segment contains a potential cloud region at the border of the segment with no hurricane eye, the eye could be present in a spatially adjacent image segment. In such cases, the new segment is defined to enclose all parts of the cloud region from the adjoining segments. Then the hurricane eye presence test is repeated for the new segment.

3. Calculation of trajectories

To determine the subsequent temporally coherent position of the already identified hurricane eye, adjacent segments of spatial and temporal coherence [4] are examined. The temporal and geographical features of the eye are stored in a database along with other hurricane-related parameters that are monitored by other sources. The trajectory of the hurricane path [3] is interpolated between temporally successive eye positions.

The process of finding the eye and the trajectory calculation is repeated temporally and spatially until the eye cannot be identified at which the hurricane has lost its strength. This completes the calculation of the hurricane trajectory The trajectories of past hurricane paths could be plotted along with the current trajectory for the comparison of hurricane paths.

4. Conclusion

In this paper, we have presented a scheme to analyze satellite images to automatically locate hurricanes and their eyes by exploiting both spatial coherence and temporal coherence. Storage of images of all past hurricanes demands enormous amount of storage space. Demand based access and analysis of these image data sets are computationally expensive. Hence our method would analyze the data once and store only the trajectory and the hurricane associated parameters.

The method described here will not provide an accurate prediction, but enables forecasters to account visually presented information of past hurricanes that are spatially and temporally relevant to a present hurricane. With sophisticated hardware support, real-time analysis of hurricane position could be computed to furnish timely information for emergency rescuers.

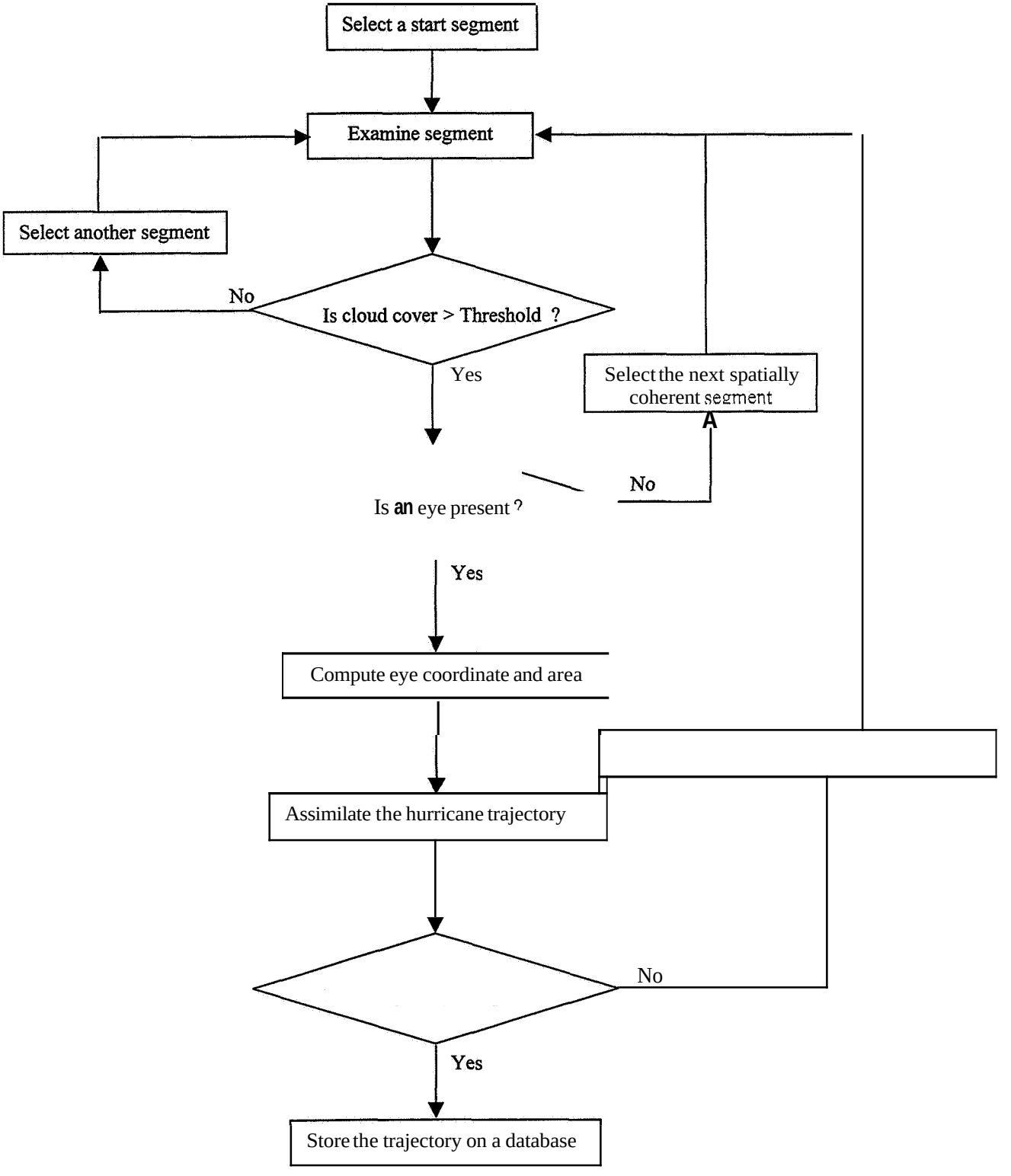


Figure 1. Proposed structure for tracking hurricane paths

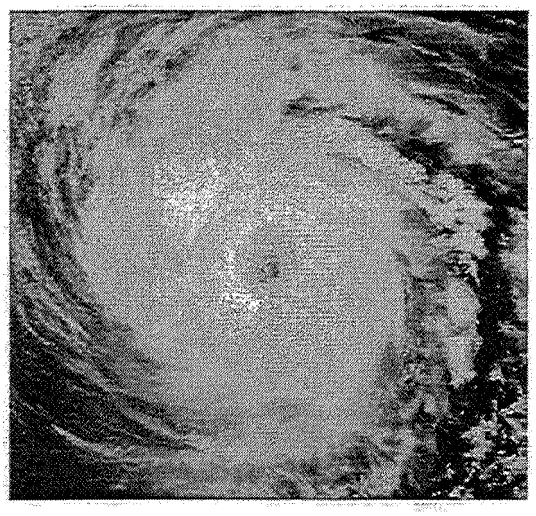


Figure 2. Gray scale image of a segment with a hurricane eye

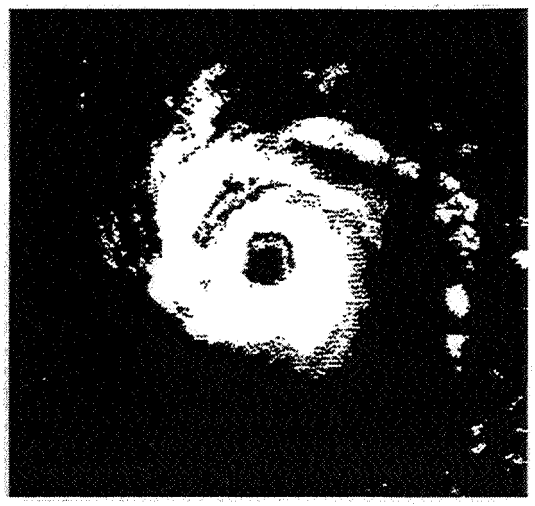


Figure 3. Image after brightness and contrast manipulations

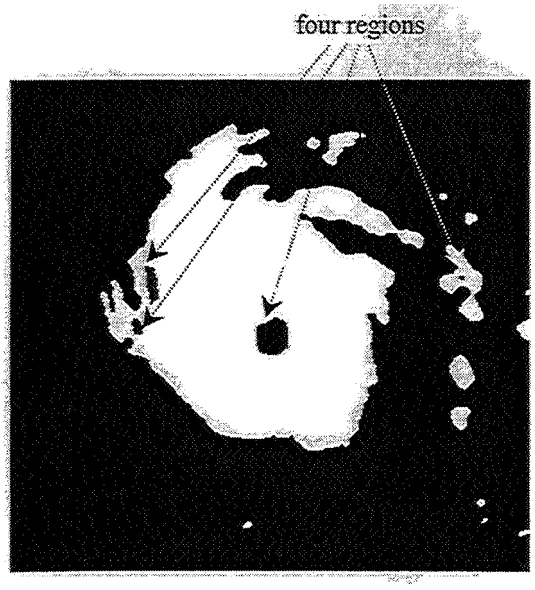


Figure 4. Image after smoothening noise

References

1. Gong, Y and Sakauchi, M. (1995) "Detection of regions matching specified chromatic features", Computer Vision and Image Understanding, 61 (2), pp. 263-269
2. Matsuyama, T and Hwang V.S. (1990) "SIGMA – A knowledge-based aerial image understanding system", Plenum Press, New York.
3. Meyer, F.G and Bouthemy, P (1994) "Region-based tracking using affine motion models in long image sequences", CVGIP: Image Understanding, 60 (2), pp. 119-140.
4. Pla, F and Marchant, J.A. (1997) "Matching feature points in image sequences through a region-based method", Computer Vision and Image Understanding, 66 (3), pp. 271-285
5. Watkins, C., Sadun, A. and Marenka S. (1993) "Modern image processing: warping, morphing, and classical techniques", Academic Press Professional, New York.

Registration of Video Sequences from Multiple Sensors

Ravi K. Sharma and Misha Pavel

Department of Electrical Engineering
Oregon Graduate Institute of Science and Technology
P.O. Box 91000, Portland, OR 97291-1000, USA

Abstract

In this paper, we describe an approach for registration of video sequences from a suite of multiple sensors including television, infrared and radar Video sequences generated by these sensors may contain abrupt changes in local contrast and inconsistent image features, which pose additional difficulties for registration. Our approach to registration addresses the difficulties caused by using multiple sensors. We use a representation for registration that is invariant to local contrast changes, followed by smoothing of the resulting error measure used for registration, for robust estimation of registration parameters. We use an iterative procedure to reduce the effect of inconsistent features. Finally, we describe a method that uses same-sensor registration to aide in performing registration of sequences of video frames across multiple sensors

1 Introduction

Image registration is a critical component of multi-sensor systems that typically use different types of sensors. An example of such a system is a fusion system which aides in vehicular navigation [11] by combining image information contained in video sequences generated by television (TV), forward looking infrared (FLIR) and millimeter wave radar (MMWR). The phrase *video sequence* or *vrdeo stream* refers to the sequence of visual images generated by an imaging sensor For performing fusion, it is necessary to ensure that the imagery from multiple sensors is registered or aligned so that different video sequences can be compared and combined.

Several techniques have been developed in the past to register images from the same sensor or images from similar types of sensors [4]. These techniques can be broadly categorized as optical flow techniques and feature mapping techniques. Optical flow techniques [8, 10, 1, 7] perform registration by estimating the optical flow field that represents the motion between images, based on the relation between motion and image brightness. Feature mapping techniques locate corresponding features such as corners and contour lines in the images [12, 9]. The images are then regstered by using the locations of these features as control points and finding the mapping that aligns these points.

Video streams captured through different types of sensors pose additional difficulties to traditional registration techniques. The gray-scale characteristics of such video streams may differ locally. For example, there can be abrupt local changes in the polarity of contrast (local polarity reversals) between TV and FLIR imagery, Local polarity reversals can degrade the performance of optical flow techniques which assume brightness constancy. The feature extraction task in feature based registration techniques can be impaired because features present in one video stream may not be present in the other stream (inconsistent or complementary features), e.g. TV and radar imagery

In this paper, we present an approach for registration of video sequences from multiple sensors based on extending the optical flow based techniques. In Section 2, we review a technique for registering images from the same sensor or similar types of sensors. Our approach consists of four stages that address the difficulties caused by multiple sensors by modifying the technique discussed in Section 2. To reduce the errors in registration caused by gray-scale disparities, we use a non-linear representation that is invariant to local polarity reversals. This is described in Section 3. In the second stage, described in Section 4, we modify the error measure used for registration to facilitate robust estimation of registration parameters. The third stage consists of an iterative process to reduce errors in registration caused by complementary features, and is described in Section 5. The final stage is described in Section 6 and consists of using same-sensor registration to aide in registration of sequences of multi-sensor video frames. Finally, experimental results are shown in Section 7

2 Gradient based approach for registration

We now review an algorithm [3, 2, 7] for registration of images obtained from similar types of sensors. This algorithm estimates the optical flow or motion field between the images using a gradient based approach. Optical flow is the apparent motion of brightness patterns between images. In this approach, the observed brightness of objects in the images is assumed to be constant over the images to be registered. This assumption gives the relationship between the motion and the gradient of image brightness. In addition, a smoothness constraint is assumed to ensure that the motion of nearby points is similar If the motion in the images consists of global motion, then the smoothness constraint can be imposed by assuming a global motion model.

Global motion [7] can be defined by a set of registration parameters $p = [p_x, p_y]$, which describe the motion field [3]. If I_1 and I_2 are the two images to be registered, then $I_2(x,y)$ can be expressed in terms of $I_1(x,y)$ as

$$I_2(x,y) = I_1(x + p_x, y + p_y) \quad (1)$$

where $p_x(x,y)$ and $p_y(x,y)$ depend upon the model used to describe the global motion. If the captured scene is assumed to be a planar surface, the motion of the scene can be approximated by a projective transformation, which is given by

$$\begin{aligned} p_x(x,y) &= p_1 + p_3 x + p_5 y + p_7 x^2 + p_8 xy \\ p_y(x,y) &= p_2 + p_4 x + p_6 y + p_7 xy + p_8 y^2 \end{aligned} \quad (2)$$

where $[p_1, .., p_8]$ are the parameters of the global motion model. When the distance between the scene and the camera is large, the motion can be approximated by an affine model, with $p_7 = p_8 = 0$ in Equation 2. Translation is a special case of the projective model when $p_3 = p_6 = 1$ and $p_4 = p_5 = p_7 = p_8 = 0$ in Equation 2.

The registration parameters are then obtained by minimizing the sum squared error,

$$E = \sum_{x,y} (I_2(x,y) - I_1(x + p_x, y + p_y))^2 \quad (3)$$

If the displacements $[p_x, p_y]$ are small, the above equation can be simplified by a Taylor series approximation of $I_1(x + p_x, y + p_y)$,

$$I_1(x+p_x, y+p_y) = I_1(x,y) + p_x I_{1x}(x,y) + p_y I_{1y}(x,y) \quad (4)$$

where

$$I_{1x} = \frac{\partial I_1(x,y)}{\partial x}, \quad (5)$$

$$I_{1y} = \frac{\partial I_1(x,y)}{\partial y} \quad (6)$$

The sum squared error can now be expressed as

$$E = \sum_{x,y} (\Delta I - p_x I_{1x} - p_y I_{1y})^2 \quad (7)$$

where,

$$\Delta I = I_2(x,y) - I_1(x,y) \quad (8)$$

The motion between the images is obtained by setting the derivatives of the error measure in Equation 7 with respect to the parameters of the motion model to zero and solving the resulting set of equations. Such an estimation is accurate only when the displacements between the images are a fraction of a pixel, so that the Taylor series approximation holds. The estimation can be improved by using an iterative alignment procedure that warps the images according to the initial estimates and then repeats the estimation to obtain residual displacements.

To capture large displacements, a coarse-to-fine refinement of the registration parameters is performed in a multi-resolution framework such as a Gaussian or Laplacian pyramid scheme [6]. The registration parameters are initially estimated at a lower resolution to obtain a coarse solution. This solution is then used to initialize the parameter estimation at a higher resolution to obtain more accurate parameters.

3 Invariant representation for registration

The algorithm described in Section 2 cannot be directly applied if I_1 and I_2 are multi-sensor images. The gray-scale characteristics of video streams from different types of sensors are different. The assumption of brightness constancy is therefore not always valid. Hence, the motion between the images cannot always be described by the apparent motion of brightness patterns. Sometimes, images from different types of sensors contain local polarity reversals which are abrupt changes in the polarity of local contrast. This means that an error measure such as the sum of squared differences which was used in the previous section may not necessarily be minimum even when the images are perfectly registered.

Figure 1(a), shows an example of the error for a pair of images that are mis-registered due to translational motion. In this case, the point of correct registration gives the maximum mor To address the problems caused by gray-scale disparities and local polarity reversals, we need to transform the images into a representation that is invariant to changes in brightness and invariant to local polarity reversals. The particular transformation that we have chosen is the absolute value of pixels in a Laplacian pyramid, The output of this transformation is a non-linear representation that results in identical representations for signals which are exact opposites of each other An advantage of using the Laplacian pyramid is that its successive levels are band-passed versions of the original signal. The band-pass operation ensures that the low spatial frequencies which contain the information about gray-scale disparities, are removed. The error measure *can* now be expressed as

$$E = \sum_{x,y} (G_2'[I_2](x,y) - G_1'[I_1](x + p_x, y + p_y))^2 \quad (9)$$

where G' denotes the non-linear representation.

Figure 1(b) shows the error surface obtained by using the non-linear representation. Since the gray-scale disparities are reduced, the minimum point on the error surface now corresponds to the parameters that give the correct registration.

The non-linear representation that we have used results in a transformation of the image intensities into features, that are invariant to gray-scale disparities. We note that well chosen invariant representations will allow the registration algorithm to benefit from both approaches – feature and motion based registration

4 Estimation of Registration Parameters

The registration parameters, $[p_1, .., p_8]$ can now be estimated by minimizing the modified error measure in Equation 9. The accuracy and convergence of the minimization depends upon the error surface. Typically, the error surface is noisy when multi-sensor imagery is used, especially at higher resolutions. This gives rise to local minima, which may lead to erroneous registration or non-convergence of the minimization procedure. This suggests that smoothing of the error surface to reduce the effect of local minima can improve parameter estimation. Smoothing of the error surface can be achieved by applying a low-pass filter to the non-linear absolute Laplacian representation. The low-pass filter operation removes the high-spatial frequency noise and causes a spatial spread of features in the absolute Laplacian representation. Figure 1(c) shows the error surface after smoothing. The error surface is smoother and the basin containing the global minimum is widened. The trade-off in performing a low-pass filtering operation is that the accuracy of the registration will be somewhat reduced since the finer details are removed by filtering. On the other hand, the possibility of erroneous matches is also reduced.

At this stage the error measure can be expressed in terms of the non-linear representation and the smoothing operation as,

$$E = \sum_{x,y} (G_2[I_2](x,y) - G_1[I_1](x + p_x, y + p_y))^2 \quad (10)$$

where G denotes the combined operations of the non-linear representation and the smoothing. The block diagram illustrating the steps in Sections 2, 3 and 4 using a multi-resolution framework is shown in Figure 2.

5 Reducing Errors caused by Complementary Features

Sometimes images from different types of sensors may contain different features. Features present in one image may be totally absent in the other image. We call such unique or sensor-specific features as complementary features. Complementary features can cause errors in registration if their impact on the error measure is significantly large.

We now describe an iterative procedure to reduce the errors caused by complementary features. The initial step consists of registering the images using the approach described in the previous sections. We then compute a discrepancy image W that represents the error due to complementary features and the error due to the remaining mis-registration. To compute W, we estimate a local, pixel by pixel, gray-level transformation between the images:

$$I_2(x,y) = \beta(x,y) I_1(x + p_x, y + p_y) + \alpha(x,y) \quad (11)$$

This transformation accounts for local polarity reversals between the two images. The parameters β and α are obtained by computing correlations in small (5 x 5) regions of the images. W is the squared difference after the transformation,

$$W(x,y) = (I_2(x,y) - \beta(x,y) I_1(x + p_x, y + p_y) - \alpha(x,y))^2 \quad (12)$$

This discrepancy image is then transformed into a Laplacian pyramid representation to obtain the discrepancy pyramid. We identify the highest ten percent of values in the two highest resolutions of the discrepancy pyramid and ignore these regions while computing the error in Equation 10. We assume that these regions contain the complementary features that cannot be matched. We then repeat the registration process with the new error

6 Registration of multi-sensor sequences

The final stage consists of registration of sequences of video frames from multiple sensors. Registration of each and every video frame across different types of sensors is computationally expensive. In many applications involving image fusion [11], it is necessary to register consecutive video frames within each sensor (i.e. motion compensation) in addition to registration across the sensors. Consecutive frames from the same sensor can be registered using the computationally less expensive same-sensor registration techniques, as described in Section 2. This registration of consecutive video frames can then be used to aide in the task of registering the sequence of frames across multiple sensors.

Regstration of multi-sensor video sequences is carried out in two steps. A frame from one sensor is first registered with a frame from the other sensor using the method described in Sections 3, 4 and 5. These registered frames are then used as reference images. Subsequent video frames from each sensor are then registered with the corresponding reference images from the same sensor, using same-sensor registration technique. The video sequences that are obtained after performing this step will now be in registration within each sensor as well as across the sensors.

7 Experimental Results

We have tested our approach with several sets of multi-sensor imagery. We have used both synthetic sequences and real sequences from different types of sensors. The images on the left and middle in Figure 3 are captured from different portions of the infrared spectrum at different times. These images contain polarity reversals in some regions, whereas other regions have the same polarity of contrast. The image on the right shows these images registered, with the features in the images aligned. An example of registration of images from sequences captured

by FLIR and TV is shown in Figure 4. The images on the left and middle are from FLIR and TV respectively, and show a runway scene as an aircraft is approaching to land. These images contain gray-scale disparities as well as local polarity reversals. The image on the right shows the registration of these images using a projective model. The edges of the runways and the taxiways are properly aligned. We have observed that it is difficult to remove the residual mis-alignment in the regions closer to the sensors (bottom of the images), which is caused by the temporal properties of the sensor and the speed of the aircraft. Figure 5 shows the registration of FLIR and MMWR imagery. The FLIR image on the left contains features which are absent in the MMWR image and vice versa. The image on the right shows the registration of these images. We have evaluated our approach using synthetic sequences where the correct registration is known. Our experiments show that the registration is accurate within a pixel.

8 Summary

Most current algorithms for image registration are based either on the optical flow approach or the feature based approach. These algorithms cannot be directly applied to register images from multiple sensors due to additional difficulties caused by multi-sensor imagery, Specifically, gray-scale disparities and complementary features give rise to errors in registration. We have described an approach for registration of multi-sensor sequences that is based on reducing the errors caused by these disparities. We have used a non-linear representation that is invariant to gray-scale variations, and performed smoothing of the error surface for robust estimation of registration parameters. We have also described a procedure for reducing the errors in registration caused by complementary features. Finally, we have used same-sensor registration techniques to aide in the registration of video sequences from multiple sensors.

9 Acknowledgments

This work was supported by a grant to the Oregon Graduate Institute from NASA Ames Research Center, NCC2-811. We are grateful to Yacov Hel-Or for providing us with the source code of his gradient-based registration package, and to Todd Leen for useful discussions.

References

[1] P Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Visron,* 2(3):283–310, 1989.

[2] J. R. Bergen, P Anandan, K. J Hanna, and R. Hingorani. Heirarchical model based motion estimation. In *Second European Conference on Computer Vision (ECCV'92)*, pages 237–252, 1992.

[3] J. R. Bergen and P J Burt. A three-frame algorithm for estimating two-component image motion. *IEEE transactions on Pattern Analysis and machine Intelligence,* 14(9):886–896, September 1992.

[4] L. G. Brown. A survey of image registration techniques, *Computing Surveys,* 24(4):325–376, 1992.

[5] M. A Burgess, T Chang, D. E. Dunford, R. H. Hoh, W F Home, and R. F Tucker Synthetic vision technology demonstration executive summary, Technical Report DOT/FAA/RD-93/40,I, Research and Development Service, Washington, D.C., 1993.

[6] P J Burt and E. H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactrons on Communications,* Com-31(4):532–540, 1983.

[7] Y Hel-Or Studies in gradient-based registration. Technical report, NASA Ames Research Center

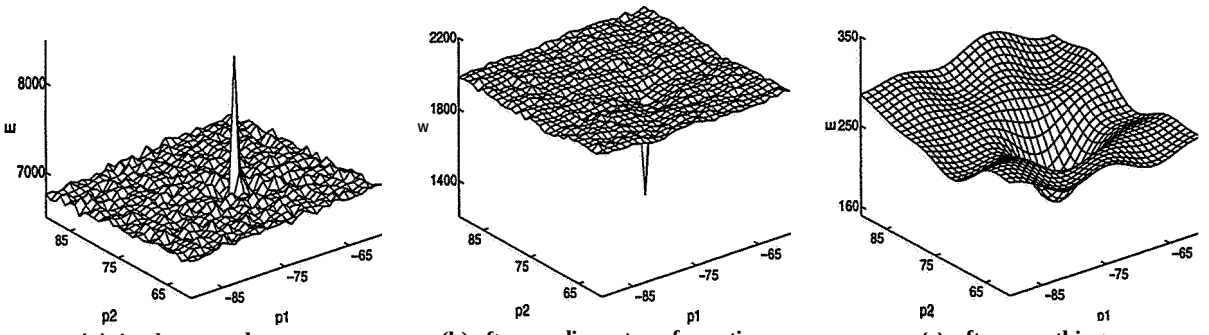
[8] B. K. P Horn and B. G Schunk. Determining optical flow. *Artificial Intellrgence,* 17:185–203, 1981

[9] H. Li and Y Zhou. Automatic visual/IR image registration. *Optical Engineering,* 35(2):391–400, 1996.

[10] B. D. Lucas and T Kanade. An iterative image registration technique with an application in stereo vision. In *Seventh Intemataonal Joint Conference on Artificial Intelligence,* pages 674–679, 1981

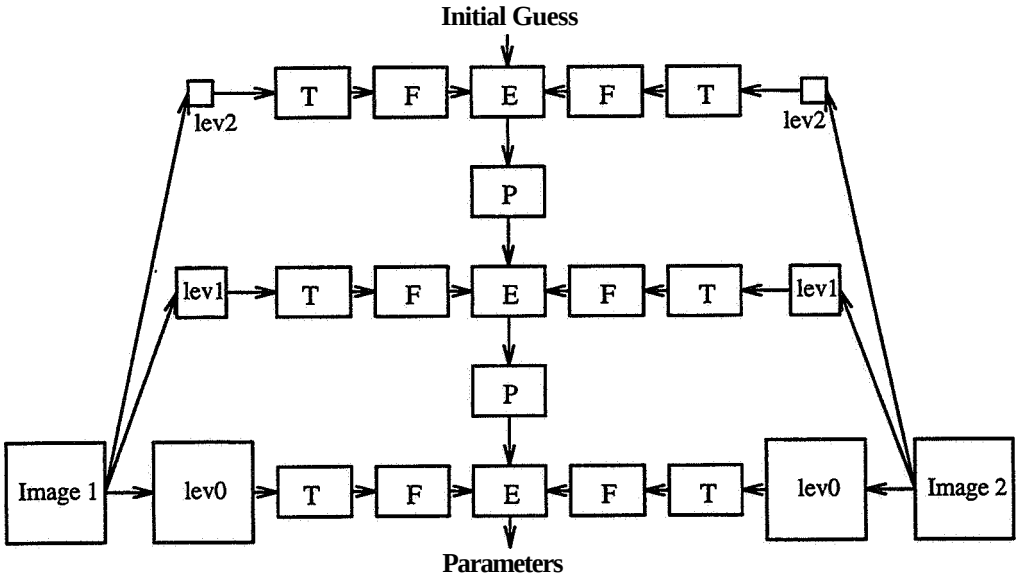
[11] R. K. Sharma and M. Pavel. Adaptive and statistical image fusion. *SID Digest of Technical Papers,* XXVII:969–972, 1996.

[12] Q. Zheng and R. Chellappa. A computational vision approach to image registration. In *Proceedings of the International Conference. on Pattern Recognitzon,* pages 193–197, The Hague, Netherlands, 1992.



(a) simple squared error (b) after non-linear transformation (c) after smoothing

Figure 1. Error surfaces for highest resolution of Laplacian pyramid.



T = Nonlinear Transformation
F = Filtering
lev = Laplacian pyramid level
E = Parameter Estimation
P = Parameter Propagation

Figure 2: Block diagram of registration process.

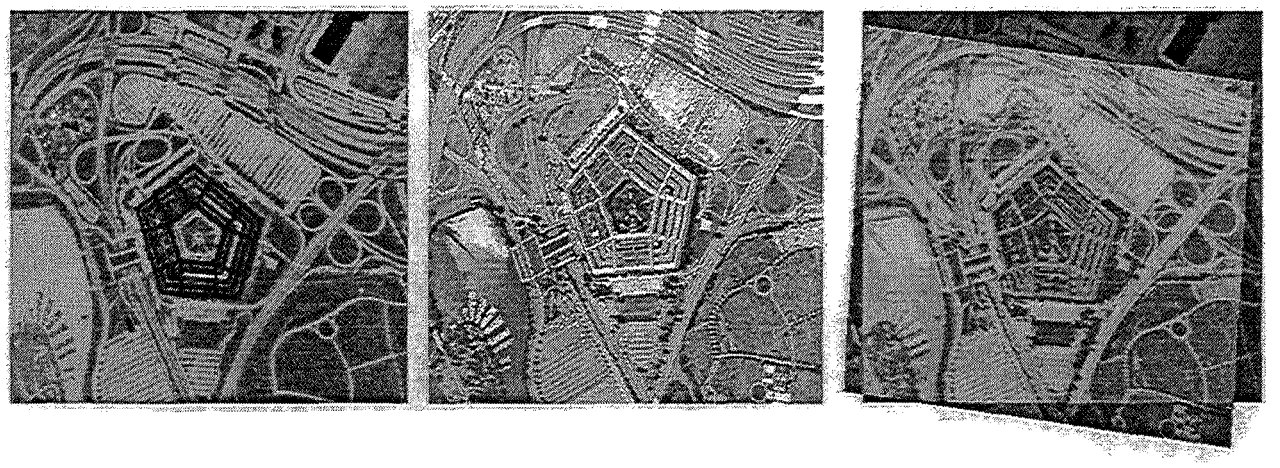


Figure *3* Registration **using** affine model (Data from **AMPS).**

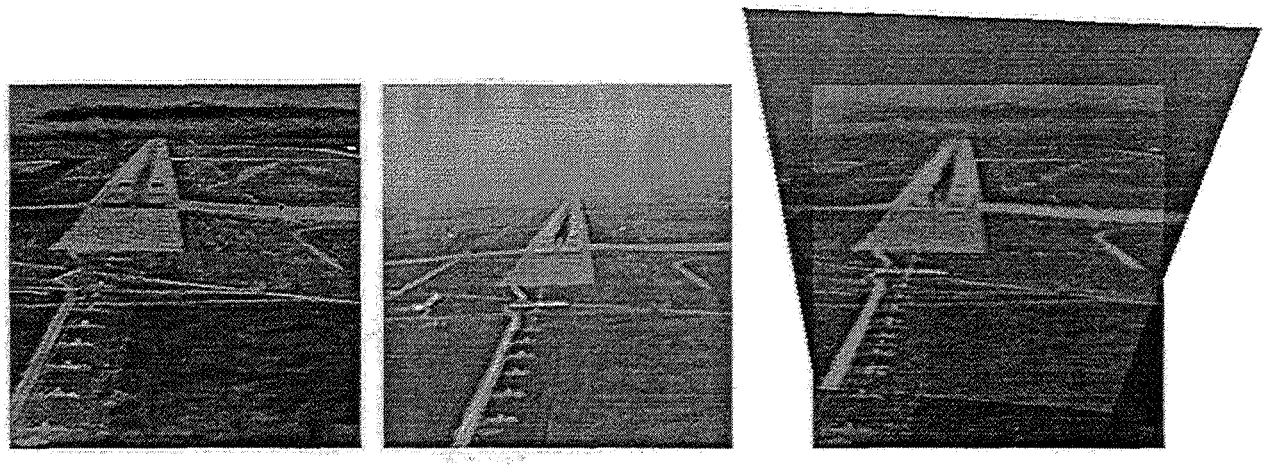


Figure **4:** Registration of FLIR **and visual** images **using** projective model (Data from SVTD[5])

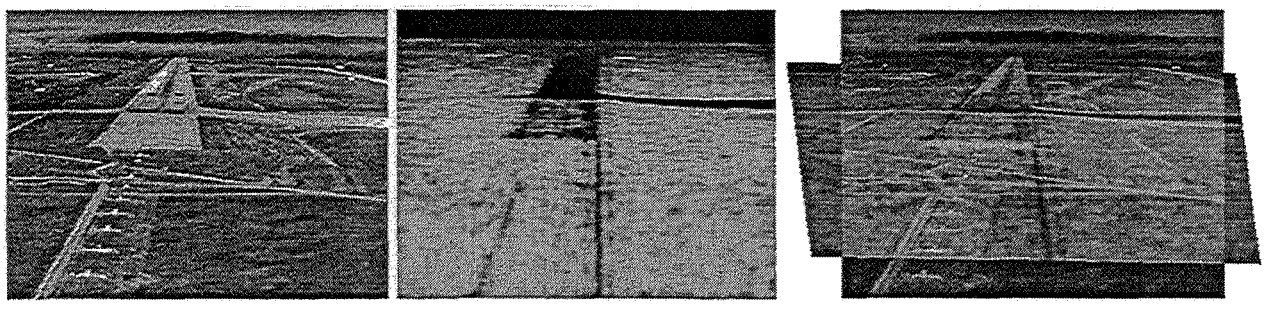


Figure 5: Registration **of** FLIR and radar images (Data from SVTD [5])

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1 AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 1997	Conference Publication

4. TITLE AND SUBTITLE

Image Registration Workshop Proceedings

5. FUNDING NUMBERS

6. AUTHOR(S)

Jacqueline LeMoigne, editor

7 PERFORMING ORGANIZATION NAME(S) AND ADDRESS (ES)

NASA Goddard Space Flight Center
Greenbelt, MD 20771

8. PEFORMING ORGANIZATION REPORT NUMBER

98B00048

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS (ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

CP-1998-206853

11 SUPPLEMENTARY NOTES

G Flanagan, Goddard Space Flight Center, Greenbelt, MD

12a. DISTRIBUTION/ AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Category 59

Availability. NASA CASI (301) 621-0390.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Current research in image registration focuses on automatic, fast, accurate, and reliable techniques whose application is feasible to the large amount of remote sensing data being generated by Earth and space missions. Image registration by point matching involves two main steps: the identification of significant features, commonly referred to as control pomts, and the matchmg of the features to identify the mapping function that brings one image into alignment with the master (reference) image. Automatically identifying control pomts is difficult, because (1) pomts significant in an mage ("hot points") need not match with corresponding points in a reference image, and (2) even when matched pairs of points are obtained, there must be a significant number of them suitably distributed spatially to compute accurately a mappmg function for the registration.

17 SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL