Multisource Data Integration in Remote Sensing

James C. Tilton, *Editor*
Goddard Space Flight Center
Greenbelt, Maryland

FOREWORD

On June 14-15, 1990, Technical Committee 7 (TC7) of the International Association for Pattern Recognition (IAPR) held a Workshop on "Multisource Data Integration in Remote Sensing" in College Park, Maryland. The original plan for this workshop was to place it in between two major, related conferences: the International Geoscience and Remote Sensing Symposium (IGARSS'90) at the end of May 1990, in College Park, MD; and the International Conference on Pattern Recognition (ICPR'90) in mid-June 1990, in Atlantic City, NJ. Such a configuration of two major conferences in geographically and temporally close range would have offered an excellent opportunity to realize a major goal of TC7: to bring together scientists from remote sensing applications and from the methodology of pattern recognition. Due to the postponement of the ICPR'90, the time between the two conferences became too long to allow participants of both conferences to attend the workshop by a short prolongation of their stay in the area. We therefore moved the workshop close to the ICPR to at least partly reach our goal.

The subject of the workshop, Multisource Data Integration in Remote Sensing, seems to have become a real challenge for the near future. New instruments and new sensors will provide us with a large variety of new views of the "real world." This huge amount of data has to be combined and integrated in a (computer-)model of this world. But also, the knowledge of how these data are gathered and what their characteristic properties are is among the useful sources of information that contribute to a meaningful interpretation. Multiple sources may give us complementary views of the world -- consistent observations from different (and independent) data sources support each other and increase their credibility, while contradictions that may be caused by noise, errors during processing, or misinterpretations, can be identified as such. As a consequence, data integration can produce results that are very reliable, and represents a valid source of information for any geographical information system (GIS).

The workshop structure consisted of three sessions of three to five individual presentations. All papers were discussed both individually and in the general context of the session. Additionally, all papers were considered under four specific aspects:

1. What are the characteristic properties of the data sources that are explored in the individual approach?
2. In which category does the used integration method fall?
3. What is the result of the integration and what are the improvements realized?
4. What are the major advantages of the proposed methods?

Four participants generously volunteered to consider the papers under one of the above aspects. Started as an experiment to encourage lively discussion, the four summaries (three of which are included herein) prove the success of this approach, and give also a useful index to the presented papers. It was an interesting experience for the authors to get a feedback on their own presentations.

Last but not least, I would especially like to thank Dr. James C. Tilton of NASA Goddard Space Flight Center for the excellent organization of this workshop. I would like to further acknowledge the sponsors of the workshop, the IAPR, NASA Goddard Space Flight Center, and the Washington/Northern Virginia Chapter of the IEEE Geoscience and Remote Sensing Society. My thanks go also to the authors that have contributed papers and to all the active participants that made the workshop more than just a sequence of presentations. I hope that activities of IAPR-TC7 will continue in the future under the leadership of the new chairman of TC7, Dr. Tilton, and that we can reassemble for another interesting workshop -- probably in 2 years -- on the occasion of the next ICPR, which will be held in The Hague, The Netherlands.

Dr. Walter G. Kropatsch

PRECEDING PAGE BLANK NOT FILMED

CONTENTS

PRESENTATIONS

SUMMARIES

APPENDIX

PRESENTATIONS

Refinement of Ground Reference Data
With Segmented Image Data

Jon W. Robinson
Image Analysis Section
ST Systems Corporation
4400 Forbes Boulevard
Lanham, MD 20706
Phone: (301) 794-5200
E-Mail: mscc@dftnic.gsfc.nasa.gov

James C. Tilton
Mail Code 936
Information Systems Development Facility
NASA Goddard Space Flight Center
Greenbelt, MD 20771
Phone: (301) 286-9510
E-Mail: tilton@chrpisis.gsfc.nasa.gov

ABSTRACT

There are a variety of ways for determining ground reference data for satellite remote sensing data. One of the ways is to photo-interpret low altitude aerial photographs and then digitize the cover types on a digitizing tablet. These digitized cover types are then registered to 7.5 minute U.S.G.S. maps that have themselves been digitized. The resulting ground reference data can then be registered to the satellite image, or, alternatively, the satellite image can be registered to the ground reference data. Unfortunately, there are many opportunities for error when using a digitizing tablet and the resolution of the edges for the ground reference data depends on the spacing of the points selected on the digitizing tablet. One of the consequences of this is that when overlaid on the image, errors and missed detail in the ground reference data become evident. This paper discusses an approach for correcting these errors and adding detail to the ground reference data through the use of a highly interactive, visually oriented process. This process involves the use of overlaid visual displays of the satellite image data, the ground reference data, and a segmentation of the satellite image data.

Several prototype programs have been implemented on the VAX computer system and the IVAS image display system to examine various methodologies for improving ground reference data. These programs provide a means of taking a segmented image and using the edges from the reference data to mask out those segment edges that are beyond a certain distance from the reference data edges. Then using the reference data edges as a guide, those segment edges that remain that are judged not to be image versions of the reference edges are manually marked and removed. We describe the prototype programs that were developed and the algorithmic refinements that facilitate execution of this task. Finally, we point out areas for future research.

INTRODUCTION

There are many ways to use multiple data sources in remote sensing. In this paper, we discuss a method for using image data to improve ground reference data sets. Reference data sets are sometimes referred to as "ground truth"; however, since the methods of creating reference data sets provide many opportunities for error, we will use the term reference data set. This terminology allows for the possibility of using multiple data sources for determining not only the contents of the reference data set but also for refining it.

3

We have obtained satellite image data from a number of investigators along with reference data sets that they created. By using the image data to create a segmented image, edges consistent with the spectral variation in the image are created. Visual inspection of the reference data edge map overlaid on the raw image or the segmented image reveals many discrepancies. These may be errors, or simply inappropriate attention to detail by the person generating the reference data (particularly if they were digitizing data from a digitizing tablet). An example exhibiting displacement errors and too coarse of a digitization scale is given in Figure 1.

While it would be desirable to have an automatic method of using the segmented image data to revise the reference data set, we chose, as a first approximation, to develop prototype methods that allow interactive selection and deletion of the segmented image edges that appeared to be too far from those edges generated from the reference data set. The end result is to leave a set of edges that are considered to be the true edges derived from the segmented image data, based on their proximity to the original reference data set edges.

METHODS

A series of programs were developed using IDL (Interactive Data Language) that allowed the user to use either an IIS system 575 terminal or an IIS IVAS terminal attached to a VAX cluster for interactive editing of the edge map generated by an image segmentation algorithm (Tilton, 1989) which runs on the MPP (Massively Parallel Processor) at Goddard Space Flight Center.

The image segmentation algorithm is an iterative process. At each iteration, those regions that are most similar by a particular criteria (e. g., minimum change in image entropy, or minimum rise in image mean squared error) are merged. As the number of iterations increase, the number of image segments decreases. A program called "edge.movie" that runs on the IIS System 575 was developed that allows the user to view various iteration steps and compare these with the original reference data set edges and bands from the image data.

This program allows the user to interactively pick an iteration and then immediately compare it with these other sources of information. It is best to select an iteration with more segments and edges than necessary to fit the reference data so that a crucial segment edge that may match a particular reference edge will not be lost. On the other hand, choosing too low an iteration and thus an image with too many segments and edges increases the work load of the analyst significantly.

The next step is extract the edges of the selected iteration from the combined edge file which contains the edges from all iterations coded by iteration number. This leaves an image of all of the edges present at iteration n with the pixel value of the edge indicative of its age.

The original reference data is a raster format image divided into regions. An IDL function, GREFEDGE.PRO was written to extract the edges from this image.

This segmented edge file is then masked with the edge file from the reference data to eliminate all edges that are beyond a specified distance from the reference set edges. This distance is selected by the analyst and depends on the image set being analyzed. Figure 2 shows an example of this process after this masking is complete.

4

Because the main algorithm for this procedure can only process images that are 128 x 128, sections of this size or smaller must be extracted from larger images for processing and then be recombined to produce a final product. In order for there to be continuity between sections it is necessary to have a certain amount of overlap between them. The degree of overlap required is being investigated in ongoing studies.

Two programs, CHGVAL5.pro and CHGVAL4.pro (for the IIS System 575 and IIS IVAS respectively) were written in IDL to provide a prototype interactive environment (listings in Appendix A). The input files (image, segment edge, original reference edge) for either of these programs can either be input one at a time as the program requests or the input file names can be read from an input file.

The CHGVAL5.pro program for the IIS System 575 image display system uses a track ball to move the cursor on the image display and the track ball function buttons to either delete an edge pixel, replace and edge pixel or go on to the next step in the program.

The CHGVAL4.pro program for the IIS IVAS system functions the same way except that it allows the display to be zoomed and roamed when looking for edges to delete. In addition, the IIS IVAS system uses a three button mouse instead of a multi-button track ball.

The CHGVAL#.pro programs start by asking whether the analyst wishes to enter the names of the image files individually or whether an input file containing all of the other input items is to be used. Since the program is usually invoked many times to complete the processing of a single section, the use of the input file saves the analyst the trouble of remembering and typing in all of the other file names each time the program is used on a section.

The edges from the original reference image are loaded into the graphics plane of the display device. Two bands of the original image data are loaded into the red and green display memories and the edges from the masked segmented edge image are loaded into the blue display memory.

In the CHGVAL5.pro program for the IIS system 575, the analyst uses the track ball to move the cursor over parts of the blue edge image that they wish to remove and click on the appropriate track ball button, producing "nicks" in the selected edge. For the CHGVAL4.pro program for the IIS IVAS system, the user can zoom and roam the image with the mouse first so it is easier to see what and where one is deleting edge segments. The terminal monitor prompts the user with the current functions of the track ball buttons or the mouse buttons. An example of this process after nicking several edges is given in Figure 3.

After having nicked a number of edges for removal, the analyst can exit this part of the program and the nicked edge image is written out to a file for processing by a batch program that runs on the MPP. After writing out the nicked edge image, the CHGVAL#.pro program submits the batch program to the MPP and waits for it to finish. The MPP program writes out the new edge image with the nicked edges completely removed. The CHGVAL#.pro program then reads in the new edge image and writes it out to the red display memory so the analyst can review the results of his work.

When the new edge image is read into the red memory bank, those parts of the original edge image that were deleted show up as blue and those parts that were not deleted show up as magenta (red "new edge image" + blue "blue old edge image"). The analyst is then given the choice of undoing some of his deletions or continuing with further nicking or

quitting and saving the result. An example of the final result from CHGVAL#.PRO for a single 128 x 128 section of data is given in Figure 4.

There are two types of edge connectivity that can be used, 8 connected and 4 connected. With 8 connected edges, unexpected deletions can chain through edge intersections. Thus it is particularly useful to have a number of ways of back tracking.

Our prototype system provides several alternative ways to recover edges. The simplest method is that while nicking lines, the user can undo a nick by placing the cursor over the nicked pixels and pushing the appropriate key on the mouse or the track ball. Alternately, the user can run a test, deleting the nicked edges and determine whether he wants to undo some of the nicks. If the user chooses to undo some of the nicks he can either fill in particular nicks using the cursor control device or he can pop deleted pixels off of the stack into their former locations. The latter method is most effective if the most questionable deletes are saved until near the end of a nicking cycle.

These methods of recovering erroneously deleted edges only work within a single nick and try cycle. If one either continues with a new nick and try cycle or saves the last result and exits the CHGVAL#.PRO program, then another method must be used to recover lost edges. In order to repair or recover edges after leaving the CHGVAL#.PRO a program that would allow pasting edge pixels from an earlier edge image into the current edge image was developed. This program puts one band of the image into the green display memory, the edge image to be fixed into the blue display memory and the master or original edge image into the red display memory. The user then uses the cursor control device to put the cursor over the master edges that he wishes to copy to the edge image to be fixed and presses the appropriate button. This puts a copy of the master pixel overlayed by the cursor into the edge image to be fixed. In this manner each pixel of an edge segment in the master image can be copied into the edge image to be fixed. An appropriate change in color takes place for each pixel that is moved into the edge image that is being repaired so the user can tell which pixels are present in both images.

The above description is summarized in Figures 5 and 6. Figure 5 illustrates the overall data flow from the original image and ground reference data to the revised reference edges (producing a revised ground reference data file). Figure 6 is a flow chart describing the CHGVAL#.PRO programs.

After several overlapping 128x128 edge images have been edited, it is necessary to rejoin them into a single image, to delete nonterminating edges in the overlap region and to join edges from overlapping segments. Another IDL procedure, COMBCHG4.PRO that runs on the IIS IVAS display is being developed to accomplish this task. Since its functions are so similar to those of CHGVAL4.PRO it is being designed to perform these functions also.

Two test images are being used. The first is a 468 x 368 pixel Thematic Mapper (TM) image of the Ridgely Quadrangle on the eastern shore of Maryland. It contains few classes and most of them are large in area. The main features of the scene are fields and wooded streams. There are small areas of water, ponds, and single urban area. These areas were digitized from a 7.5 United States Geological Survey quadrangle map that had had its reference data boundaries drawn on a plastic overlay from 1977 aerial photographs using a digitizing tablet (Gervin, et al., 1985). In the original study, the ability of AVHRR and MSS data to distinguish Level 1 land cover classes was examined. This involved registering the MSS data to the Ridgely Quadrangle and resampling it to 60 meter pixels. This lead to the ground reference data being rasterized to 60 meter pixels. Since the pixel size of the original reference data set was 60 meters and TM data

is 30 meters, the disparity between the TM boundaries and the reference data boundaries, and lack of detail in the reference data are understandable. Examples shown in Figures 1 through 4 are from a 128 x 128 section in the upper left corner of this data set.

The second image is a portion of the Washington D.C. metropolitan area. This area was broken down into more classes than the Ridgely quadrangle and the classes are smaller in size. Tests with this data set were not yet completed as of this writing.

RESULTS & DISCUSSION

In practice, the prototypes worked well, allowing the edge maps to be trimmed by nicking those edges to be deleted. Problems were primarily ones of speed in loading the original reference data into the graphics plane.

The use of 8 connected edges lead to unexpected chains of deletions. Because this was so severe, the processing procedures were started over using 4 connected edges.

Joining the segments with overlapping boundaries was not a serious problem with the Ridgely quadrangle. The large areas and concomitant small number of edges contributed to the each in joining the segments together.

Looking again at Figure 4, note that refined ground reference boundaries (obtained by selecting boundaries from the image segmentation) follow very closely visible boundaries in the image data. In particular, note the very coarse, misregistered boundary from the original ground reference file the top middle of the image. The refined ground reference boundary is perfectly registered, and follows the actual variations in the boundary very closely.

We plan to use this and other data sets in comparative studies of various algorithms designed to extract spatial information from imagery. We expect that the refined ground reference data will help to more accurately evaluate the behavior of these algorithms than would the often too coarse and misregistered original ground reference data.

REFERENCES

Gervin, J. C., A. G. Kerber, R. G. Witt, Y. C. Lu and R. Sekhon, 1985, "Comparison of Level I Land Cover Classification Accuracy for MSS and AVHRR Data," *International Journal of Remote Sensing*, 6(1), pp. 47-57

Tilton, J. C., "Image Segmentation by Iterative Parallel Region Growing and Splitting," *Proceedings of the 1989 International Geoscience and Remote Sensing Symposium*, Vancouver, BC, Canada, July 10-14, 1989, pp. 2420-2423.
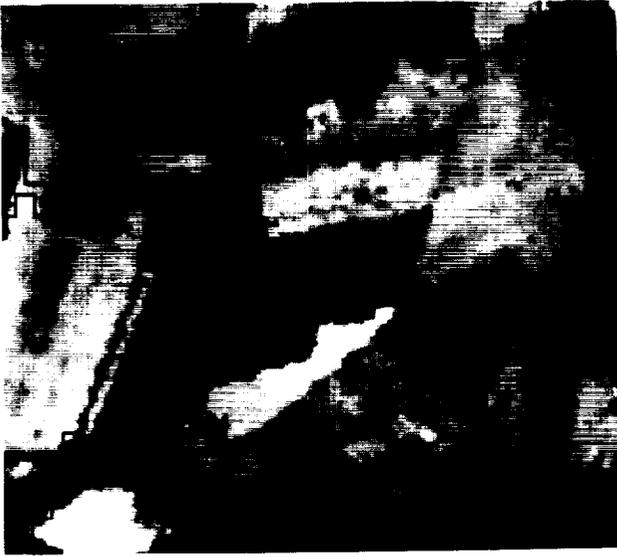
Figure 1. Edges from a ground reference file (black) overlaid upon the corresponding Landsat Thematic Mapper (TM) image. Note the displacement errors and coarse digitization scale compared to the Landsat data.
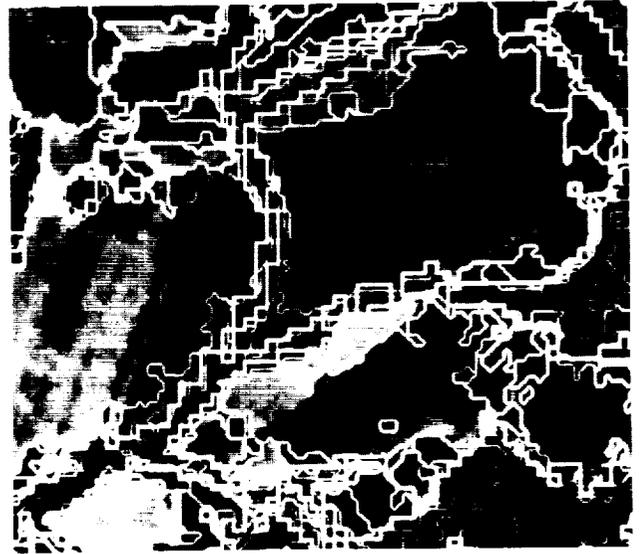
Figure 2. Edges from a ground reference file (white) and edges from an image segmentation (gray) overlaid upon the corresponding Landsat TM image. Image segmentation edges further than 6 pixels from a ground reference edge have been masked out.
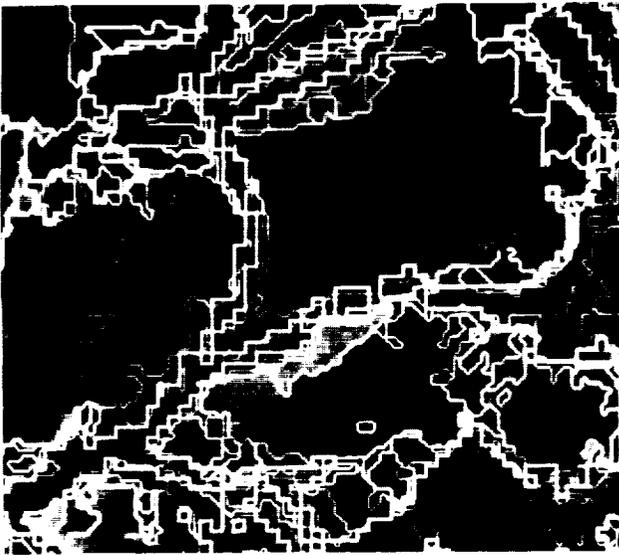
Figure 3. Edges from a ground reference file (white) and edges from an image segmentation (gray) overlaid upon the corresponding Landsat TM image. Image segmentation edges that are shown to be "nicked" in this image will be deleted by the next processing step (connected components labeling).
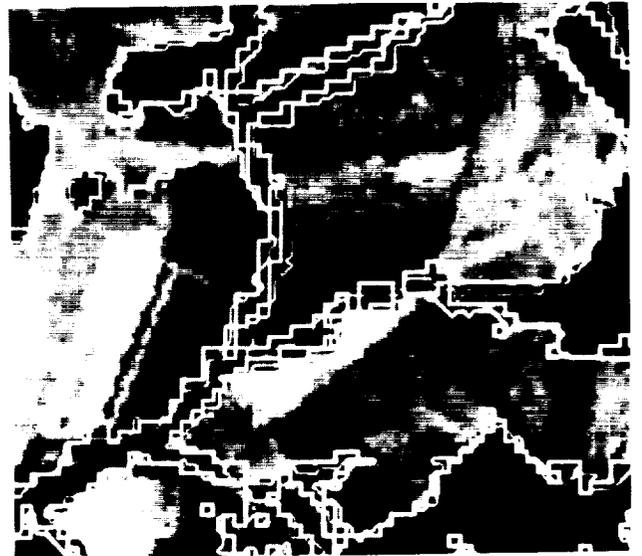
Figure 4. Edges from a ground reference file (white) and the final selection of corresponding edges from an image segmentation (gray) overlaid upon the Landsat TM image. This final selection of edges can now be used to generate a label map that can be used as a substitute for the original ground reference file.

8
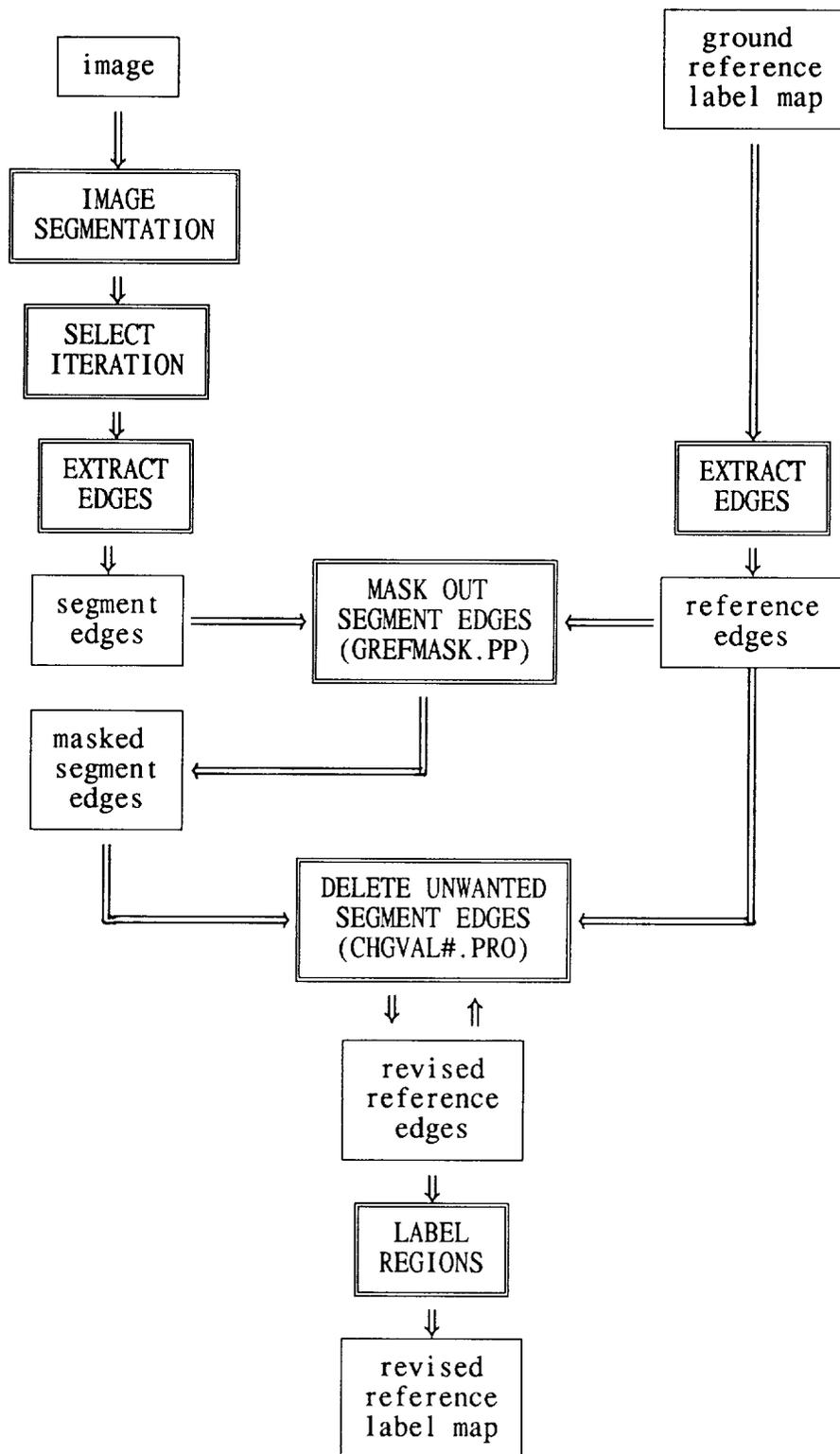
Figure 5. Overall data flow for producing a revised ground reference label map based upon an image segmentation.
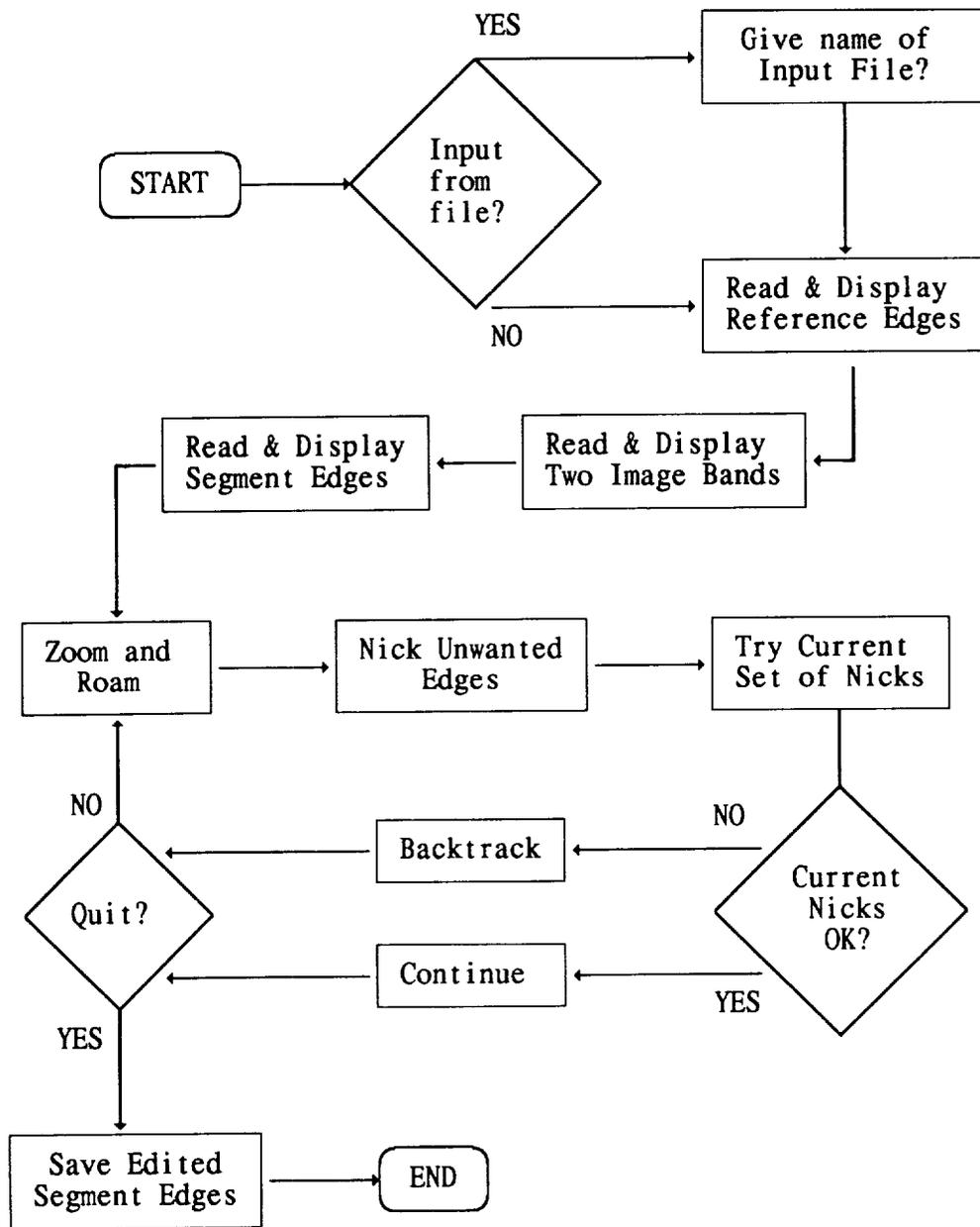
Figure 6. Flow chart outlining the steps carried out by the interactive CHGVAL#.PRO program.

Near Ground Level Sensing for Spatial Analysis of Vegetation

Tom Sauer, John Rasure, Charlie Gage

University of New Mexico
Department of Electrical and Computer Engineering
Albuquerque, NM, 87131

Abstract

Measured changes in vegetation indicate the dynamics of ecological processes and can identify the impacts from disturbance. Traditional methods of vegetation analysis tend to be slow because they are labor intensive; as a result, these methods are often confined to small local area measurements. Scientists need new algorithms and instruments that will allow them to efficiently study environmental dynamics across a range of different spatial scales. Presented is a new methodology that address this problem. This methodology includes the acquisition, processing and presentation of near ground level (NGL) image data and its corresponding spatial characteristics. The systematic approach taken encompasses a feature extraction process, a supervised and unsupervised classification process, and a region labeling process yielding spatial information.

1. Introduction

1.1. Motivation

During the 1990's NASA will establish a new remote sensing system, the Earth Observation System (EOS), with a variety of sensors and resolutions. Interpretation of the data at different resolutions will require ground level validation and correlation studies that quantify the heterogeneity of the environment over the range of spatial scales. Both transect sampling (NGL sensing) and remote sensing (satellite sensing) provide data that can identify changes in landscape[1]. Changes in species populations represent shifts in community organization that typically show temporal and spatial variation. Changes in organization among species can occur randomly or in response to governing biotic and abiotic factors[2]. These types of changes can not be detected accurately at the satellite sensing level, and currently the NGL methods used to determine change are typically labor intensive and slow. Thus, there exists a need to develop a new methodology to analyze the NGL sensed data.

This new methodology, also should provide the scientist with information that correlates satellite imagery with NGL imagery. For example, the spectral signature for a pixel in a satellite image provides a single, integrated measure of the ecological patterns within the ground surface area represented by the pixel. The same pixel value may be the result of diverse ground conditions. Without finer resolution imagery it is impossible to determine whether this signature

corresponds to a uniform cover of vegetation or various combinations of vegetation patterns. For the ecological community to make full use of remotely sensed data, it is critical to provide a way to relate the integrated reflectance values to the variety of vegetation patterns that occur at different scales.

1.2. Summary of the NGL Methodology

The NGL sensing system provides absolute and relative measurements of ground level vegetation. The NGL measurement process starts with the acquisition of 35mm color slide images of field plots. The field plot images, ranging in resolution from 1mm to 1cm, and varying in size from 0.5 m^2 to 10 m^2, are obtained using a camera gimbal mounted on a boom. Each rectangular plot is corner marked for later spatial registration. The NGL image is digitized with a high resolution, 4000 by 6000 pixels, slide scanner, and then image analysis is performed using a workstation based software system called Khoros (see Appendix).

The NGL images are comprised of only three spectral bands in the visual region of the spectrum: red, green and blue (RGB). Since the NGL images do not contain spectral information in the infrared region, the image processing analysis that allows differentiation between plant species and the differentiation of above ground biomass and bare ground is more difficult.
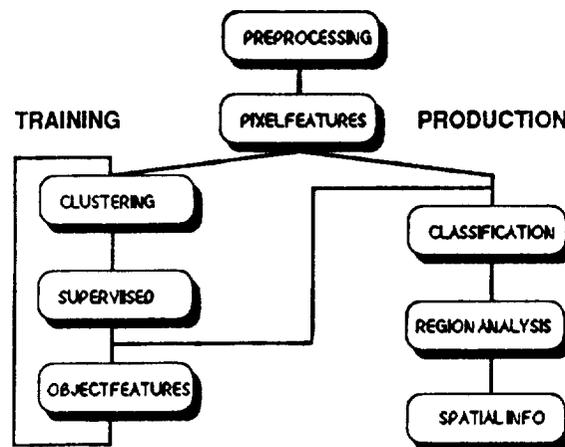
Figure 1. Block diagram of NGL methodology.

The NGL image analysis employs four related components: preprocessing, feature extraction, classification, and region analysis, see Figure 1. Preprocessing can involve warping the image to achieve spatial registration, median filtering to reduce noise, and image cropping. Once image preprocessing is complete, pixel features can be extracted. Pixel features include spectral information, local statistical measures, and various texture measures such as Hurst fractal dimension or the Laws' texture metrics. Representation of the RGB triples in other color spaces often allows for a better segmentation. Local statistical operations that can be computed using Khoros include: mean, variance, contrast, second angular moment, zero order entropy, and dispersion. Spatial feature extraction can be performed on one, all, or a ratio of the preprocessed RGB data bands.

The original spectral bands (RGB) and the feature bands of data are then combined to produce a multiband image. The concept of a multiband image is analogous to a multispectral image; each pixel in the image now contains many elements or attributes (spectral reflectance and features). Since a pixel contains many elements it can be thought of as a vector, where each element in the vector represents a different attribute associated with the pixel. This data organization lends itself to general classification methods.

The overall classification process first involves a one-time training phase that produces a mapping that is then used in the production phase. The training phase is a two part process, unsupervised classification followed by supervised classification. The unsupervised classification portion of the training phase is used to, (1) reduce the complexity and the dimensionality of the multiband image, and (2) determine the inherent structure of the data based solely on reflectance and texture measurements, which are unconstrained by external knowledge of the data. The supervised classification step allows an analyst to map the *clusters* determined by the unsupervised classifier to specific desired classes. The motivation for using the unsupervised classifier first is to reduce the complexity associated with the supervised classification. It has been found that combining the two types of classifiers in this manner produces relatively accurate decision boundaries, and therefore near minimal classification error[3][4].

After a single pass of the training process, object features are obtained that can be added to the original multiband image. This multiband image is used as a new input to the training process, see Figure 1. Object features such as geometric moments, fractal dimension and morphology supplement the pixel features used in the previous pass to produce a more accurate classification. The final result of the training phase will produce data vectors that represent the different cluster means and variances.

The second phase of the classification process uses the results, cluster means and variances, obtained in the training phase to classify other images that fit in the same representative set used in the training phase. Algorithms as simple as a minimum distance classifier, or as robust as the approximated likelihood ratio detector are available in Khoros. The classification process is followed by spatial analysis. Percent coverage of above ground biomass and individual plants is calculated. This information is the basis for the time series analysis that then can be correlated with changes seen at the remote sensing level.

This methodology has been applied to the analysis of images for the Sevilleta Long Term Ecological Research, LTER, project. The National Science Foundation LTER program supports research on long-term ecological phenomena at a national network of sites. One major goal of the LTER project is to study long-term trends in natural ecosystems that have not previously been systematically monitored. The NGL methodology is capable of analyzing image data acquired from large transect plots and small fertilizer plots. A time series analysis of this data can be accurately tracked and eventually correlated with changes seen at the global level. This methodology provides a link that will allow ecological phenomena that occur on large time scales to be investigated.

2. Theory of the NGL Classification System

2.1. Image Preprocessing

Image preprocessing for the NGL images acquired at the Sevilleta LTER site only require geometric correction and image cropping. Median filtering was originally used to reduce noise artifacts, but the final spatial measurements exhibit distortion caused by the smoothing effects of

the filter.

Since the NGL data are captured using a 35mm camera gimbal mounted on a boom, the image will be distorted because of camera position and terrain topology. The corner markers in the image provide tie points that will allow the image to be warped back to the correct geometry. Since the acquisition system uses 35mm slides, the xy pixel ratio in the image is 2/3, and must be corrected back to a 1/1 pixel ratio when the slides are digitized. The Khoros interactive image editor allows a user to select the image corner points and record the xy locations as source tie points. The user must then specify the distance between the tie points. From this information the destination tie points may be computed. The following table and equations describe the destination tie point computations

$$\alpha = \frac{L}{(x'^2+y'^2)^{\frac{1}{2}}} \tag{1}$$

Where: $x' = k(x_{p_1}-x_{p_2})$ $y' = k(y_{p_1}-y_{p_2})$

p_1 *is some tie point*

p_2 *is some other tie point*

k *is a pixel aspect ratio constant*

L *is the actual distance between* p_1 *and* p_2 *in meters*

α *is the new coordinate position translation factor*

Source tie points	Destination tie points
(x_1,y_1)	$(0,0)+(x_1,y_1)$
(x_2,y_2)	$(\frac{1}{\alpha},0)+(x_1,y_1)$
(x_3,y_3)	$(\frac{1}{\alpha},\frac{1}{\alpha})+(x_1,y_1)$
(x_4,y_4)	$(0,\frac{1}{\alpha})+(x_1,y_1)$

The next step in the registration process is to use the four source and destination tie point pairs to compute the coefficients for two first-order equations that will be used to perform the image registration. In some cases, a wide angle lens causes severe image distortion. This requires the use of more than four tie points, resulting in a higher order warping polynomial. The original image is warped using the computed polynomial equation and bilinear interpolation, and then cropped using the destination tie points.

2.2. Feature Extraction

Texture measures are typically computed on a single band image, necessitating the reduction of the RGB image to a single band image. This reduction is performed using a color quantizer that reduces the image from 16.7 million colors (3 bands) to 256 colors (1 band). Alternatively, the RGB image can be converted to the HSV (Hue, Saturation, and Value) color space with the spatial measures computed on the value band.

Simple statistical parameters based on local area measurements over a small moving window are commonly used to provide texture information[5]. The statistical parameters, mean and variance, are based on the central moments and are used to provide an indication of how uniform or regular a region is. Contrast provides a measure of the dissimilarity of the intensity values in the image, and angular second moment yields a measure of uniformity or homogeneity of the gray level values. An indication of the texture nonuniformity is provided by a measure of the entropy. Texture measures based on these statistical parameters did not yield any new information that aided in the classification process. For this reason, features based on simple statistical parameters were not used.

Although numerous texture measures have been proposed to characterize the spatial texture features in an image, good results have been obtained using a set of spatial convolution masks proposed by K. I. Laws[6]. The Laws' texture masks are comprised of a set of 5 by 5 masks that are convolved over the entire image[7]. The masks are intended to be sensitive to visual structures such as edges, ripples, and spots.

Each of the Laws' texture masks are derived from a set of five basic vectors. There are a total of 25 possible masks, each formed by multiplying two of the five vectors together. They are designed to act as matched filters for certain types of quasiperiodic variations commonly found in textured regions.

Various texture masks were tried in order to achieve good discriminating power between adjacent regions in the image. The set of texture masks that provided the best results include the L5E5 and E5L5 masks. The L5E5 and E5L5 masks are constructed by multiplying the L5 and E5 vectors, yielding the following texture masks:

$$L5E5 = \begin{bmatrix} -1 & -4 & -6 & -4 & -1 \\ -2 & -8 & -12 & -8 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 8 & 12 & 8 & 2 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} \qquad E5L5 = \begin{bmatrix} -1 & -2 & 0 & 2 & 1 \\ -4 & -8 & 0 & 8 & 4 \\ -6 & -12 & 0 & 12 & 6 \\ -4 & -8 & 0 & 8 & 4 \\ -1 & -2 & 0 & 2 & 1 \end{bmatrix}$$

The L5E5 mask tends to detect edges arising from horizontal changes in texture, while the E5L5 mask detects texture changes in the vertical direction.

Once the spatial texture features are extracted by convolving each mask with the gray level image, an additional feature selection step is used to reduce the dimensionality of the classification process. This involves a 50% blending of the two texture bands into one texture band that contains the information extracted by each of the texture masks. By using one texture band, the overall weighting of the texture features relative to the spectral image information is reduced. This provides a more representative weighting of the original spectral information relative to the spatial texture information in the classification process.

2.3. Classification

As was mentioned above, the classification process is a two phase process. The first phase is considered the training phase, while the second phase implements the actual classifier and is referred to as the production phase. The training phase is further broken into two parts, unsupervised classification and supervised classification. The unsupervised training determines the inherent structure of the data, unconstrained by external knowledge about the vegetation patterns, while the supervised training imposes the analyst's knowledge of the vegetation patterns to constrain the results. The final objective of the classification process is to reduce a large data set

(the multiband image) into a few classes in a single band image.

2.3.1. Training Phase

The goal of the training phase is to produce an ensemble of data that characterizes a representative set of NGL images, so that other images with similar characteristics can be *automatically* classified (the production phase). In the unsupervised classification portion of the training phase the algorithm maps areas on the ground that have similar texture and spectral reflectance characteristics to the same cluster. The resulting clusters assigned to the image pixels therefore represent different *classes* that may or may not correspond to the classes of ground objects that we are ultimately interested in mapping. A good example of such a situation is the mapping of shadow areas and wet or dark soil areas. The analyst may want to ultimately consider both of these classes as bare ground, but each may represent a separate cluster as produced by the unsupervised classifier. The output of the unsupervised classifier is a single band pseudo colored image that represents a map of the clustered pixel vectors, the cluster centers (means), and variances. The mean and variance data represent the ensemble of data that characterizes a specific set of NGL images.

Image data that represents specific areas to be classified are submitted to the unsupervised classifier. The unsupervised classifier is implemented as a clustering algorithm that will determine the *natural* groupings of clusters of the data in K-dimensional feature space. The cluster centers represent an estimate of the probability density function. The cluster centers are then assigned to classes during the supervised classification. The determination of clusters is accomplished by the K-means clustering algorithm[8].

The K-means algorithm is a partitional algorithm that attempts to minimize the sum of squared errors in its cluster assignments. The similarity measure used is the Euclidean distance. The K-means algorithm partitions the data space by using a search method where patterns are moved from one cluster to another until all patterns belong to a cluster. Each cluster is identified by a single cluster center (mean) and cluster variance. Since the K-means algorithm uses the Euclidean distance as a similarity measure, it is vital that the features previously determined are weighted so as not to bias the results produced by K-means. The performance of K-means is improved if the feature pixel vectors are orthogonal. In practice, however, this is rarely the case. Therefore, it is best to *over-cluster* the pixel vectors resulting in a less refined classification.

Experiments show that the number of clusters produced by K-means should be about four to seven times the final number of classes desired. The cluster centers provide the location in K-dimensional space for each cluster, while the variance describes the size and orientation of each cluster. This information is used in the supervised classification described below.

The output of the unsupervised classifier provides a mapping of pixels in the original image to different clusters. The clusters produced by the unsupervised classifier are usually not the desired classes; the object of the supervised classifier is to map each cluster to a desired class. The supervised classification process is performed manually using the Khoros image editor. The Khoros image editor allows the analyst to display both the clustered image and the original image. Cluster numbers in the clustered image can then be assigned to specific desired classes. The resulting mapping of clusters to specific desired classes will be used in the production phase of the classification process.

Often the data in a cluster may need dividing because it is spread over multiple desired classes. The P(m,L) fractal algorithm can be used to help determine the splitting of the clusters.

The P(m,L) distribution is obtained from the unsupervised classification image data. Frequency distributions for each class m, are determined for a series of different window sizes, L. The resulting probability distribution should provide valuable information describing the aggregation of classified pixels in the NGL image.

The P(m,L) probability density function has moments that vary with the measurement scale. This scale dependent characteristic of the moments provides a framework for transforming plant coverage estimates from one scale to another. It has been found that natural landscapes often exhibit consistent changes in the fractal dimensions over a range of moments[9]. This provides a way of measuring the degree of relationship from one scale to another.

Once the moment bands have been determined, they will be appended to the multiband image containing the spectral and texture bands. This image will then be reprocessed by the training phase. The result of this iterative processing will produce statistics (cluster means and variances) that better describe the desired classes.

2.3.2. Production Phase

The object of the this phase is to take the mapping obtained in the training phase and allow unsupervised classification of subsequent images that are considered to be in the same representative data set as used during training. It is required that the same feature extraction process is performed on the new images as was performed on the training set. The unsupervised classifier used in this phase is the approximated likelihood ratio detector (ALRD). The ALRD uses the cluster centers, cluster variances, and cluster to class mapping to classify new images. This robust unsupervised algorithm is not limited to detecting whether a pixel vector belongs to a single class. A pixel vector can be assigned to multiple classes and through a thresholding test determine to which class the pixel best belongs. If a pixel vector does not have a high enough probability to belong to any class, then it is considered an outlier, and thus unclassifiable. This algorithm uses the ratio of the distance of a pixel vector to a cluster center to each diagonal element of the covariance matrix (variance elements), to determine to which class a pixel belongs. In other words, the algorithm computes the probability density function of all clusters that belong to a class and then determines if a data point has a high enough probability to belong to that class. The diagonal of the covariance matrix is computed by Equation (1) while Equations (2) and (3) perform the likelihood ratio test.

$$diag\,(C_i) = \begin{bmatrix} \dfrac{1}{N_i} \sum_{l=1}^{N_i} (X_{0il} - m_{0i})^2 \\ \dfrac{1}{N_i} \sum_{l=1}^{N_i} (X_{1il} - m_{1i})^2 \\ \cdot \\ \cdot \\ \dfrac{1}{N_i} \sum_{l=1}^{N_i} (X_{nil} - m_{ni})^2 \end{bmatrix} = \begin{bmatrix} \sigma_{0i}^2 \\ \sigma_{1i}^2 \\ \cdot \\ \sigma_{ni}^2 \end{bmatrix} \quad (2)$$

where N_i is the number of points in the ith class, m_i is the cluster mean, l is an index into the list of data points belonging to ith cluster, and n is the dimensionality of the vector.

$$P_i = \sum_{j=0}^{D} \frac{||X_{ji} - m_i||^2}{K \sigma_{ji}^2} \quad (3)$$

17

P_i is the likelihood ratio of the ith class, D is the dimension of the unclassified data vector X, and K is a tuning parameter.

$$class_X = \begin{cases} no\ class\ \text{if } P_i > 1 \\ class\ i\ \text{if } P_i < 1\ minimum(P_i) \end{cases}$$

(4)

The diagonal of the covariance matrix and the cluster means are computed during the training process. The similarity measure used by the ALRD is the same as that used in the K-means algorithm, the Euclidean distance. The tuning factor adjusts the likelihood ratio, which either increases or decreases the number of outliers detected.

The ALRD is used rather than the minimum distance classifier because it allows for outlier or unclassifiable pixels. This reduces the size of the training set because it eliminates the need to classify every possible pixel vector. The ALRD also uses both the size and orientation of the classes in K-dimensional feature space to aid in classifying new pixels.

2.4. Region Analysis

The final step in the NGL image analysis is the calculation of class and region moments [10]. For example, in the case of a two class image (above ground biomass and bare ground), the area calculations result in percent vegetation cover. More detailed information can be obtained by labeling the individual objects in the two class image and then calculating moments.

Labeling of individual objects is based on the splitting and merging of regions, where the decision metric is the gradient between eight-connected neighbors. The labeling algorithm uses either the difference between the gray levels of adjacent pixels or the Euclidean distance between adjacent pixel vectors as the gradient value. If the gradient value is less than a threshold the regions are merged. The moment calculations (standard, central, and invariant) on the resulting labeled regions give detailed spatial information on each object. This information provides the analyst with the necessary information to track individual plant changes over time.

The region analysis algorithm generates two images; (1) an axis image that contains a cross for each region with the cross centroid located at the center of the object, and (2) a region outline image or contour image. Overlaying the outline image upon the original RGB image or the axis image, provides the analyst with a means of visual interpretation and verification. This assists in the time series analysis since it allows the analyst to visually track the vegetation changes.

3. Discussion of a Specific Example

The NGL methodology has been applied to helping ecologists at the Sevilleta LTER site track the vegetation change in both transect and fertilizer plots. The transect plot dimensions are usually 10m by 5m, and fertilizer plots are usually 1m by 0.5m. The following example will illustrate the results obtained by using the NGL methodology on transect images. A representative image of a transect area in the field is used as the training pattern, then another image is classified based on the results from the training. In this example, vegetation is segmented from all other matter, thus a two class problem.

This example begins with a representative transect plot image, Figure 2, that has been spatially registered and cropped.
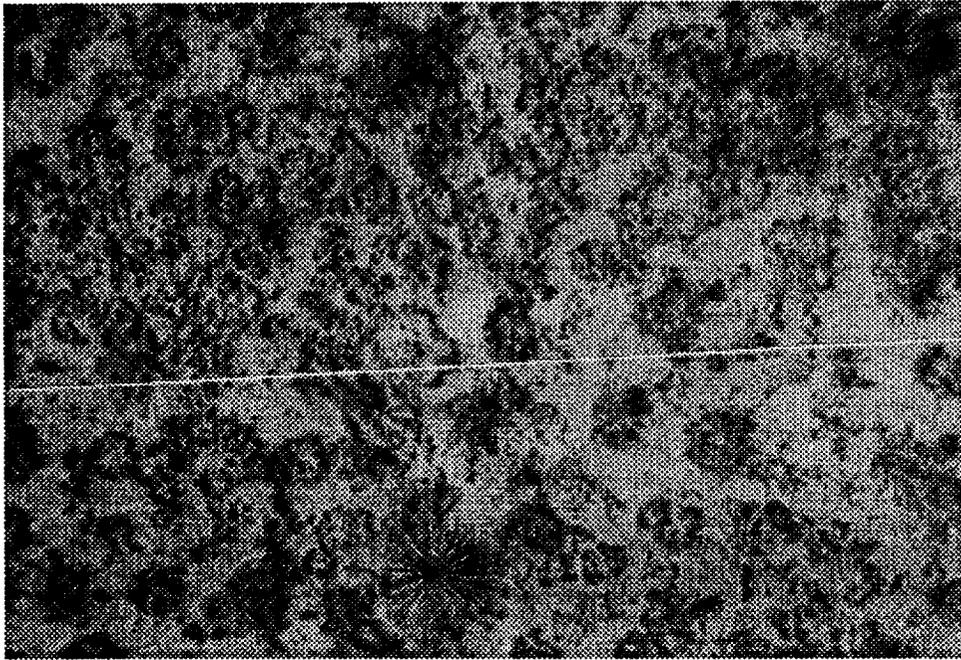
Figure 2. Representative transect plot image.

The next step in the process is to compute pixel features. Pixel features are computed using the Laws' texture metrics. Since these metrics only work an a single band image, the original RGB image must be compressed down to a single band. This is performed using the color quantizer method. Two different Laws' texture kernels, E5L5 and its transpose L5E5, are convolved with the one band image producing two single band images that are blended together producing another single band image, shown in Figure 3.

The texture band is then appended to the end of the original RGB image. This new multi-band image is used in the classification training process. The K-means algorithm, produces a single band cluster number image shown in Figure 4.

Figure 5 illustrates a plot of the distribution of the cluster centers. Each row of impulses represent a different set of cluster center values. This plot gives a visual interpretation of the correlation between different cluster centers.

The next step is the supervised classification phase of the training. Cluster numbers are assigned to specific classes using the Khoros interactive image editor. The result of the supervised classification is shown in Figure 6.

At this point the training can stop if all the clusters have been mapped to the desired classes. Otherwise, object features are computed and the system is retrained. In this example all clusters have been mapped to the two desired classes. This ends the training phase.
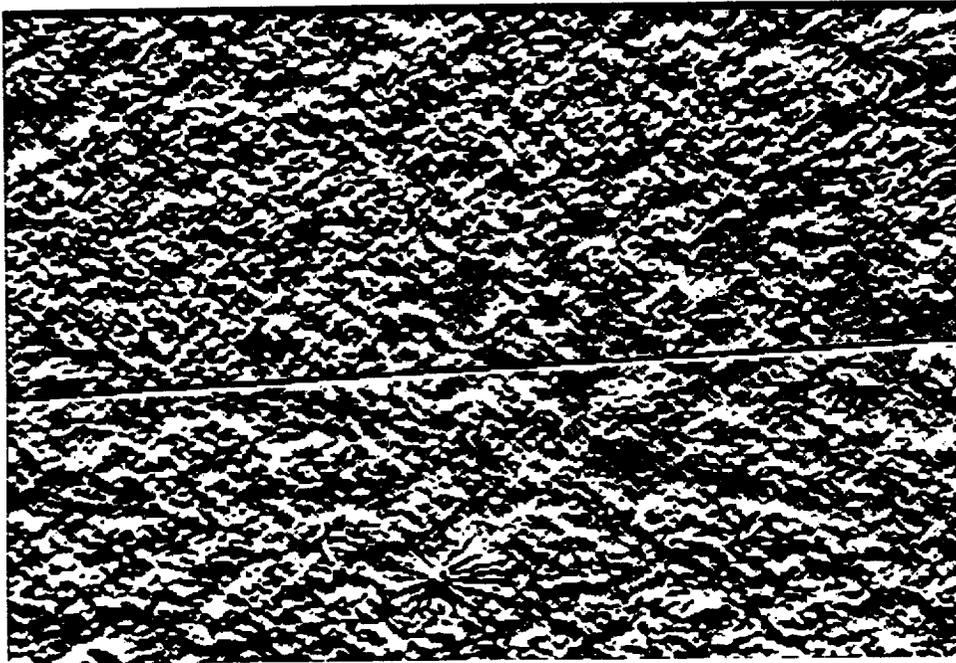
Figure 3. Laws' texture band.
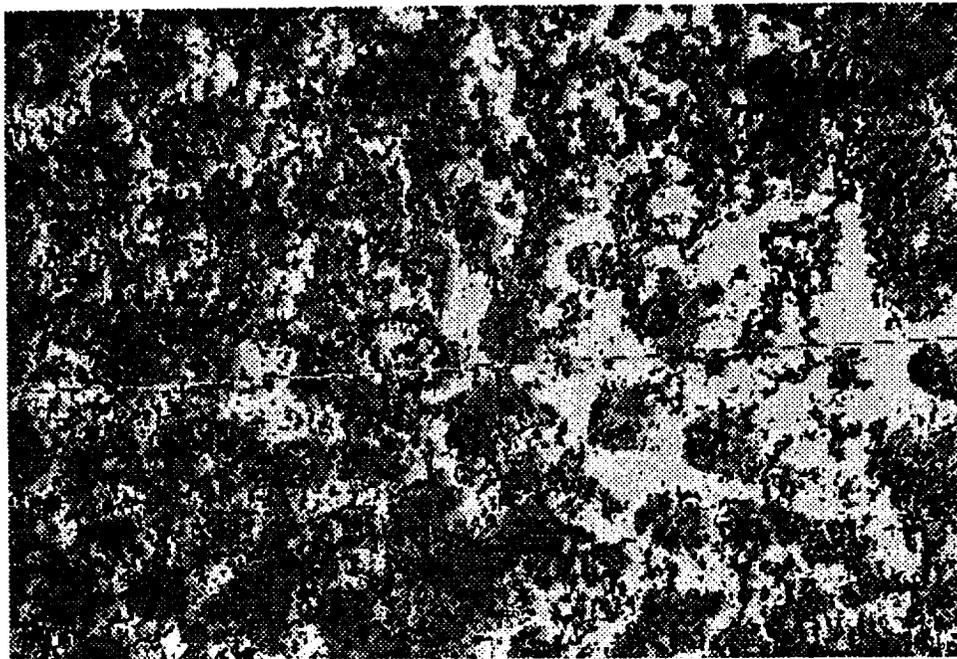
Figure 4. Cluster number image.
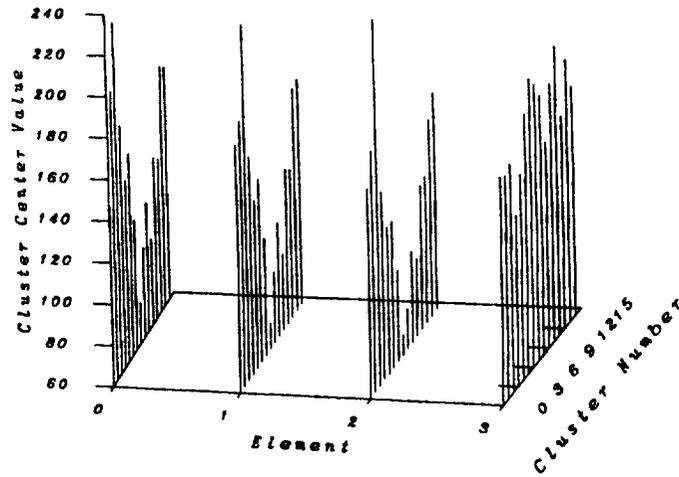
Figure 5. Plot of cluster centers.
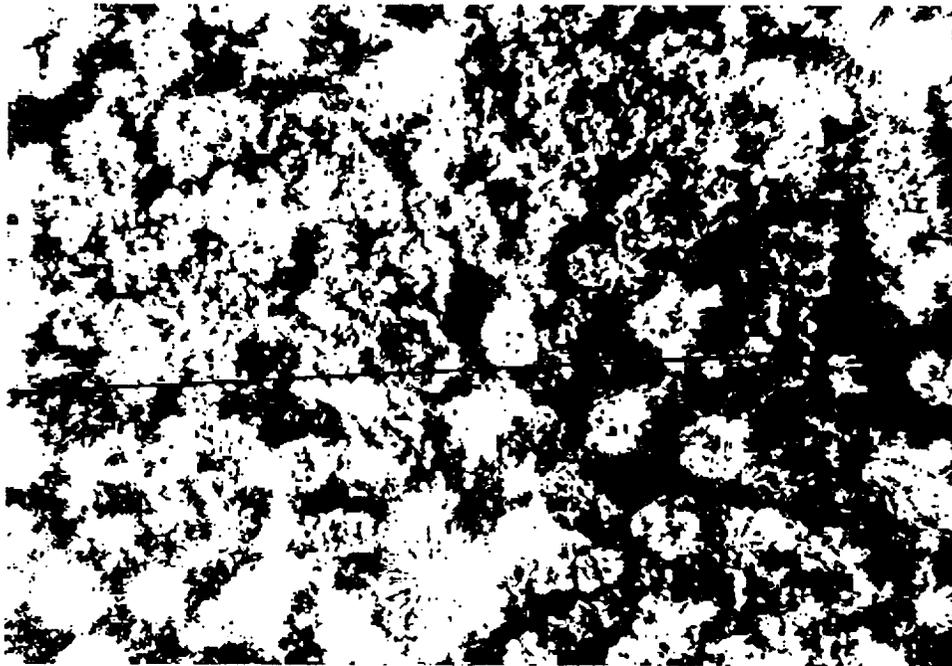
Figure 6. Resulting class image after training phase.

21

With the results produced by the training phase, other images from the same representative set can be classified using the approximated likelihood ratio detector. The new image to classify is shown in Figure 7.
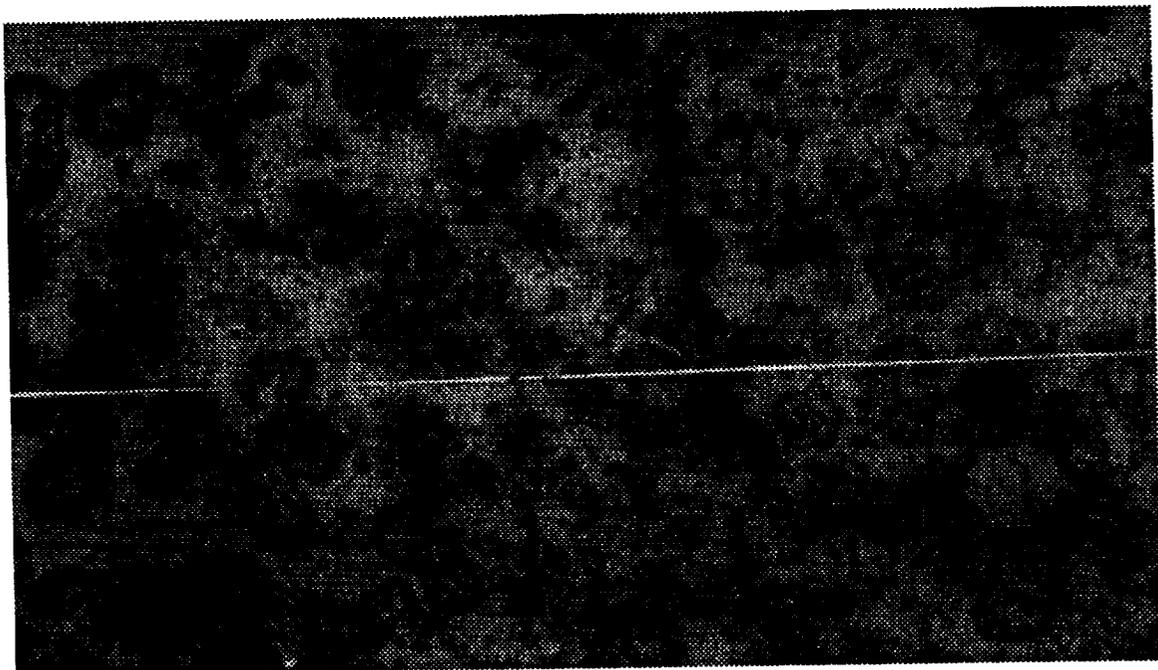
Figure 7. Transect image to be classified.

The same image preprocessing and feature extraction is performed on this image as on the training image, resulting in a multiband image. The result of the unsupervised classification is shown in Figure 8.

The final step is to perform region and class analysis to determine the desired spatial information. Figure 9 illustrates the result of the analysis procedure. This image shows the size of the regions by outlining them and the orientation by the crosses in each outlined region. In this example the percent coverage of vegetation is 38.43%.

This example illustrates the applicability of the NGL methodology for the Sevilleta LTER project. The system is planned for production use by the end of the year. The ecologists see this approach as critical to the successful and timely analysis of the thousands of transect and fertilizer plot images required by the project.
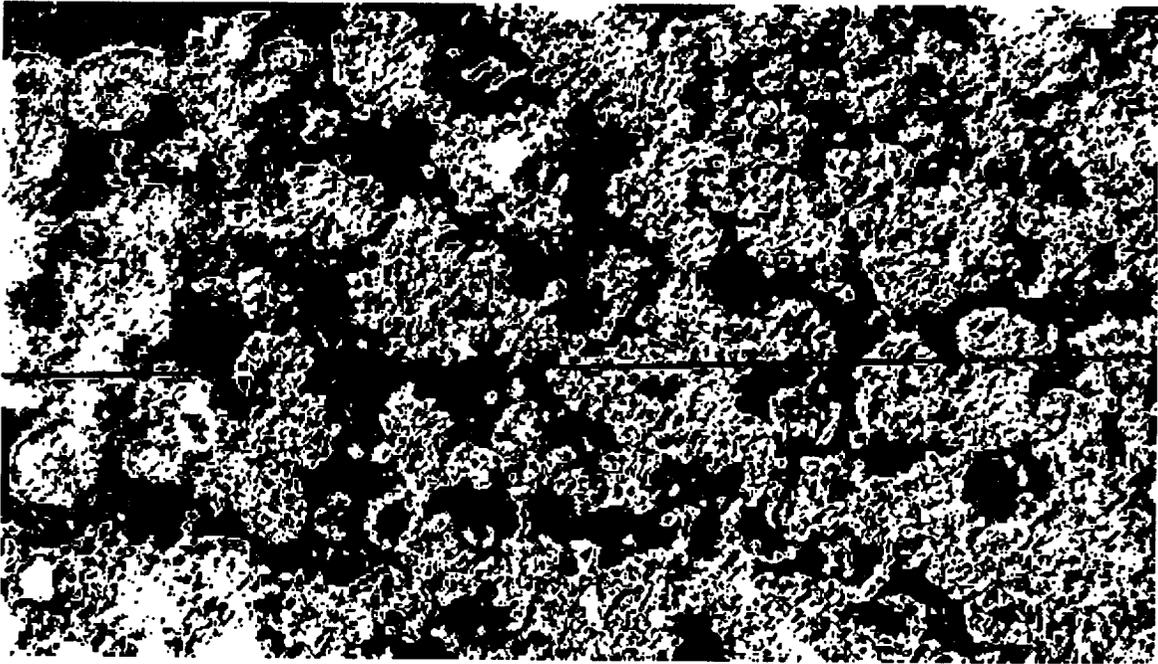
Figure 8. Resulting image after classification.
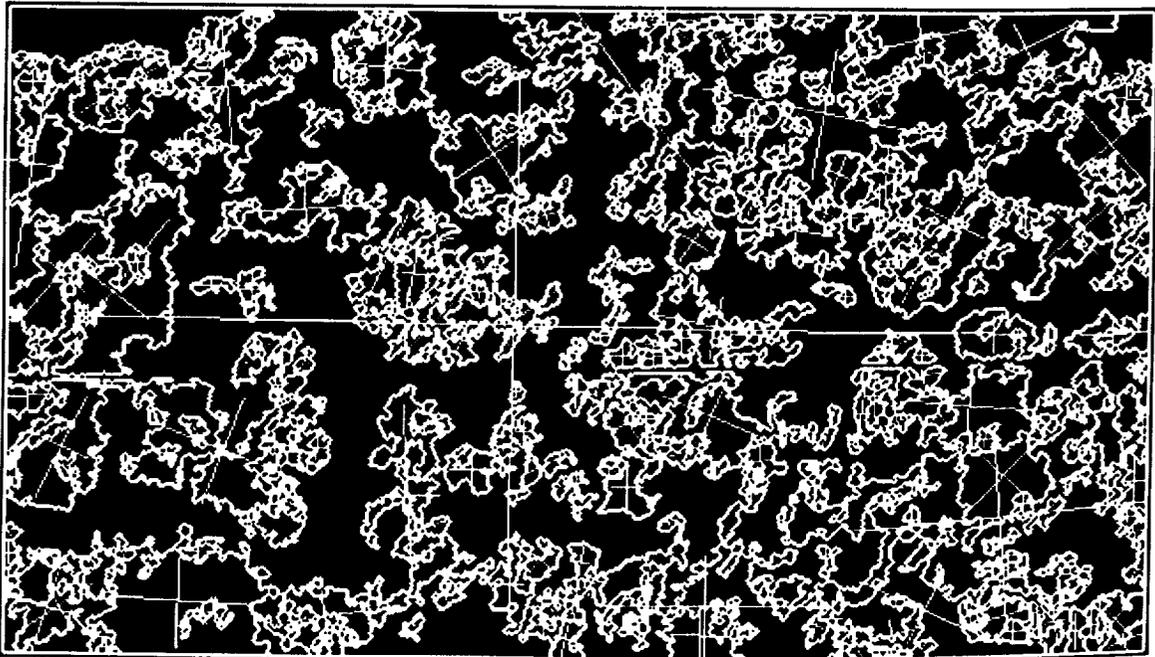
Figure 9. Region outline image.

23

4. Conclusion

This paper presents a new methodology for ground level vegetation analysis. It emphasizes an integrated approach using existing algorithms and introduces a new classifier, the approximated likelihood ratio detector. Some of the techniques used in the analysis of the NGL imagery include preprocessing, feature extraction, classification, and region analysis. The goal is to allow the scientist to correlate information obtained at the satellite sensing level with more detailed information contained in the NGL imagery. Existing techniques based on satellite imagery do not provide enough detailed information for a complete vegetation analysis. The approach presented here provides a means of accurately tracking and quantifying the vegetation changes across a range of different scales. Future development of this system includes the integration of spatial results of NGL images into GIS.

Acknowledgments

The authors thank the following members of the Khoros research group: Pascal Adam, Danielle Argiro, Ramiro Jordan, Donna Koechner, Carla Williams, Scott Wilson, and Mark Young; as well as the following members of the Sevilleta LTER project: James Brunt, James Gosz, Troy Maddux, Bruce Milne, Doug Moore, and Greg Shore.

Appendix

The Khoros system integrates multiple user interface modes, code generators, instructional aids, data visualization, and information processing to produce a comprehensive image processing research tool. This system can easily be tailored to other application domains because the tools of the system can modify themselves as well as the system. This attribute is important in a system that is designed to be extensible and portable.

The Khoros infrastructure consists of three major components: a high level user interface specification, methods of software development embedded in a code generation tool set, and an interoperable data exchange format and algorithm library. These basic facilities have been used to build a set of applications for performing image processing research, algorithm development, and data visualization. One of the most powerful features of the system is its high-level, abstract visual language.

Khoros is a successful demonstration of how development programming, end-user applications programming, information processing, data display, instruction, documentation, and maintenance can be integrated to build a state-of-the-art image/data processing and visualization software environment.

References

[1] R.A. Schowengerdt, *Techniques for Image Processing and Classification in Remote Sensing*, Academic Press, New York, 1983.

[2] J.W. Brunt and W. Conley, "Behavior of a Multivariate Algorithm for Ecological Edge Detection", *Ecological Modeling*, 49, 1990, pp. 179-203.

[3] D. Hush and J. Salas, "Classification with Neural Networks: A Comparison", *Proc. 11th Annual ISE Conference*, Albuquerque, NM, May 1989.

[4] D. Hush, "Classifiers: Results for I4-4I Case with 8 Dimensions", *Sandia National Laboratory Report*, Department 9133, January 5, 1989.

[5] R.M. Haralick, "Statistical and Structural Approaches to Texture", *Proceedings of the IEEE*, vol. 67, no. 5, May 1979, pp. 786-804.

[6] J.Y. Hsiao and A.A. Sawchuk, "Supervised Textured Image Segmentation Using Feature Smoothing and Probabilistic Relaxation Techniques", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 12, December 1989, pp. 1279-1292.

[7] K.I. Laws, "Rapid Texture Identification", *Proc. SPIE*, vol.238, 1980, pp. 376-380.

[8] J.T. Tou and R.C. Gonzales, *Pattern Recognition Principles*, Addison-Wesley Publishing Co., Reading, Massachusetts, 1974.

[9] B.T. Milne, "Spatial Aggregation and Neutral Models in Fractal Landscapes", Submitted to *The American Naturalist*, December 1989.

[10] M.D. Levine, *Vision In Machine and Man*, McGraw Hill, New York, 1985.

Integration of SAR and DEM data - Geometrical Considerations

Walter G. Kropatsch

Institute for Informatics

University Innsbruck

Technikerstr. 25/7

A-6020 INNSBRUCK

Austria / Europe

Abstract

General principles for integrating data from different sources are derived from the experience of registration of SAR images with DEM data. The integration in our case consists of establishing geometrical relations between the data sets that allow to accumulate information from both data sets for any given object point (e.g. elevation, slope, backscatter of ground cover, etc.).

Since the geometries of the two data are completely different they cannot be compared on a pixel by pixel basis. The presented approach detects instances of higher level features in both data sets independently and performs the matching at the high level. Besides the efficiency of this general strategy it further allows the integration of additional knowledge sources: world knowledge and sensor charateristics are also useful sources of information.

The SAR features layover and shadow can be detected easily in SAR images. An analytical method to find such regions also in a DEM needs in addition the parameters of the flight path of the SAR sensor and the range projection model. The generation of the SAR layover and shadow maps is summarized and new extensions to this method are proposed.

1 Introduction

Synthetic Apertur Radar (SAR) images differ strongly from conventional optical images by their image formation principle. Since SAR is an active sensor, image acquisition does not depend on local wheather conditions, which is a major advantage over all optical sensors especially in areas of the world that are often covered by clouds. Hence, many of the planned remote sensing systems include a SAR sensor. On the other side, SAR images are still very noisy data and are difficult to interpret by a photo-interpreter.

One of the reasons are the complex geometric distortions that are introduced by mapping the earth with a range projection. There exist several possibilities to remove these systematic distortions and to transform the SAR image into a map projection which should be easier to interpret. This process is called 'geocoding'. Several geocoding transformations are based on digital elevation models (DEM), especially in moutainous areas. Domik (1985) used image simulation; Raggam, Strobl, and Triebnig (1986) used squint angle condition and bundle adjustment; Meier and Nüesch (1986) used doppler information and target point velocity; Kwok, Curlander, and Pang (1987) used doppler information and a three pass resampling.

Most of the current approaches determine in a first step control points. A control point identifies the locations of one feature in reality in both data sets. It is represented by a pair of coordinates that is used to establish the geometrical correspondences between the two data sets. The set of all control points could be input to the geometric rectification procedure which combines interpolation and resampling or it could be simply used to localize a given object point in both data sets.

If both data sets show similar characteristics then similarity of image features can be used to find control points. But this is not possible if the data look completely different. In some cases one data set can be transformed such that the result shows optical resemblance with the other data set (at least locally). Then corresponding features can be detected by local similarity measurements (e.g. correlation). Most of the common methods compare pixel values of both images. Guindon (1987) and Sasse (1989) use automatic correlation to find the control points; Strobl (1986) uses manual matching for that purpose.

In the geocoding of SAR images, DEM data are often used to simulate the SAR geometry. But such geometric image transformations are computationally very expensive operations and have to be repeated not seldom to further adjust transformation parameters.

In Kropatsch and Strobl (1990) we have designed a different strategy to integrate SAR images and DEM data. The idea is based on the fact that real world objects are mapped differently in both data sets. Therefore we need also different operators to detect instances of the same object in the two data sets. Such feature detectors produce sets, F1 and F2, of image features with individual properties and with feature-to-feature relations in both data independently. Knowing the formation principles of the data sources, properties and relations of the features can be derived from properties and relations of the real world objects. Hence features from F1 and F2 can be matched via the corresponding objects in reality. This general approach has two main advantages:

1. Since image operators are derived from properties of real objects, the resulting features also relate to the semantics of these objects. Real world knowledge as a third information source can be efficiently applied for further processing.

2. The image-to-feature mapping reduces the data amount considerably, while, at the same time, the expressive power of the vocabulary to describe the image contents (e.g. the fea-

tures) increases. Less data can be matched with less computation and more features can be differentiated by the greater vocabulary. Moreover, this generalization (or abstraction) process can be repeated.

2   SAR processing

Global characteristics of a SAR image include the parameters of the range projection, the resolution, and the flight path of the sensor. DEM data do not contain information about the type of the surface cover which is an important constituent of the SAR image. SAR images of mountainous areas often show characteristic features that do not severely depend on the backscattering of the surface cover: **layover** and **shadow**. Due to the multiplication of signals, layover regions appear brighter than the surrounding regions. Shadow regions, which appear as dark regions in the SAR image, are independent from the backscattering of the imaged terrain.
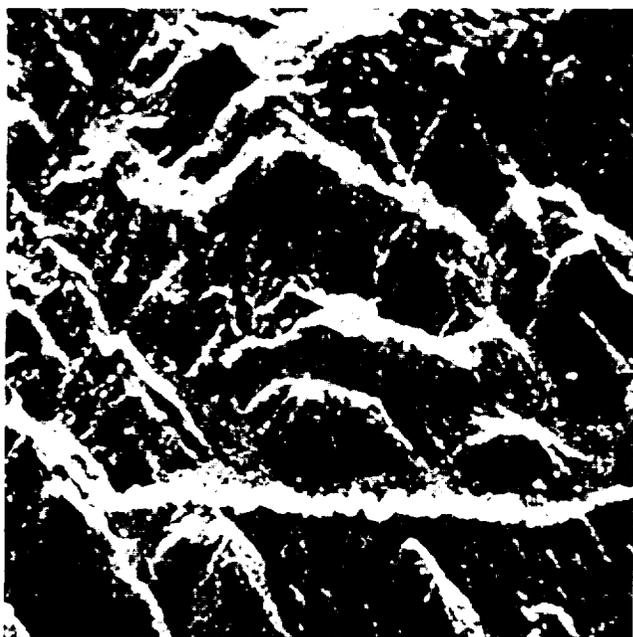
Figure 1: SAR-image of Iceland
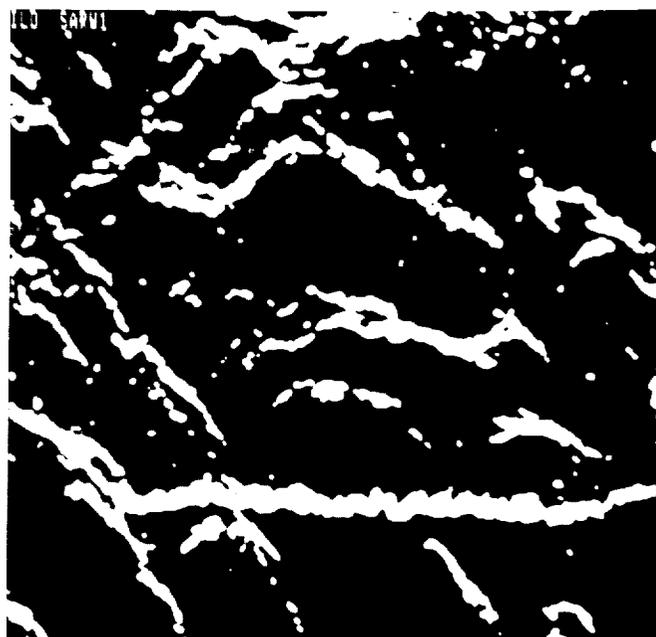
Figure 2: Layover regions

The flight path allows to distinguish between foreslopes and backslopes of the mountains. Foreslopes are oriented towards the sensor's path on the ground and are the (only) areas where layover can occur. Backslopes face the opposite direction and cannot be reached by the radar beam under certain imaging conditions (low sensor position or steep slope). Both types of features can be detected in a SAR image (Fig. 1) by a combination of noise elimination (i.e. filtering), thresholding, and connected component labeling (Plößnig, Billington, and Kropatsch, 1989; Plößnig, Kropatsch, and Strobl, 1989). The result is a set of layover and shadow regions in the SAR image (e.g. layover regions in Fig. 2).

29

3 DEM processing

Global parameters of a DEM are a reference coordinate system and the resolutions both in the ground plane $(\Delta x, \Delta y)$ and in the elevation (Δz). Using the parameters of the SAR flight path, a sequence of operations is applied to the DEM data in order to detect layover and shadow regions also in the DEM geometry. The mathematical model (Kropatsch and Strobl, 1990) is based on differential analysis of the range and look angle functions derived from a continuous terrain. The discrete implementation includes local differentiation and a pointwise computation of decision functions. A search is necessary to complete the layover and shadow regions by their passive parts.

3.1 Layover in the DEM

The characteristic radar measurements are **range** - the distance from the sensor to an object point - and **time** - the position of the sensor along its flight path where the data are collected. They define the two dimensional SAR image space.

Layover is called the radar mapping, where different object points having the same time and the same range are mapped into one image point, i.e. more than one points on the Earth's surface are mapped into one image point (many to one mapping). SAR mapping is mainly an integration of reflected signals having the same doppler frequency (azimuth or along track measurement) and the same distance (range or across track measurement).

In the object space the layover region splits into **active** and **passive** subregions. Active layover regions are the sources for layover (points that produce layover in the SAR image) whereas layover passives are only part of the layover because the active parts lay over them. The active layover region is embedded in two passive regions (called 'near passive' and 'far passive'). The calculation of the passive regions needs sequential search when no image simulation technique is used. In the image space there is no such distinction.

Since layover only occurs in an across track line (imaging time $t = const$) it necessitates to study profiles along iso-azimuth curves. At any time t, the (x, y, z)-DEM coordinate system can be transformed into a sensor ground-centered coordinate system (Fig. 3), where the sensor receives coordinates $(0, z_t)$ and any object point is located on a curve $(s, z(s))$. s measures the length of the ground projected iso-azimuth curve $(z = 0)$ between the nadir point (x_t, y_t) $(s = 0)$ and the object point (x, y).

The relevant SAR mapping equation for the slant ranges $r(s)$ is defined by

$$r(s) = \sqrt{s^2 + (z_t - z(s))^2} \tag{1}$$

In the iso-azimuth curve $z(s)$ of a SAR image the phenomenon of layover occurs, when the range $r(s)$ decreases by increasing nadir distance s (Fig. 3). If a plane terrain (parallel to the x, y-plane) is mapped this function is continuously increasing since the height difference of the sensor and the imaged object point is constant $((z_t - z(s)) = const)$. In hilly or mountainous terrain the height $z(s)$ takes different values. This circumstance produces layover in the SAR image when the height $z(s)$ increases faster than the nadir distance s. This region is bounded by a local maximum $r(B)$ and a local minimum $r(C)$ in the range (i.e. $\frac{\partial r(s)}{\partial s} = 0$). It is called the active layover region. It can

30
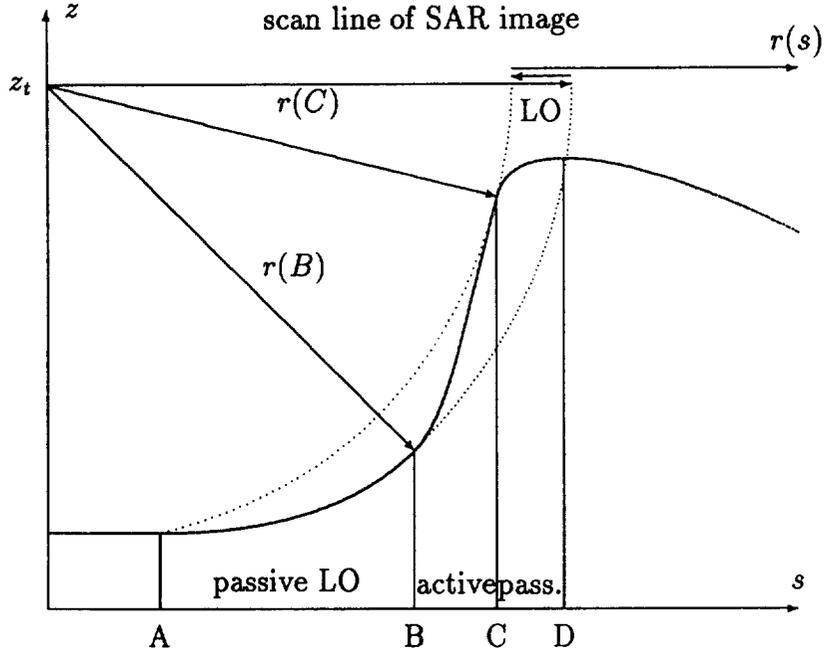
Figure 3: The elevation-range diagram of a SAR layover (LO).

be calculated by differentiating

$$\frac{\partial r(s)}{\partial s} = \frac{s - (z_t - z(s))\frac{\partial z(s)}{\partial s}}{r(s)} \qquad (2)$$

Since $r(s) > 0$ for all $s > 0$, a decision function $R(s) = r(s)\frac{\partial r(s)}{\partial s}$ can be determined which decides for a given object point $(s, z(s))$ whether it belongs to an active layover region or not:

$$R(s) = s - (z_t - z(s))\frac{\partial z(s)}{\partial s} \qquad (3)$$

$R(s) \leq 0$ defines the active layover subregions. $R(s) = 0$ defines the exact boundaries $s = B$ and $s = C$ of the active layover region.

Since layover occurs, when more than one object point is mapped into a single image point, regions $s < B$ and $C < s$ are also part of the layover. Knowing the maximum range $r(B)$ and minimum range $r(C)$ of a layover interval, the ranges of passive regions $[A, B)$ and $(C, D]$ have to be within this range interval too.

Figure 4 shows the layover regions that have been extracted from the DEM data. The area corresponds to the windows in Figures 1 and 2.

3.2   SAR shadow in the DEM

Shadow in a SAR image is called the region, where an object point is not reached by any radar beam. Such object points produce a 'zero' signal in the image. Therefore shadow regions appear in the SAR image as dark areas corrupted by noise.

31

Figure 4: Layover from DEM
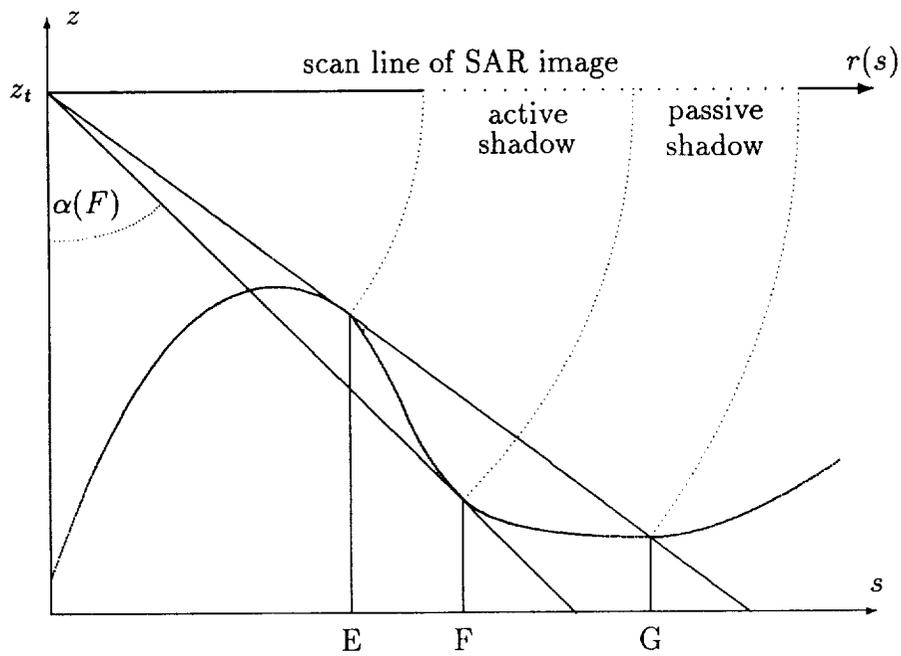
Figure 5: The elevation-range diagram of a SAR shadow

Among object points which are part of the shadow region we distinguish points belonging to the **active** (own shadow) region of an object or belonging to the **passive** (cast-shadow) region, which is produced by another object located closer to the sensor (Fig. 5). In the image space there is no such distinction.

Since shadow only occurs in an across track line (time $t = const$) it also necessitates to study iso-azimuth curves. The relevant SAR mapping equation for the look angle $\alpha(s)$ is defined by

$$\alpha(s) = arctan\left(\frac{s}{z_t - z(s)}\right) \tag{4}$$

In the iso-azimuth line of a SAR image the phenomenon of shadow occurs, when the look angle $\alpha(s)$ decreases by increasing nadir distance s (Fig. 5). If a plane terrain (parallel to the x,y-plane) is mapped, this function is increasing when the height difference of the sensor and the imaged object point is constant. In hilly and mountainous terrain, shadows appear in the SAR image wherever the height $z(s)$ decreases faster than the nadir distance s increases. This region is bounded by a local maximum $\alpha(E)$ and a local minimum $\alpha(F)$ (i.e. $\frac{\partial \alpha(s)}{\partial s} = 0$), and is called active shadow region. It can be calculated by considering the sign of the derivative

$$\frac{\partial \alpha(s)}{\partial s} = \frac{z_t - z(s) + s\frac{\partial z(s)}{\partial s}}{r^2(s)} \tag{5}$$

A simple decision function $A(\alpha)$ can be defined similar to $R(s)$, to decide whether a given object point $(s, z(s))$ belongs to an active shadow region or not.

$$A(s) = z_t - z(s) + s\frac{\partial z(s)}{\partial s} \tag{6}$$

$A(s) \leq 0$ defines the active shadow parts. $A(s) = 0$ defines the exact boundaries $s = E$ and $s = F$ of the active shadow region.

In contrast to layover where two passive regions occured here we have only one additional passive shadow region. It is located at the end of the active shadow, where the nadir distances s are increasing.

Both layover and shadow regions may also overlap. Two or more overlapping layover regions cause the multiplicity of the signals to further increase. Overlapping shadows result in the union of the single shadow regions in a mixture of active and passive parts. A shadow in a layover effects the multiplicity of the signal. All possible interactions between layover and shadow can be found in Strobl (1989) and Kropatsch and Strobl (1990).

3.3   Layover and Shadow Map

The discrete implementation of the above decision functions marks each cell of the DEM with labels layover, shadow or none, with further distinction between active and passive parts (Fig. 9 codes active layovers in white and passive parts in gray). Subsequent connected component labeling of this 'Layover and Shadow Map' (LSM) delivers the lists of regions corresponding to the regions in the SAR image.

We now summarize the algorithmic steps to calculate the LSM. The complete derivation is given in Kropatsch and Strobl (1990).

Let the DEM grid be defined by (i,j) with $i = 1, \ldots, n; j = 1, \ldots, m$; let the position of a grid cell be $(x_{i,j}, y_{i,j})$ with height $z_{i,j}$. The parameters of the flight path $(x_t(t), y_t(t), z_t(t))$ allow to compute the imaging time $t_{i,j}$ for every grid location by trilinear interpolation (Raggam 1988). The nadir distance $s_{i,j}$ can be calculated by

$$s_{i,j} = \sqrt{(x_t(t_{i,j}) - x_{i,j})^2 + (y_t(t_{i,j}) - y_{i,j})^2} \tag{7}$$

Using the spacing of the DEM grid $(\Delta x, \Delta y)$, the iso-azimuth direction $\phi_{i,j}$ can be approximated by central difference of the first derivative of imaging times

$$\cos \phi_{i,j} = \frac{t_{i+1,j} - t_{i-1,j}}{2\Delta x} \quad , \quad \sin \phi_{i,j} = \frac{t_{i,j+1} - t_{i,j-1}}{2\Delta y} \tag{8}$$

$$\tag{9}$$

Using central differences again, the terrain slope in across track direction becomes

$$\frac{\partial z_{i,j}}{\partial s} = \frac{z_{i+1,j} - z_{i-1,j}}{2\Delta x} \cos \phi_{i,j} + \frac{z_{i,j+1} - z_{i,j-1}}{2\Delta y} \sin \phi_{i,j}. \tag{10}$$

The above computations deliver the values $s_{i,j}$, $z_t(t_{i,j})$, and $\frac{\partial z_{i,j}}{\partial s}$ to compute the decision functions $R(s_{i,j})$ and $A(s_{i,j})$ for the active parts of layover and shadow respectively.

If the passive parts are needed, the precise boundary ranges, $r(B)$ and $r(C)$, $r(E)$ and $r(F)$ resp., must be interpolated for every range profile that crosses such regions. A search across track must be performed in order to delineate the boundaries of the passive parts.

4 Matching

The comparison and matching of the two sets of regions (i.e. three layover regions in Figures 6 and 7) includes the following measurements:

- For a single region:

  - type: layover or shadow;
  - the center (of gravity);
  - the size and orientation;
  - shape characteristics like
    * medial axis or
    * segments of the boundary.

- For a local configuration of regions:

  - the distances between pairs of regions;
  - the relative positions between pairs of regions;
  - the adjacencies.

Figure 6: 3 SAR-layover regions
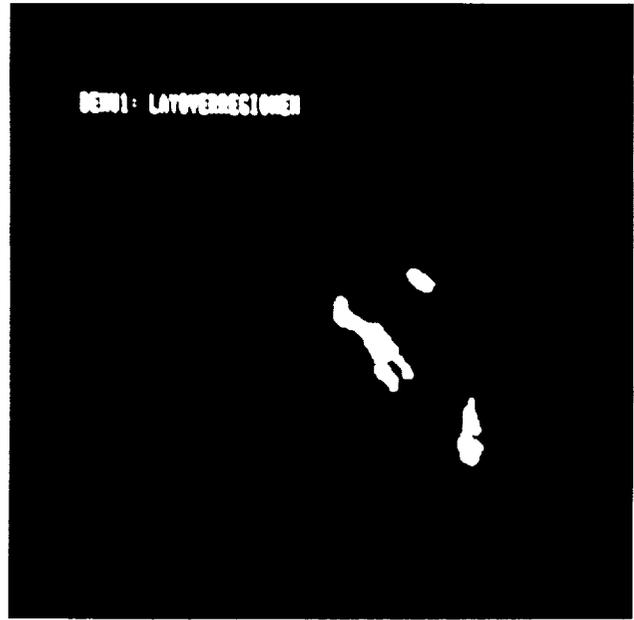
Figure 7: match 3 DEM-layover regions.

Base level (1): 512 × 512
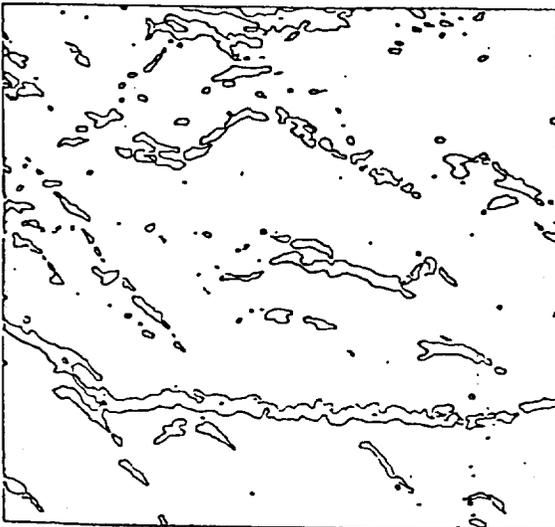
Level 5: 128 × 128
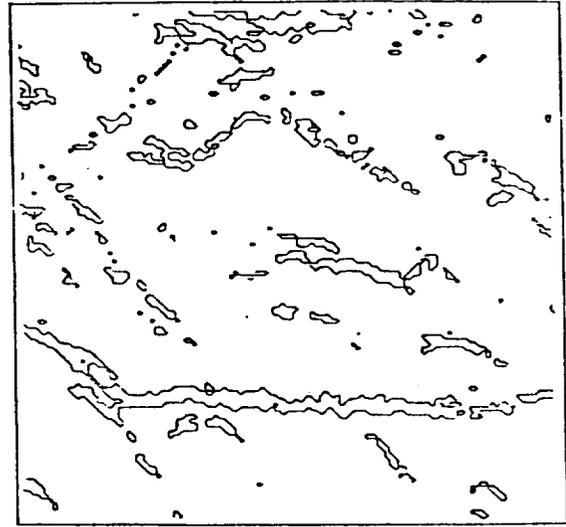
Figure 8: Levels 1 and 5 of layover curve pyramid

- For the entier image:
    - the parameters of the geometric mapping;
    - the accuracy of the resulting transformation.

An experiment using a $2 \times 2/2$ curve pyramid demonstrates a coarse-to-fine strategy for efficient layover matching (Fermüller and Kropatsch, 1989 and 1990).

In a curve pyramid, the boundaries of all layover (or shadow) regions are stepwise reduced in resolution. Such curve reduction (Kropatsch, 1985) preserves the connectivity but shrinks the length of the curve (Kropatsch, 1987). In the bottom-up building process closed boundaries survive until a resolution cell completely covers the corresponding region. We therefore continue reducing the resolution until only a few boundaries of large layover regions remain. Fig. 8 shows the base level (1) of a curve pyramid derived from the SAR layover regions of Fig. 2 and level 5 of this pyramid. All major shape characteristics are preserved while a lot of small detail, which is mostly due to noise, disappeared.

Building this curve pyramid for both the SAR- and the DEM-regions reduces the complexity of a rough matching to a couple of large regions that have to be compared with each other. The accuracy of that match is then stepwise refined in a top-down process, that uses the match approximation of the level above to match the higher resolution curves. If implemented on parallel hardware this automatic control point determination algorithm would require only $\mathcal{O}(\log n)$ computational steps.

5 Possible extensions and drawbacks of the method

There are several possible extensions to the proposed method. We just enumerate a few of them without investigating the details.

- The calculation of the shadow regions in the DEM could also be interesting for other types of images, e.g. optical images with shallow sun angle.

- Using smoothness constraints and backscattering characteristics from the surrounding of layover regions, the integral information in (small) layover regions could be separated into its constituent parts in the ground reference.

- The proposed method depends on the knowledge of the sensor's flight path for the calculation of both the layover and the shadow regions. To relax this requirement, the DEM could be preprocessed to preselect the potential shadow and layover points by only rough estimations of the flight altitude and flight direction. A variation of this preselection was useful in the acceleration of the sequential implementation of the algorithm.

- The weakest (computational) component of the algorithm is the (sequential) search for the boundaries of the passive regions. Although the search in parallel across track curves could be done in parallel it still depends on the diameter of the region.
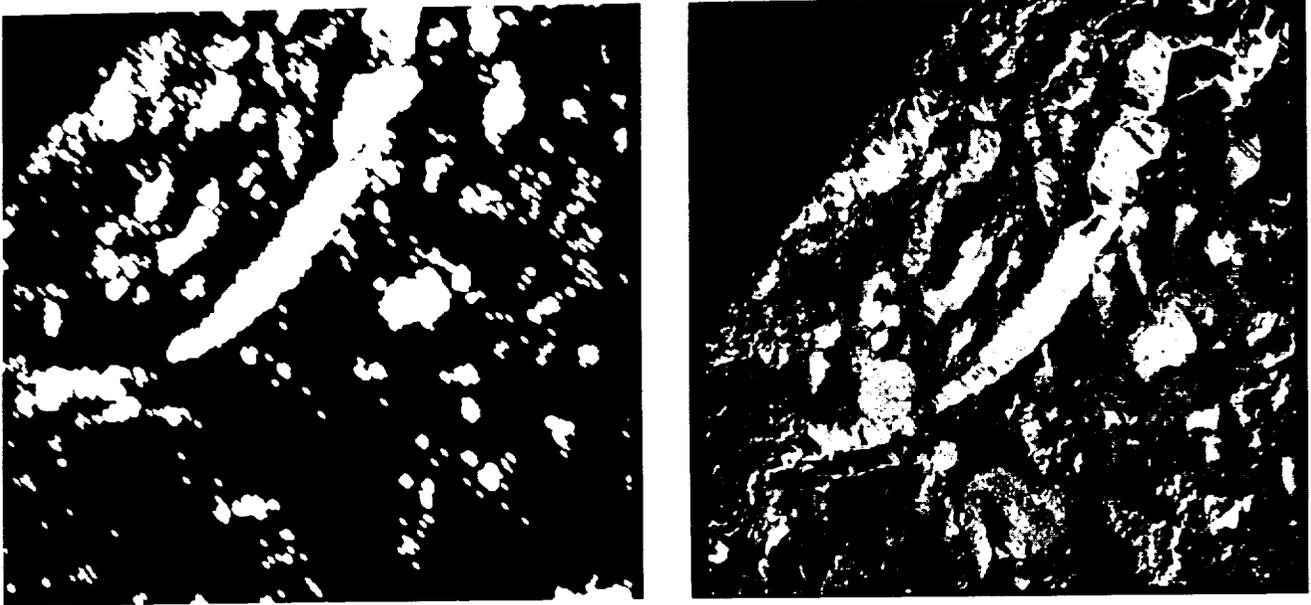
Figure 9: Layover and shadow map and geocoded SAR-image.

6 Conclusion

A feature-based approach to the integration of SAR-images and DEM data is presented. The features **layover** and **shadow**, which are characteristic for SAR images, are recognized independently in both data sets using properties of these features that are specific for the respective data set. Properties and relations of the resulting sets of layover and shadow regions allow to match both data sets. A curve pyramid of the region shapes is an example for an efficient coarse-to-fine strategy for matching. The resulting geometric correspondences allow the SAR image to be transformed into the (map-) geometry of the DEM ('geocoding', Fig. 9), or, they allow to measure properties of other (local) features directly in the SAR image, thus avoiding consequences of resampling errors, and relating these measurements to the corresponding location in the DEM.

7 References

[1] G. Domik. "Verfahrensentwicklung zur Analyse von digitalen Seitsicht-Radarbildern gebirgigen Geländes mittels digitaler Höhenmodelle und Bildsimulation." Ph.D. thesis, Tech. Univ. of Graz, Graz, Austria, Apr. 1985. Forschungsgesellschaft Joanneum, Graz, Austria, DIBAG Rep. 21.

[2] C. Fermüller and W. G. Kropatsch. "Hierarchische Kontur-Beschreibung durch Krümmung." In A. Pinz, editor, *Wissensbasierte Mustererkennung*, pages 171–187, OCG-Schriftenreihe, Österr. Arbeitsgemeinschaft für Mustererkennung, R. Oldenburg, 1989. Band 49.

[3] C. Fermüller and W. G. Kropatsch. "Multi-resolution shape description by corners." *Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990.

[4] B. Guindon. "Methods for automated control point acquisition in SAR images," in Proc. Intl. Workshop on SAR Image Rectification Techniques." Jan.14-16. 1987, pp. 29-32. Forschungsgesellschaft Joanneum, Graz, Austria, DIBAG Rep. 29.

[5] W. G. Kropatsch. *"Hierarchical Curve Representation in a New Pyramid Scheme."* Technical Report TR-1522, University of Maryland, Computer Science Center, June 1985.

[6] W. G. Kropatsch. "Curve Representations in Multiple Resolutions." *Pattern Recognition Letters*, Vol. 6(No. 3):pp.179–184, August 1987.

[7] W. G. Kropatsch and D. Strobl. "The Generation of SAR Layover and Shadow Maps From Digital Elevation Models." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 28, No. 1, pp.98–107, January 1990.

[8] R. Kwok, J. c. Curlander, and S. S. Pang. "Rectification of terrain induced distortions in radar imagery," Photogrammetr. Eng., vol. 53, pp. 507-513. May 1987.

[9] E. Meier and D. Nüesch. "Geometrische Entzerrung von Bildern orbitgestützter SAR-systeme." Bildmessung und Luftbildwesen, vol. 54, no. 5, pp. 205-216, 1986.

[10] M. Plößnig, B. Billington, and W. G. Kropatsch. "SHERLOCK - an intelligent system to locate control points in SAR" images and DEM. In *Proc. of the 2nd intl. Geosar-Workshop on 'Image Rectification for Spaceborne Synthetic Aperture Radar - Geocoded and Value-Added Products'*, Loipersdorf, Austria, January 1989.

[11] M. Plößnig, W. G. Kropatsch, and D. Strobl. "SHERLOCK Supports the Geocoding of SAR Images." In *IGARSS'89: Quantitative Remote Sensing: An Economic Tool for the Nineties*, pages 856–859, Vancouver, Canada, July 10-14 1989. Book 2.

[12] J. Raggam. "An efficient object space algorithm for spaceborne SAR image geocoding." in Proc. ISPRS Comm. II Symp. (Kyoto, Japan), July 1988, vol. 2, pp. 393-400.

[13] J. Raggam, D. Strobl, and G. Triebnig. "The rectification of SAR image data using a digital elevation model and image simulation techniques. Phase-B study for ERS-1 processing and archiving facility," Forschungsgesellschaft Joanneum, Graz, Austria, ESA Contract Rep. 6292/85/HGE-I, Tech. Note 17, Aug. 1986.

[14] V. Sasse. "Correlation of SAR data with optical data and with simulated data," in Proc. Intl. GEOSAR Workshop on Geocoded and Value-Added Products. Forschungsgesellschaft Joanneum, Graz, Austria, DIBAG Rep. 40, Jan. 1989.

[15] D. Strobl. "The effects of control point location errors on SAR geocoding in mountainous terrain," Diploma thesis, Technical Univ. of Graz, Austria, Jan. 1989. Forschungsgesellschaft Joanneum, Graz, Austria, DIBAG Rep. 37.

[16] D. Strobl. "Paßpunktbestimmung in Radarbildern," in Mustererkennung 1986, ÖCC-Schriftenreihe der Österreichischen Arbeitsgemeinschaft für Mustererkennung, vol. 36, pp. 206-235, 1986.

TOWARDS OPERATIONAL MULTISENSOR REGISTRATION

ERIC J.M. RIGNOT, RONALD KWOK, and JOHN C. CURLANDER

Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109, U.S.A.
Ph (818)-354-1640
telex 675-429

Abstract. To use data from a number of different remote sensors in a synergistic manner, a multidimensional analysis of the data is necessary. However, prior to this analysis, processing to correct for the systematic geometric distorsion characteristic of each sensor is required. Furthermore, the registration process must be fully automated to handle a large volume of data and high data rates. In this paper, a conceptual approach towards an operational multisensor registration algorithm is presented. The performance requirements of the algorithm are first formulated given the spatially, temporally and spectrally varying factors that influence the image characteristics and the science requirements of various applications. Several registration techniques that fit within the structure of our algorithm are then presented. Their performance was evaluated using a multisensor test data set assembled from the Landsat TM, SEASAT, SIR-B, TIMS and SPOT sensors. The results are discussed and recommendations for future studies are given.

1. Introduction

In future years a number of spaceborne remote sensing instruments will be operational. These instruments will gather data over a broad range of the electromagnetic spectrum allowing scientists to study the physical, chemical and electrical properties of the Earth's environment on a global scale and over an extended period of time. To derive geophysical parameters of interest for each of the planned science applications, the data collected by these sensors must be combined and analyzed in a multidimensional manner. However, the sensors may be on different platforms and in different orbits, have different physical characteristics, viewing geometries, and data collection and processing systems. Consequently, systematic and nonsystematic registration errors will exist between coincident multisensor data samples. It is a prerequisite for synergistic analysis of these data to remove such errors. Furthermore, because of the anticipated large data volume and high data rates of future high resolution sensors, it is necessary to develop an automated multisensor registration process that requires no or little operator supervision.

Considerable experience has already been accumulated in the operational registration of Landsat data (e.g. Grebowsky, 1979). However, these techniques are not well adapted to the registration of image data from multiple sensors of significantly different characteristics operating at different wavelengths. A robust and adaptable automated multisensor registration technique must be developed.

In this paper, a high-level algorithm that integrates several registration techniques is presented. First, a formulation of the performance requirements for development of an operational algorithm is given. These requirements are derived from the needs of several key science applications as well as a review of practical limitations given the image characteristics. We then describe the multisensor test data set that has been assembled for evaluation of several computational techniques that fit within the structure of our algorithm. One registration technique that has been evaluated uses high resolution digital elevation models (DEM) of the areas to be registered. Others, which operate in the absence of ancillary data, are based on the extraction and matching of scene features across the different images to be coregistered. The results are discussed and recommendations for future work are given.

2. Performance Requirements

2.1 Characterization of the Input Data

A number of spatially, temporally and spectrally varying factors influence the image characteristics and the registration accuracy.

Due to the finite precision of the estimate of the platform ephemeris and attitude, absolute location errors and geometric distortion affect the geometric quality of the imagery. Such errors can typically be removed by the use of tiepoints. However, nonsystematic errors and tiepointing bias the image location, and a final step of precision registration is required to achieve sub-pixel level accuracy. The geometric quality of the data is also affected by the presence of topography in the observed scene. For an active sensor like a synthetic aperture radar (SAR), predominant terrain-induced geometric distortions such as foreshortening and layover (Lewis and Mc Donald, 1970) constitutes additional difficulties. Rectification of these distortions is essential before registration of the data. As an illustration, a perspective view of geocoded and rectified multisensor imagery is shown in Figure 1 using a technique described in (Kwok et al., 1987).

For sensors on different platforms and in different orbits, the acquired data are intially sampled to grids that are more natural to the sensor geometry than that of multisensor registration. A common grid for image coregistration, such as an Earth-fixed grid, is required. The process of mapping image data into this grid is

40

known as geocoding and has been developed for a variety of sensors including SAR (Curlander et al., 1987).

Multisensor registration will be affected by the large variability in spatial resolution of the data to be registered (from tens of meters (SAR, HIRIS) to kilometers (MODIS, HMMR) in the future NASA Earth Observing System (EOS) platform). Since resolution defines the ability of a system to discriminate small details within a scene, it establishes a limit for the achievable registration accuracy.

System noise (i.e. thermal noise, quantization noise, bit error noise, etc.) will also affect the registration accuracy, because many of the techniques used for registration are very sensitive to noise. While all sensors are corrupted by additive noise from the receiver electronics, SAR images are additionally corrupted by multiplicative noise known as image speckle. Thus the multisensor registration techniques must be robust to noise of a variety of statistics.

An additional consideration in any registration scheme is the scene composition. In cases where only a few features can be positively identified across the various sensors, the registration accuracy may be seriously inpaired. Furthermore, identifiable features are inherently space, time and frequency dependent. Therefore, it is necessary to develop robust automated techniques of selection of invariant features across the multisensor data.

In view of the above remarks, the input and output data requirements for an operational algorithm can be formulated. They define the operational domain and conditions under which the multisensor algorithm is expected to operate, and can be used as a basis for the evaluation of candidate algorithms.

2.2 Input and Output Data Requirements

The input data shall be corrected from the geometric distortion characteristic of each sensor using the best information available, geocoded onto a preselected grid common to all sensors (e.g. UTM), and resampled to the same pixel spacing. The signal to noise ratio of the data shall be better than 5 dB. The geodetic accuracy of the input images shall be better than 500 meters or 10-50 pixels. It is expected that most sensors will do better than this since most of them will have an accurate geographical location system on board.

The output products shall have a registration accuracy of less than one resolution element. This requirement is derived from a subset of application being considered for multisensor data analysis.

In the case of change detection, sub-pixel accuracy is desired to compare the response of individual pixel elements. Depending on the scene characteristics (presence of identifiable features) this requirement may be very difficult to achieve. In other cases, for example in the global study of hydrological cycles (which includes tasks such as sea-ice identification and dynamics, determination of moisture content of soil and vegetation, vegetation identification, areal extent and growth, etc.) a registration accuracy of several resolution elements may be sufficient.

It is important to point out that although the accuracy requirements have typically been well defined for each individual instrument, little or no accuracy requirements have yet been clearly defined for multisensor registration by the scientific community (EOS, 1987). More work is clearly needed in this area for each interdisciplinary science application.

3. Multisensor Test Data

A multisensor test data set has been assembled using image products from SEASAT SAR, SIR-B SAR, Landsat TM, SPOT, and TIMS (Kwok et al., 1989). Information on each sensor, including look angle, spectral range, polarization and spatial resolution is given in Table 1. Geocoding of the images to a common UTM Earth-grid has been performed and the data have been resampled to the same pixel spacing of 25 meters. Several sub-images of reduced size (512 x 512, 1024 x 1024 pixels) were selected from the areas where the sensors have coincident coverage. The characteristics of the original image data and of the selected sub-images are presented in Table 2. This table includes information about the geographic location of the data, the initial sample spacing and size, the revolution number and date of acquisition, the number of selected sub-images and the type of map projection used for coding. A summary list of the natural features present in the imaged scenes is also indicated.

For each selected sub-image, manual registration was performed, resulting in an estimated relative misregistration uncertainty of less than ± 2 pixels, roughly equal to the largest resolution element (40 meters). This uncertainty results from the differences in resolution between the various sensors. This estimate is used as a basis for the true registration for quantitative evaluation of the performance of various automated registration techniques.

4. Automated Multisensor Registration

The structure of the candidate multisensor registration algorithms is presented on Table 3. The input data satisfy the requirements as formulated in the previous section. The first processing step consists of automatically selecting sub-frames from each input image to define local areas of multisensor coincident coverage where pre-

cision registration can be performed with a high confidence of success. Depending on the availability of ancillary data (DEM or cartographic maps) a registration mode is selected. For the case where DEM is available the multisensor data are coregistered to the common grid provided by the DEM. Otherwise, invariant features are extracted from the sub-images and correspondence is established across the data to be registered. To reduce the computational complexity of the algorithm and obtain several estimates of the misregistration per sub-image, feature matching is performed at multiple locations and the results are then filtered to evaluate their relative spatial consistency within the selected patch (local constraints). If the match can be labeled as statistically significant (e.g. satisfies some goodness measure), the misregistration error of the selected sub-image is estimated and the multisensor data are then registered. Otherwise, the result is rejected and the selection and matching process is repeated with different parameters. At a higher-level of processing, the combined results from different features and from registered neighborhood patches (global constraints) can be used to produce a more accurate and more reliable solution. In effect, a cooperative process can be established where the results from different stages of the processing are used as reinforcements for the entire process.

Several candidate techniques which are effective within this structure are presented in the remainder of this section. They have been selected based on compatibility, robustness and adaptivity to the various sensors. Each matching algorithms' performance is assessed using the multisensor test data set described in the last section.

4.1 Automated Selection of Sub-images

Selection of the patches where fine registration is desired must be based on the extraction of stable features that can be unambiguously identified across the entire multisensor image data set. The difficulty is to formulate an approach without $a-priori$ knowledge of the scene content. One possible technique has been described in (Davis and Kenue, 1978) where binary edge maps are used to compute a figure of merit for candidate control points. The results obtained with images from different sensors are then cross correlated to retain valid candidates.

4.2. Automated registration to digital terrain data

Our approach is to simulate multisensor imagery from a digital elevation model (DEM) of the area where the sensors have a coincident coverage and register this simulated imagery with the actual imagery, thereby inducing coregistration of the multisensor data on the common grid provided by the DEM.

Using elevation data, viewing geometry and a model of the scene reflectance, the appearance of the scene for any given sun angle and viewing angle can be simulated as in (e.g. Horn and Bachman, 1978) for passive sensors operating in the visible and near-visible part of the spectrum. An example of a synthetic image generated using this technique is shown in Figure 2. The illumination parameters were matched to a Landsat TM image data acquired over the same area. A simple matching technique (area-correlation) is then used to establish the correspondence between the images. The registration error is approximately 80 meters for the images shown.

Our approach to generate simulated SAR image from the DEM is similar to that described above. The sensor imaging geometry, the elevation data and a model of the radar backscatter are all required to produce the image shown in Figure 3. The imaging geometry simulates that of a SEASAT image acquired over the same area. An area correlation scheme is then used to match the radar and simulated images. Using tiepoint measurements of identifiable features not within the image shown, a misregistration error of 60 meters was obtained.

Several potential error sources affect the registration accuracy, including the the uncertainty in the actual imaging geometry, the geometric accuracy of the DEM data (height), and the reflectance model used for the optical data.

4.3 Computational Approaches

In the absence of reference maps, elevation data, geographical information or correlative ground truth information, blind techniques based on the identification of invariant features across the data can be used for image registration.

A. Feature extraction

Candidate features commonly used in digital imagery include edges, regions, lines, vertices of line intersection, shapes, etc. These features must be robust to change in sensor geometry, wavelength, SNR and noise statistics. Two particular types of features, region boundaries and edges, were examined using our multisensor data set.

Multisensor region boundaries extraction

Region boundaries are one of the simplest invariant low-level features than can be used to characterize the misregistration.

Even though many unsupervised segmentation techniques exist for optical images, most of them are not effective for SAR images because of the presence of speckle

noise. One unsupervised technique that seems to work reasonably well is a scheme based on a clustering algorithm to segment the images into several regions of similar intensity and texture (Kwok et al., 1989). The region boundaries are then established where a class transition occurs.

A resulting segmentation map, using 3 classes, is shown on Figure 4 together with the original images from SEASAT, Landsat, and SPOT. A 3 x 3 pixels window was used at each pixel location to compute the mean grey level and grey level texture via a simple standard deviation measure. The results obtained by matching these region boundaries are usually less accurate than those obtained with other techniques. However region segmentation can still be refined, especially in the case of SAR imagery, to provide information that complements results from other techniques.

Multisensor edge detection

An extensive literature exists on the subject of edge detection in optical imagery. However, in the case of SAR images the detection process is complicated since the images are corrupted by speckle noise. Techniques based on an approximation of the first and second directional derivatives (e.g. Sobel, or Robert operators) perform poorly, especially in terms of localization of the edges since they tend to produce large responses. Statistical edge operators (Touzi et al., 1988, Frost et al., 1982) in a lot of cases suffer from the same limitation.

This problem is solved by regularization techniques, specifically using a two-dimensional Gaussian smoothing operator as in a Marr-Hildreth operator (Marr and Hildreth 1980) or a Canny edge detector (Canny 1983). These operators typically have good detection and localization properties without multiple responses to a single edge, the three performance criteria for evaluation of edge detection algorithms. Theoretically, these techniques are compatible to almost all types of remote sensor data. Their performance with optical data have been documented in the literature (Marr and Hildreth 1980, Canny 1983).

The performance of these two operators was quantitatively compared in (Kwok and Rignot, 1989) in the case of synthetic SAR images as well as actual SAR images. It was shown that the gradient operator outperforms the Laplacian operator in both detection and localization of edges in image speckle.

Significant improvments in the performance of the ∇G operator can result from optimizing the parameter selections. In particular, the value of the filter spatial width σ must be adapted to the spatial resolution of the different sensors. Automatic thresholding is another important factor. In our implementation, a threshold with

hysteresis as in (Canny, 1983) is used to eliminate insignificant edges. Further post-processing such as thinning and contour-filling techniques have been shown to improve the quality of subsequent matches. Another possible improvement of the edge detector uses multiple operator widths and combines the resulting edges using a technique called feature synthesis, where the responses of the smaller operators are used to predict the response of a larger operator. Some results with optical images have been presented in (Canny 1983).

For illustration, one example of edge-map using SEASAT, Landsat TM and SPOT data and the Canny edge detector with a spatial width of 2 pixels (40 m) and adaptive thresholding is presented in **Figure 5**.

B. Feature Matching

Candidate feature matching techniques include binary cross correlation, distance transform / Chamfer matching, dynamic programming, and structural and symbolic matching.

In the case of region boundaries and edges a convenient binary representation of the feature maps can be used, a grey level of one at location of a feature-point and zero otherwise. This representation reduces the computational complexity of feature matching since computational cost becomes proportional to a linear dimension as opposed to area correlation where computational cost is proportional to an area.

Binary correlation

The binary feature-maps of each of the images to be registered can be cross correlated for various relative image shifts. The shift corresponding to the peak of the cross correlation will be an estimate of the actual misregistration between the images. The process is fast and can be efficiently implemented on an array processor or vectorizing computer.

Distance transform and Chamfer matching

The distance transform and Chamfer matching are described in (Barrow et al. 1977). In this method feature-points are matched by minimizing a generalized distance between them.

A distance transform is first applied to a binary feature-map, arbitrarilly refered as the source image. The result of this transformation is a distance map where the grey level of each pixel is a measure of the distance between the pixel and the nearest feature-point. For various values of the relative shift between the source and

the target images, the total distance between the feature points of the two images can be computed. This measure is the sum of the distance values read from the source image at each location of a feature-point in the target image. If matching were perfect, this distance would be zero. The relative shift that produces the smallest total distance corresponds to an estimate of the actual relative translational misregistration between the images to be registered.

This method is more robust to distortion or residual rotation effects than a binary correlation method.

Comparison of binary correlation and Chamfer matching

The time of computation of the binary correlation is less than the time of computation of Chamfer matching, typically in the ratio 1 to 4 for a search area of 100 x 100 pixels using a 512 x 512 pixel image. The tolerance to residual rotation effects is of 1 degree in the case of the binary cross correlation based on a maximum registration accuracy of 2 pixels. This tolerance is improved to 3 degrees when thicker edges are used (3 pixels wide instead of 1) (Wong, 1977). In the case of Chamfer matching the rotation tolerance is of 3 degrees. Better registration results (10 to 20 %) were consistently obtained by binary correlation as compared to Chamfer matching. The reason is that the quality metric used during Chamfer matching does not perform as well as expected with multisensor data due to the presence of non-matchable edges across the data, i.e. edges that appear in one image and not in the other. Their presence biases the total distance between feature points and significantly affects the accuracy, whereas the binary cross correlation is not affected by non-matchable edges.

Dynamic Programming

This iterative method, combined with an autoregressive model (AR), was used in the work by (Maitre and Wu, 1989) to register severely distorted optical images to a reference map without *a-priori* knowledge of the distortion. The two processes work at a different level. The AR model defines the deformation of the image at a pixel scale, and dynamic programming optimizes the search for best registration of an ordered sequence of features or primitives (usually edges) with a comparable sequence of features extracted from a reference map. The technique is robust in the presence of non-matchable edges. Good results are shown in (Maitre and Wu, 1989) using NOAA-7 satellite data.

This method has not been tested yet using the multisensor test data set, but offers good potential.

C. Constraint Filtering

In practice, matching is performed on small areas (typically 256 x 256 pixels or less) to minimize the distortion. Thus, the time of computation is also reduced and the number of estimates of the misregistration between the two images is increased. The resulting data must therefore be filtered to eliminate false matches. A clustering technique can be used where the cluster centroid corresponds to the estimated misregistration of the images.

At a higher level, results obtained from several feature matches are used to improve clustering of the data. Results from neighborhood patches can also be included.

4.4 Experimental Results

Twelve 512 x 512 pixel images corresponding to 3 different geographic areas have been registered. Each image was divided into 4 sub-blocks, and the search area for the local registration shift was 101 x 101 pixels in each sub-block.

In the case of the images from SEASAT and SPOT, the rate of success of the binary correlation of edges was 87 %, and increased to 92 % after constraint filtering, with no false matching. In the case of images from SEASAT and Landsat TM, the rate of success of the same technique was 85 % before constraint filtering, and 86 % after. Registration was qualitatively more difficult in that case because of the lower resolution of the Landsat images as compared to SPOT images, and also because a few additional scenes where registration was more difficult was used.

The registration accuracy of the multisensor data was approximately ± 2 pixels (40 m). The achievability of sub-pixel accuracy seemed difficult to establish by visual inspection of our multisensor test data set, a fact that is a common problem when comparing digital imagery from multiple remote sensors.

5. Conclusions and recommendations

It is of considerable importance to develop automated multisensor registration tools for synergistic use of the data from a variety of spaceborne sensors. A high level algorithm that integrates a variety of registration techniques in a systematic manner was presented in this paper. It was tested using a somewhat limited multisensor test data set. A more complete study would enlarge this data set to include more instruments and more scene types. Additional techniques for feature extraction and feature matching also need to be evaluated in a follow-on study.

The performance of a multisensor registration algorithm is dependent on the sci-

ence requirements of the particular applications as well as the characteristics of the instrument, the imaged surface and the environmental conditions. This very complex task cannot be solved with just a single technique, but will require combining several techniques together that work in a competitive-cooperative mode of interaction. For this reason it seems logical that a rule based artificial intelligence approach may be necessary for the high level algorithm to select the optimal techniques and parameters from a particular multisensor application.

Acknowledgment

References

Barrow, H. G., J. M. Tenenbaum, R. Bolles, H. C. Wolf, 1977, Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching. *Proc. 5th Joint Conf. on Artificial Intelligence, Cambridge, Mass., 1977*, 659-663.

Canny, J. F., 1983, Finding edges and lines in images. Artif. Intell. Lab., Mass. Inst. Technol., Cambridge, MA. Tech Rep. AI-TR-720.

Curlander J. C., R. Kwok and S. Pang, 1987, A Post-Processing System for Automated Rectification and Registration of spaceborne SAR Imagery, *Int. Journal of Remote Sensing*, 8, 4, 621-638.

Davis, W. A. and S. K. Kenue, 1978, Automatic Selection of Control Points for the Registration of Digital Images, *Proc. of the Int. Joint Conference on Pattern Recognition, 4th, Kyoto, Japan, Nov 7-10 1978*, 936-938.

EOS Earth Observing System Reports, NASA Techn. Memorandum 86129, Vol. I and Vol II, NASA Publish., 1987.

Frost V. S., K. S. Shanmugan, and J. C. Holtzman, 1982, Edge Detection for Synthetic Aperture Radar and other Noisy Images. *Digest from the Intl. Geoscience and Remote Sensing Symposium.*

Grebowsky, G. J., 1979, Lacie Registration Processing, *The LACIE Symposium, Proceedings of the Technical Sessions, July 1979*, 87-97.

Horn B.K.P. and B. L. Bachman, 1978, Using synthetic images to register real images with surface models, *Comm. ACM, 11 Nov. 1978,* **21**, 914-924.

Kwok, R., J. C. Curlander and S. Pang, 1987, Rectification of Terrain Induced Distortion in Radar Imagery, *Photogrammetric Engineering and Remote Sensing,* **53**, 5, 507-513.

Kwok, R. and E. Rignot, 1989, Comparison of the $\nabla^2 G$ and ∇G Operators for Edge Detection in Speckle Noise, unpublished.

Kwok, R., E. Rignot, J.C. Curlander and S. Pang, 1989, Multisensor Image Registration: A Progress Report, Jet Propulsion Laboratory Internal Report JPL D-6697, September 1989.

Lewis, A.J. and H.C. Mac Donald, 1970, Interpretive and mosaicking problems of SLAR imagery, *Remote Sens. Environ.,* **1**, 213.

Maitre. H. and Wu, Y., 1989, Dynamic programming algorithm for elastic registration of distorted pictures based on autoregressive model, *IEEE Trans. on Acoust., Speech, and Sign. Proc.,* **37**, 2, 288-297.

Marr, D. and E. Hildreth, 1980, Theory of edge detection. *Proc. Royal Soc. London,* B, **207**, 187-217.

Touzi, R., A. Lopes and P. Bousquet, 1988, A Statistical and Geometrical Edge Detector of SAR Images. *IEEE Trans. Geosci. and Remote Sens.,* **26**, 6, 826-831.

Wong, Y. R., 1978, Sequential Scene Matching using Edge Features. *IEEE Trans. on Aerospace and Elect. Systems,* **14**, 1, 128-140.
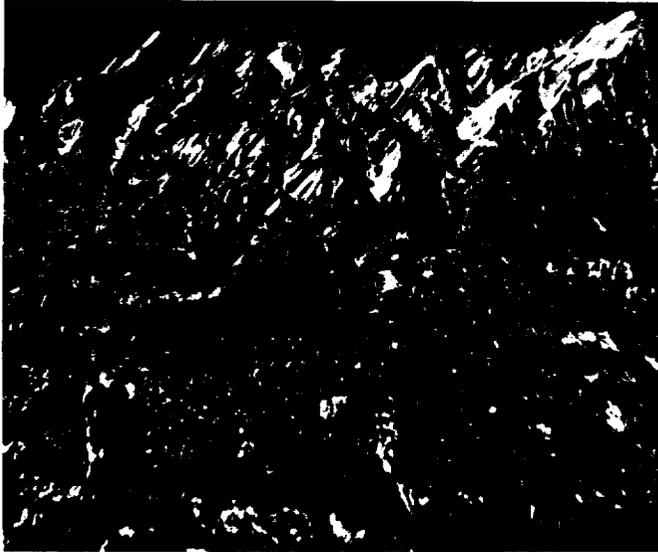
(a)   **(b)**

Fig. 1. Perspective viewing of multisensor geocoded and rectified images of an area near Los Angeles, California: (a) SEASAT radar image; (b) Landsat TM, band 4, image.

(a)   **(b)**

Fig. 2. Comparison of simulated versus actual Landsat TM image: (a) simulated image: (b) actual image.

(a)   **(b)**

Fig. 3. Comparison of simulated versus actual SEASAT radar image: (a) simulated image: (b) actual image.
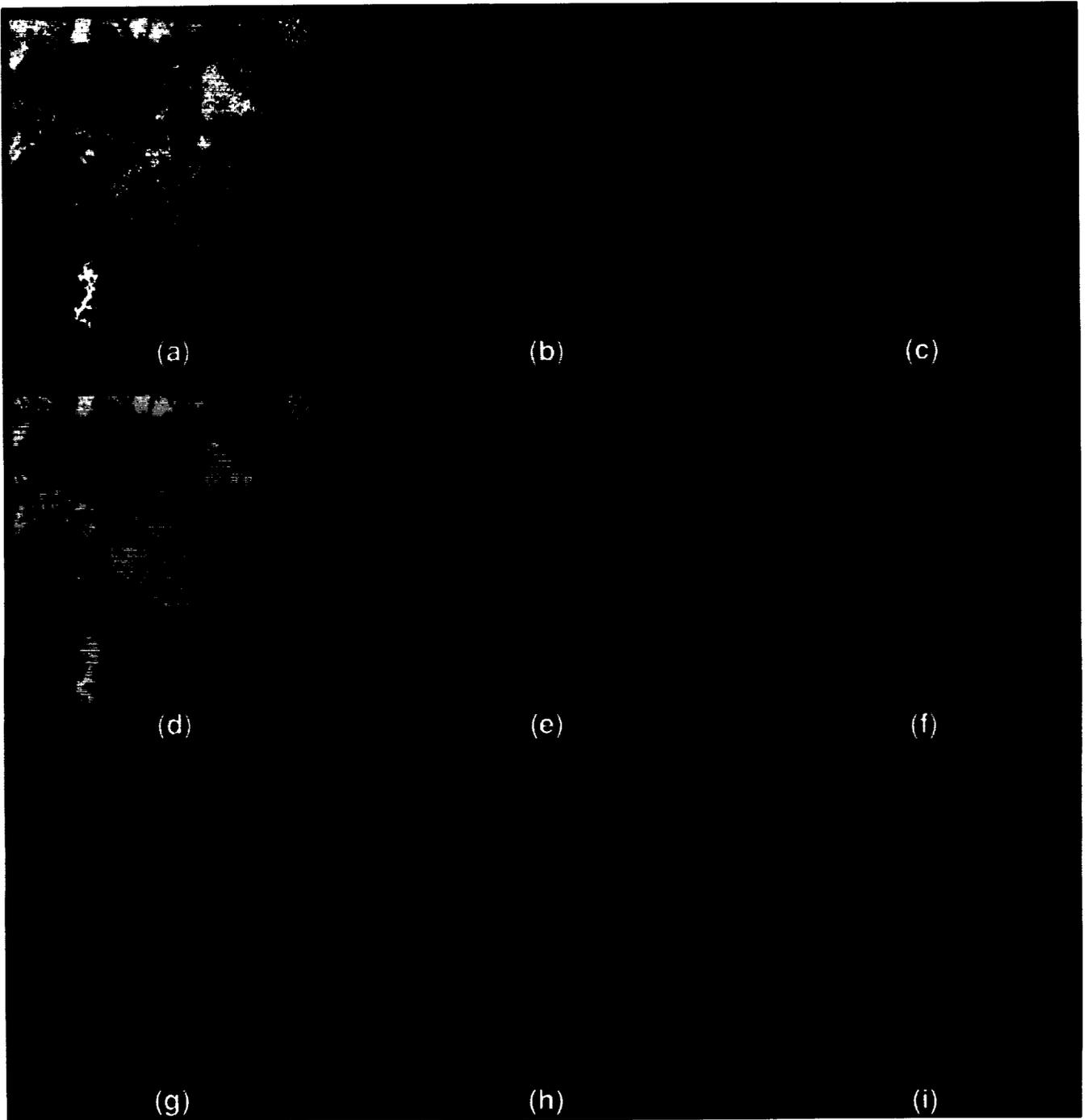
Fig. 4. Unsupervised segmentation of Optical and SAR images from an area near the Altamaha River, Georgia. (a) SEASAT, (b) Landsat TM, and (c) SPOT images are segmented into 3 regions represented in (d), (e), and (f) respectively. The corresponding images of region boundaries are (g), (h), and (i) respectively.

54
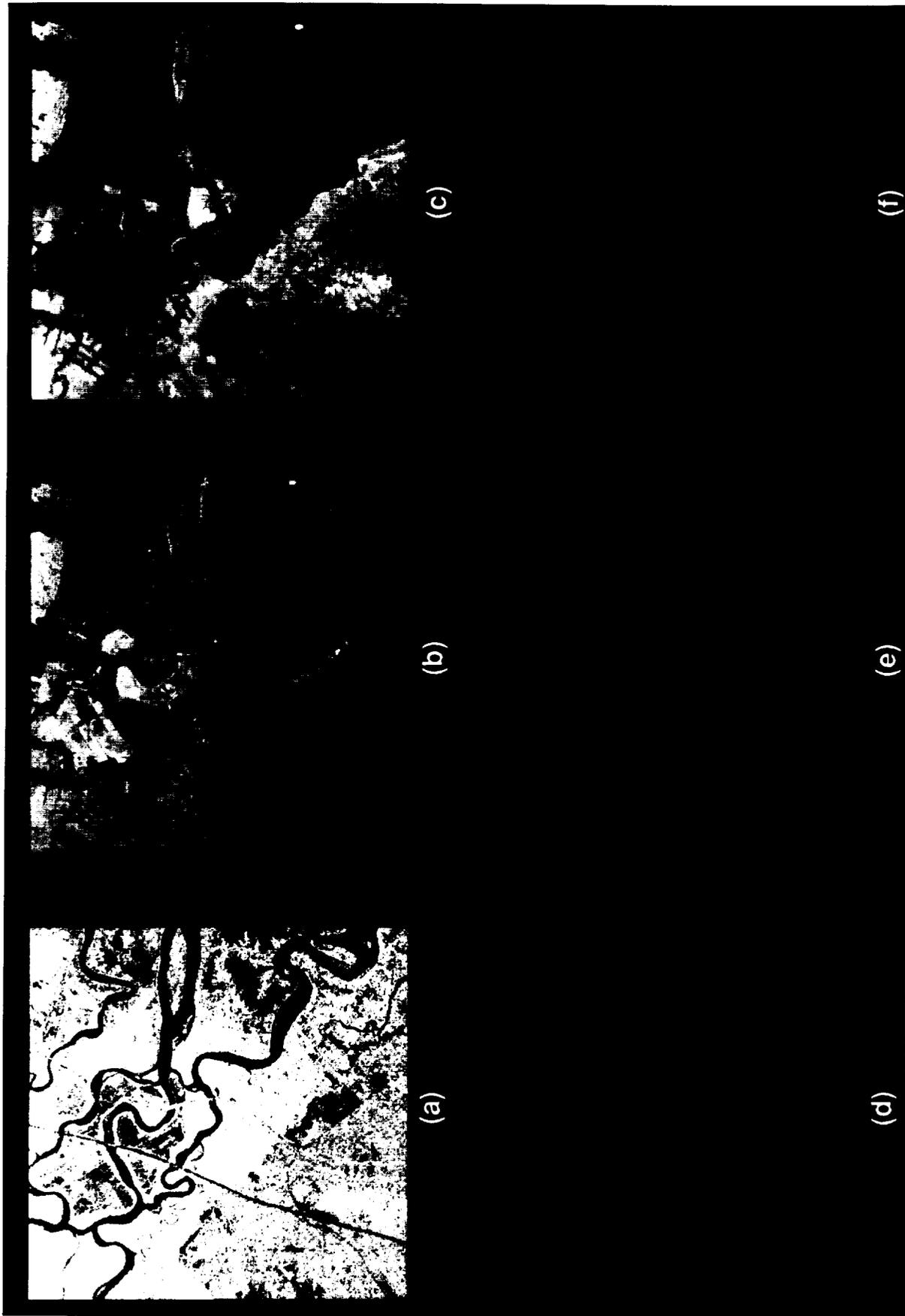
Fig. 5. Edge detection in (a) SEASAT, (b) Landsat TM, and (c) SPOT images of an area near the Altamaha River, Georgia. The edge-maps obtained from Canny's edge detector are represented in (d), (e), and (f) respectively.

55

Table 1. Description of the characteristics of the different sensors involved in the constitution of the multisensor test data set.

JPL   MULTISENSOR TEST DATASET

NAME OF SENSOR	TYPE	FREQUENCY	RESOLUTION (ORIGINAL DATA)
SEASAT	SAR ACTIVE	L BAND, HH POLARIZATION 23° LOOK-ANGLE	25 m
SIR-B	SAR ACTIVE	L BAND, HH POLARIZATION 15 TO 60° LOOK-ANGLE	25 m
TIMS	RADIOMETER PASSIVE	THERMAL-INFRARED	30 m
LANDSAT TM	OPTICAL PASSIVE	7 BANDS SPECTRAL RANGE: 1 .45 μm - .52 μm 6 1.04 μm - 1.25 μm 2 .52 μm - .60 μm 3 .63 μm - .69 μm 7 2.08 μm - 2.35 μm 4 .76 μm - .90 μm 5 1.55 μm - 1.75 μm	28.5 m (pixel spacing) IFOV
SPOT	OPTICAL PASSIVE	3 BANDS SPECTRAL RANGE: 1 .50 μm - .59 μm 2 .61 μm - .68 μm 3 .79 μm - .89 μm	20 m (pixel spacing) IFOV

Table 2. Description of the multisensor test data set.

MULTISENSOR TEST DATASET

IMAGE FRAMES LOCATION	SEASAT	LANDSAT	SPOT	TIMS	FEATURES IN PATCHES
ALTAMAHA RIVER, GEORGIA (pixel size = 20 m)	Rev : 407 Date: Jul 78 Size: 5 K x 5 K Map Proj.: UTM # of patches selected for testing: 12	Date: Jul 84 Size: 3 K x 4 K Rotated to North # of patches selected for testing: 12	Date: Sept 84 Size: 3 K x 5 K Rotated to North # of patches selected for testing: 12		Rivers, Lakes, Fields, Roads, Coasts.
WIND RIVER BASIN, WYOMING (pixel size = 30 m)	Rev : 781 Date: Aug 78 Size: 5 K x 5 K Map Proj.: UTM # of patches selected for testing: 5	Date: Jun 84 Size: 5 K x 5 K Rotated to North # of patches selected for testing: 13			Mountains, Rivers, Lakes, Fields, Roads, Cities.
YUMA, ARIZONA (pixel size = 25 m)	Rev : 681 Date: Aug 78 Size: 3 K x 3 K Map Proj.: UTM # of patches selected for testing: 13	Date: Jun 84 Size: 3 K x 4 K Rotated to North # of patches selected for testing: 13			Mountains, Rivers, Fields, Roads, Cities, Dunes.
DEATH VALLEY, CALIFORNIA (pixel size = 25 m)	Rev : 882 Date: Aug 78 Size: 5 K x 5 K Map Proj.: UTM # of patches selected for testing: 4	Date: Nov 82 Size: 3 K x 4 K Rotated to North # of patches selected for testing: 4		Date: Jul 83 Size: 1 K x 1 K Rotated to North # of patches selected for testing: 4	Mountains, Fields.

JPL

57

Table 3. Flow chart of the multisensor registration algorithm.

JPL   AUTOMATED MULTISENSOR REGISTRATION

```
                                              ┌──────────────────────────┐
                                              │                          │
                                              ▼                          │
┌─────────────────────┐                 ┌──────────────┐   ┌──────────┐  │  ┌──────────────┐
│  INPUT IMAGE DATA   │   ┌──────────┐   │   FEATURE    │──▶│ FEATURE  │──▶│  CONTRAINT   │
│ (AFTER GEOMETRIC    │   │ AUTOMATED│   │  EXTRACTION  │   │ MATCHING │   │  FILTERING   │
│  CORRECTION,        │──▶│SELECTION │──▶│              │   │          │   │              │
│ GEOCODING AND       │   │ OF SUB-  │   └──────────────┘   └──────────┘   └──────────────┘
│  RESAMPLING)        │   │ IMAGES   │        ▲                                    │
└─────────────────────┘   └──────────┘   ┌──────────────┐                         │
                                         │FINE REGISTRA-│                         ▼
                                         │TION MODE     │              ┌──────────────────────┐
                                         │SELECTION     │              │   MISREGISTRATION    │
                                         └──────────────┘              │      ESTIMATE        │
                                              │                        └──────────────────────┘
                                              ▼                                    ▲
┌─────────────────────┐              ┌──────────────┐              ┌──────────────┐│
│  ANCILLARY DATA     │              │ REGISTRATION │              │  CONTRAINT   ││
│      BASE           │─────────────▶│ TO ANCILLARY │─────────────▶│  FILTERING   │┘
│ (D.E.M., MAPS, ..)  │              │    DATA      │              │              │
└─────────────────────┘              └──────────────┘              └──────────────┘
                                          ▲                               │
                                          └───────────────────────────────┘
```

58
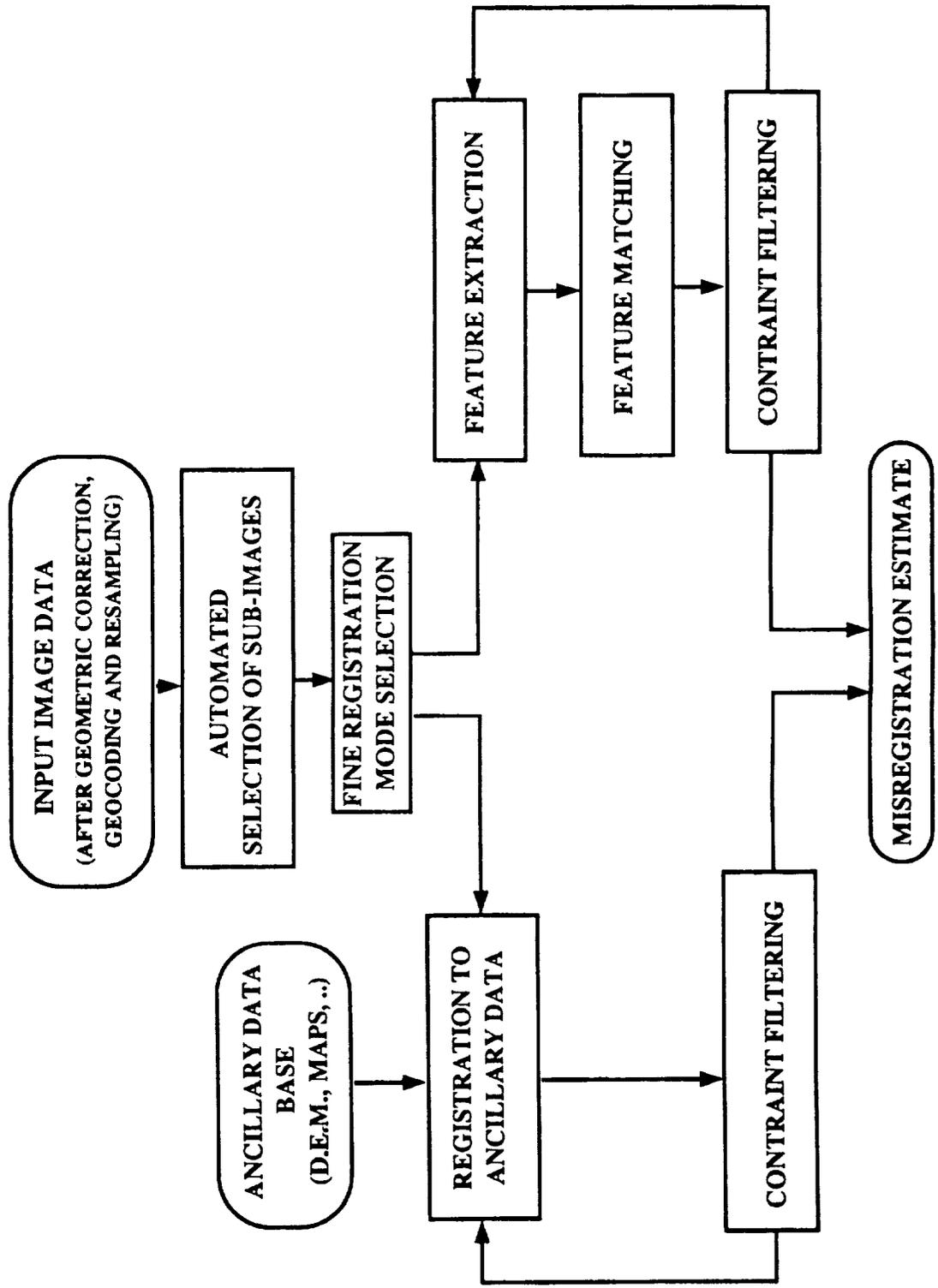
COMBINED FLUORESCENCE, REFLECTANCE, AND GROUND MEASUREMENTS OF A STRESSED NORWAY SPRUCE FOREST FOR FOREST DAMAGE ASSESSMENT

C. Banninger

Institute for Image Processing and Computer Graphics,
Joanneum Research
Wastiangasse 6, A-8010, Graz, Austria
Tel. (316) 8021, Telefax (316) 8021-20,
E-Mail: poelzleitner@rzj.fgj-graz.ada.at

The detection and monitoring of stress and damage in forested areas is of utmost importance to forest managers, who require timely and accurate information on the state of health and vitality of this natural resource for planning purposes. The extensive and often difficult-to-assess nature of many of the world's forested regions makes remote sensing the most suitable means to obtain this information. This requires that remote sensing data employed in a forest survey be properly chosen and utilised for their ability to measure canopy spectral features directly related to key tree and canopy properties that are indicators of forest health and vitality.

Plant reflectance in the visible to shortwave infrared regions (400-2500 nm) provides information on its biochemical, biophysical, and morphological make up, whereas plant fluorescence in the 400-750 nm wavelength region is more indicative of the capacity and functioning of its photosynthetic apparatus. A measure of both these spectral properties can be used to provide an accurate assessment of stress or damage within a forest canopy. What is necessary is to define the specific wavelengths within these spectral regions that provide optimal information on a plant's health and vigor.

Foliar chlorophyll and nitrogen are essential biochemical constituents required for the proper functioning and maintenance of a plant's biological processes. Chlorophyll-a is the prime reactive centre for photosynthesis, by which a plant converts carbon dioxide and water into necessary plant products. Nitrogen forms an important component of the amino-acids, enzymes, proteins, alkaloids, and cyanogenic compounds that make up a plant, including its pigments. The measurement in a canopy by remote sensing methods would allow the rapid appraisal of a forest's state of health. Both chlorophyll and nitrogen have characteristic absorption features in the visible to shortwave infrared region that furnish information on their content within a canopy. By measuring the wavelength position and depth of these features and the fluorescence response of the foliage, the health and vitality of a canopy can be ascertained.

For a stressed Norway spruce forest in southeastern Austria, foliar chlorophyll-a and nitrogen content and leaf area indices (LAI) were derived and foliage reflectance and fluorescence measurements obtained from samples collected at 50-m grid intervals over the test site. The discrete data sets were transformed into continuum data sets in the form of isopleth maps, and resolution cells corresponding to the size of the ground-projected pixels of the Landsat Multispectral Scanner (MSS) and Thematic Mapper (TM), NS001 Thematic Mapper Simulator (TMS), Thermal Infrared Multispectral Scanner (TIMS), Airborne Imaging Spectrometer (AIS-2), and Fluorescence Line Imager/Programmable Multispectral Imager (FLS/PMI) sensor systems derived for each ground data set. These cellularised data sets served as the reference data for evaluating the capabilities of the various remote sensing data used to differentiate stress or damage in the test site forest, which served as a test model for the development of remote sensing-based algorithms for more universal application. In addition to chlorophyll-a, nitrogen, and LAI data, other canopy biochemical constituents, canopy biogeochemistry, soil geochemistry, and site slope, aspect, and elevation information has been incorporated into the cell data bases. Examples of these will be presented in the paper.

A Phenomenological Approach to Multisource Data Integration: Analysing Infrared and Visible Data

N. Nandhakumar

Electrical Engineering Dept., Univ. of Virginia, Charlottesville, VA 22903

Abstract

A new method is described for combining multisensory data for remote sensing applications. The approach uses phenomenological models which allow the specification of discriminatory features that are based on intrinsic physical properties of imaged surfaces. Thermal and visual images of scenes are analyzed to estimate surface heat fluxes. Such analysis makes available a discriminatory feature that is closely related to the thermal capacitance of the imaged objects. This feature provides a method for labelling image regions based on physical properties of imaged objects. This approach is different from existing approaches which use the signal intensities in each channel (or an arbitrary linear or nonlinear combination of signal intensities) as features - which are then classified by a statistical or evidential approach.

1 Introduction

Multispectral/multisource data acquired via remote sensing have been shown to be useful for a variety of applications such as urban land-cover assessment, rain-rate classification, crop assessment, geophysical investigation, and surveillance and monitoring for national defence activities. Various techniques have been developed for combining the information in the different sensing modalities. These techniques typically use statistical or evidential rules to achieve the desired classification.

The usual statistical approach consists of first forming a feature vector wherein each element corresponds to the signal value (pixel gray level) from each sensor. This feature vector is then classified by a statistical decision rule. Other features such as the mean intensity level in a beighborhood, contrast, second and higher order moments, entropy measures, etc. have also been used as elements of the feature vector, e.g. [1]. In such approaches, interpretation of the imaged scene based on the fusion of information from the different sensors may be said to occur at the lower levels of analysis. In some techniques, linear and/or nonlinear combinations of signal values from different sensors form a feature, several of which are then fed to a classifier, e.g. [2]. In the latter case, interpretation may be said to occur at higher levels of analysis, after an earlier stage of information fusion which extracts discriminatory features. Other extensions to the standard statistical approach have been reported, e.g., a fuzzy relaxation labelling approach for image interpretation has been reported [3] wherein a Gaussian maximum likelihood classifier is used to provide initial probability estimates to the relaxation process.

Different optimal classification rules have been developed for interpreting multisource data for each of a variety of statistical models assumed for the data. The classifiers however do not address the problem of choosing *sufficiently* discriminatory features from the infinite number of available features. Such approaches therefore suffer from the disadvantage that the global optimality of the feature set is impossible to guarantee. Also, training of such classifiers is difficult since very large training data sets are warranted for achieving a reasonable error rate. It is also not clear what physical properties of the imaged objects are being utilized by the classifier during the discrimination process.

Evidential approaches have also been developed for combining information from multiple sensing
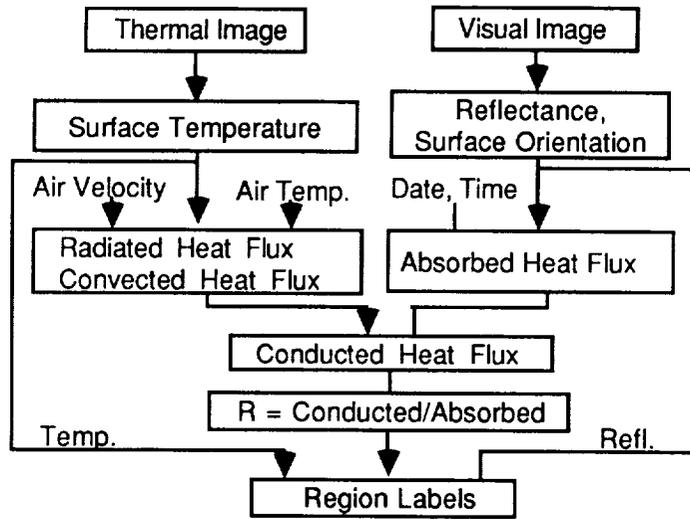
Figure 1. Combining thermal and visible data for surface heat flux estimation.

modalities, e.g. [4]- [6]. Such methods rely on a large set of heuristics rules which examine local contrast measures for each sensor and compare outputs from different sensors to provide varying degrees of support (certainty values) for a hypothesized class. A non-probabilistic framework is used for updating these uncertainties to reach a final classification. Interpretation in such systems is attempted at multiple levels of analysis. The rules, however, are based on manifestations of the differences in the intrinsic physical properties of objects rather than on direct measures of the physical properties themselves. Such approaches therefore do not fully exploit the synergy available in multisensor data fusion.

Due to these reasons, it is desirable to first combine information from the different sensors based on a physical model of the scene with the objective of evaluating intrinsic physical properties of the imaged objects. Such an analysis allows for specification of physically meaningful and discriminatory features which may then be used for scene interpretation by a probabilistic or evidential classifier at higher levels of analysis.

This paper discusses the development and use of phenomenological scene-sensor models for the fusion of information from infrared (IR) and visible data. A computational model is established in which principles of heat transfer are used along with computer vision techniques to derive a map of heat sinks and sources in the scene. The approach uses infrared imagery sensed in the $8\mu m - 12\mu m$ band, monochrome visual imagery, and knowledge of ambient conditions at the imaged surface to estimate surface heat fluxes in the scene. A feature which quantifies the surface's ability to sink/source heat radiation is derived and is shown to be useful in discriminating between different types of material classes such as vegetation and pavement.

It is assumed that the thermal image is segmented into closed regions by a suitable segmentation algorithm (e.g., [7]) and that the thermal and visual images are registered. The thermal image is processed to yield estimates of object surface temperature. This process requires the formulation of an appropriate model which relates scene radiosity to surface temperature, and received irradiation at the thermal camera to scene radiosity. Several object and scene parameters such as surface reflectivity,

emissivity, reflected scene radiosity are incorporated in the model. The visual image, which is spatially registered with the thermal image, yields information regarding the relative surface orientation of the imaged object. The Lambertian reflectance model is used along with the shape-from-shading principle for this purpose. The above information along with information regarding ambient temperature, wind speed, and the date and time of image acquisition is used in a computational model that allows estimation of surface heat fluxes in the scene. The estimated surface heat fluxes are used to evaluate a feature that is closely related to the lumped thermal capacitance of the object. This feature is shown to be a meaningful and discriminatory feature for scene interpretation. A block diagram of the approach is shown in figure 1.

Multisensory images, and in particular - thermal and visual images, have been used in the past for evaluating a rough estimate of thermal inertia for various remote sensing applications [8] - [10]. The previously developed methods use very simple models of the scene and of the energy exchange phenomena ocurring at the imaged scene. In contrast to these past approaches, the technique developed in this paper is based on an explicit and more detailed physical model of the energy exchange in the scene and provides more meaningful and discriminatory features for classification.

The remainder of this paper is organized as follows. Section 2 describes an approach for extracting accurate surface temperature estimates from infrared imagery. Section 3 discusses the computation of relative surface orientation from visual imagery. Section 4 describes the estimation of surface heat fluxes at the imaged scene. Section 5 discusses two different ways of using the surface heat flux estimates for scene interpretation. Section 6 presents experimental results using real data, and section 7 contains a summary of the ideas presented in this paper.

2 Estimating Temperature from Thermal Images

A quantitative model has been derived for estimating the surface temperature of a viewed object using the thermal image. Details of the derivation may be found in reference [11]. The salient points of this model are presented below. The model is based on observations that are unique to the situation where outdoor scenes are illuminated by solar radiation. The derivation of the model rests on the following observations and results:

1. Most surfaces found in outdoor scenes may be considered to be diffuse emitters in the $8\mu m - 12\mu m$ band. Furthermore, they possess high emissivities in this band - in the range of 0.82 to 0.96. Hence, a constant value of 0.9 may be assumed for the IR emissivity of all imaged surfaces in outdoor scenes.

2. The radiosity of an object's suface in a natural scene comprises surface emission, reflected solar radiation, and reflection of radiation that emanates from other surfaces. These components contribute to the total irradiation at the IR detector. Only 0.1% of the total solar energy lies in the $8\mu m - 12\mu m$ band. Furthermore, the surface reflectivities to IR radiation are very low. On the other hand, a large percentage of emission from scene objects lies in the $8\mu m - 12\mu m$ band since their surface temperatures lie typically between 250K and 350K. Since IR emissivities are also high the scene irradiation at the IR detector is dominated by emission from the surface of the imaged object. The components due to reflected solar radiation and reflected emissions from other objects may be safely ignored.

3. The view factor F_{oc} between the camera and imaged surface depends on the viewing geometry and typically involves the evaluation of complex integrals [12]. A reasonable approximation to
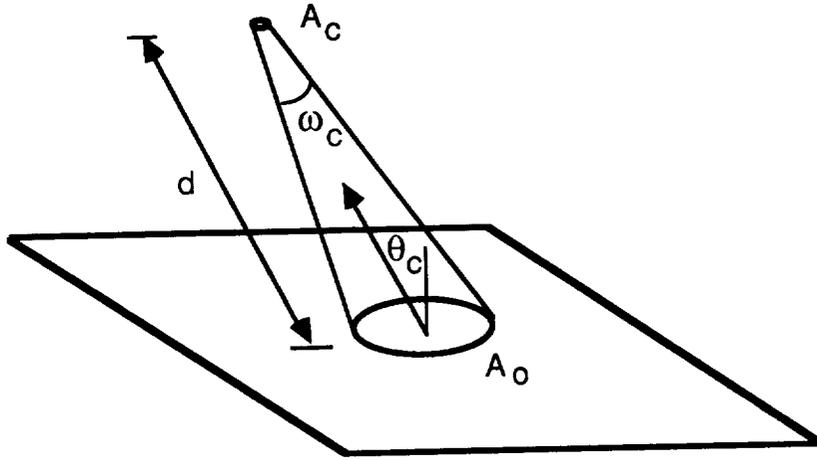
Figure 2. View factor between camera and imaged surface. A_c is the circular viewing surface of the detector, ω_c is the solid angle subtended by the detector, A_o is projection of A_c onto the imaged surface, d is the distance between the camera and the imaged surface, and θ_c is the angle between the surface normal and the viewing direction.

F_{oc} is arrived at by making the following observations. Since the solid angle subtended by the detector ω_c is usually very small (on the order of 2 mrad) we can approximate the projection of the detector's viewing surface onto the imaged surface to be a planar circular patch, denoted by A_o in figure 2.

The approximations indicated above allow for the derivation of a simple model that relates the surface temperature of the imaged object to the digitized value of the IR sensor's signal due to irradiation at the sensor [11]. The resulting model is expressed as:

$$0.9 \int_{\lambda_1}^{\lambda_2} \frac{C_1}{\lambda^5(exp(C_2/\lambda T) - 1)} d\lambda = K_a L_t + K_b \tag{1}$$

where, C_1 and C_2 are constants in Planck's equation and have values: $C_1 = 3.742 \times 10^8\ W\mu m/m^2$ and $C_2 = 1.439 \times 10^4\ \mu m K$. T is the surface temperature of the imaged object, λ is the wavelength of energy, $\lambda_1 = 8\mu m$ and $\lambda_2 = 12\mu m$. L_t is the pixel gray level value of the digitized thermal image. K_a and K_b are constants for a particular imaging setup and are obtained by appropriate camera calibration as described below.

The model established above provides a simple algorithm for surface temperature estimation. At the outset a table of values of $F(T_i) = 0.9 \int_{\lambda_1}^{\lambda_2} \frac{C_1}{\lambda^5(exp(C_2/\lambda T_i)-1)} d\lambda$ is created for different values of T_i via numerical integration of this expression. A scene containing two objects at two different known temperatures is imaged. The corresponding gray level values are used in equation (1) to solve for the constants K_a and K_b. Thereafter, the temperature of other surfaces in other scenes can be determined by first evaluating the right-hand side of equation (1) using the corresponding gray level. Then the table of values of $F(T_i)$ created above is looked up for a matching value. The value of the associated index T_i now provides the surface temperature estimate.

64

In general, an exact match will not be found and a linear interpolation is performed as follows to acquire a reasonably accurate estimate of the surface temperature. Consider a particular gray level value L_t in the thermal image which corresponds to a surface temperature of T_s. Let the right-hand side of equation (1) evaluate to G, i.e., $G = K_a L_t + K_b$. On searching for a match in the table of values of $F(T_i)$ assume that G is found to lie between adjacent entries $F(T_m)$ and $F(T_n)$ such that $F(T_m) < G < F(T_n)$. The desired value of surface temperature is then computed as:

$$T_s = T_m + \frac{T_n - T_m}{F(T_n) - F(T_m)}(G - F(T_m)) \tag{2}$$

In deriving the above approach for temperature estimation, the effect of atmospheric attenuation has been ignored. This is justifiable for the following imaging situations:

1. The surfaces whose temperatures are to be estimated appear in the same scene that was used for calibration.

2. The distance between the calibration surfaces and the thermal camera is the same as the distance between the surfaces whose temperatures are to be estimated and the camera.

3. The distance between imaged surfaces and the thermal camera is on the order of only a few hundred meters [13].

If neither of the above conditions apply, appropriate models need to be applied to account for atmospheric attenuation loss [13]- [15].

3 Inferring Surface Reflectivity and Relative Orientation

In order to estimate heat fluxes it is necessary to estimate not only surface temperature as described above but also surface reflectance to visible radiation and also surface orientation relative to the incident (solar) radiation. The visual image of the scene provides clues to both these quantities. In the following discussion it is assumed that the infrared and visual images of a scene are spatially registered, and that the images are segmented *a priori* into regions by a method such as that described in [7].

The use of shading information to recover the shape of an object has been addressed by several researchers, e.g. [16]-[20]. These techniques, however, can be applied only if certain conditions are satisfied. The bi-directional reflectance distribution function of the surface must be known *a priori*. Image resolution must be high enough to allow the rendition of several surface patches near the occluding boundary, or there need to exist background patches of known surface orientation surrounding the region of unknown surface orientation. These conditions are difficult to satisfy when imaging objects in a natural scene. We also note that while the aforesaid efforts attempt the problem of determining the (x, y, z) direction cosines of the surface normals of the imaged surface, our problem is a much simpler one, i.e., to arrive at an accurate estimate of $cos\theta_i$, where θ_i is the angle between the surface normal and the direction of incident radiation. A simpler method may be used for this purpose as described below.

Real surfaces in outdoor scenes exhibit a combination of diffuse and specular reflectivities. The diffuse component has been found to dominate in commonly occurring surfaces [21]. Hence, it is reasonable to assume that the imaged surfaces are Lambertian reflectors. If L_v represents the gray

level of a pixel in the visual image, the relative surface orientation of the surface patch corresponding to that pixel is related to the brightness value by:

$$L_v = K_\rho \; cos\theta_i \; + \; C_v \tag{3}$$

where, $K_\rho = \rho \; K_v$, ρ is the surface reflectance, K_v, C_v are constants of the visual imaging system and are determined via calibration. The calibration process simply consists of imaging two different surfaces at known orientations to solar radiation and of known reflectivities, whence the constants K_v and C_v are easily computed.

It is possible to obtain via stereoscopic image analysis, laser radar imagery, or from registered digital terrain data, the orientation of one elemental surface patch in the entire surface that is represented by a given image region. This orientation is best acquired for the elemental area that provides a reliable estimate, e.g., one that lies within a large planar patch. The region reflectivity ρ is then computed using equation (3). Knowing ρ, the value of $cos\theta_i$ is easily computed at each pixel in that region using equation (3). The surface is assumed to be opaque, hence, the absorptivity is computed as $\alpha_s = 1 - \rho$. The above procedure is applied to each region in the image to provide estimates of ρ and $cos\theta_i$ at each pixel in the entire image.

The assumption that viewed surfaces are opaque and are Lambertian is sometimes violated by the presence of transparent objects (e.g. glass windows, lakes), or regions of specular reflection (e.g. a polished surface). It is assumed that such regions in the imagery are detected by means other than that presented in this paper.

4   Estimating Surface Heat Fluxes

In this section, the various heat fluxes at the surface of the object are identified and the relationship between them is specified. A method for estimating these heat fluxes is then presented. This method uses values of surface temperature deduced from the thermal image, and surface reflectivity and relative orientation deduced from the visual image. Section 5 describes methods for interpreting imaged scenes using these heat flux estimates.

Consider an elemental area on the surface of the imaged object. Assuming one-dimensional heat flow, the heat exchange at the surface of the object is represented by figure 3. W_i is the incident solar radiation, θ_i is the angle between the direction of irradiation and the surface normal, the surface temperature is T_s, and W_{abs} is that portion of the irradiation that is absorbed by the surface. W_{cv} denotes the heat convected from the surface to the air which has temperature T_{amb} and velocity V, W_{rad} is the heat lost by the surface to the environment via radiation and W_{cd} denotes the heat conducted from the surface into the interior of the object. Irradiation at the object surface also includes that emanating from other scene components. As discussed in section 2, the magnitude of this absorbed irradiation is small when compared to total solar irradiation absorbed in visible and IR bands. The contribution of the former may therefore be ignored.

At any given instant, applying the law of conservation of energy to the heat fluxes flowing into the surface of the object and those flowing out from the surface. we have,

$$W_{abs} \; = \; W_{cd} \; + \; W_{cv} \; + \; W_{rad} \tag{4}$$

where, $W_{rad} \; = \; 0.9 \; \sigma \; (T_s^4 - T_{amb}^4)$,

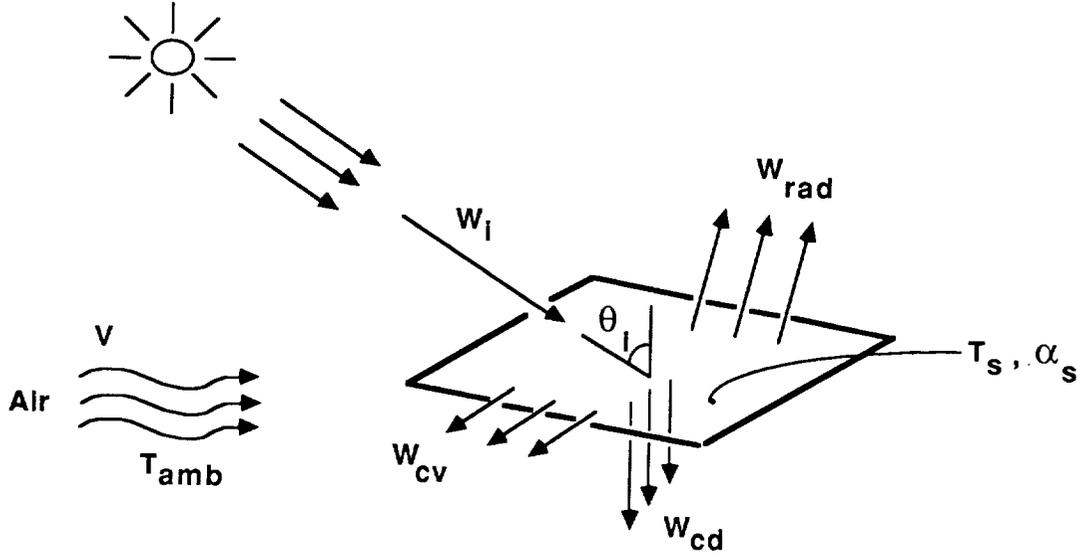$$W_{abs} \; = \; W_i \; cos\theta_i \; \alpha_s \; , \tag{5}$$

66

Figure 3. Exchange of heat fluxes at the surface of the imaged object.

σ denotes the Stefan-Boltzman constant, and α_s denotes the solar absorptivity of the surface. The convected heat transfer is given by

$$W_{cv} = h(T_s - T_{amb}) \tag{6}$$

where, h is the average convected heat transfer coefficient, and depends on the properties of the surrounding air (e.g. velocity, viscosity, temperature, etc.), and on the geometry and the nature of the object's surface. We note that W_{rad} is immediately available when T_{amb} is known since T_s is deduced from the thermal image as discussed in section 2.

In order to estimate the heat flux absorbed by the surface, it is first neccesary to determine the magnitude of the incident radiation on a horizontal surface and then compensate for the orientation and the reflectivity of the imaged surface. One approach is to directly measure the incident solar radiation using a pyrheliometer. Alternately, as was done in the experiments described later, an appropriate analytical model may be used to estimate this quantity. The variation (with day of the year and time of day) of the intensity of solar radiation incident on a horizontal surface on the ground has been modelled by Thepchatri, et al. [22] based on the data presented by Strock and Koral [23]. The empirical model accounts for diurnal and seasonal variations. This model is used for specifying W_i. Thus, knowing $cos\theta_i$ and α_s as described in section 3, W_{abs} may be computed using from equation (5).

The convective heat flux is obtained by using equation (6). The temperature for the object's surface is obtained from the thermal image as described in the previous section. The ambient temperature T_{amb} is known. The problem therefore lies in estimating the average convected heat transfer coefficient h. A plethora of empirical correlations have been established for computing h for various thermal and hydrodynamic conditions [24]. The simplifying assumption that the portion of the surface being viewed is flat allows the use of convecion correlations developed for external flow over flat plates [24]. The procedure for estimating the convected heat flux is as follows: knowing the wind velocity and the air temperature, the Reynolds number is computed, where the characteristic length of the object is

67
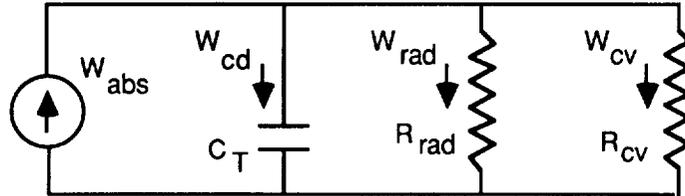
Figure 4. Equivalent thermal circuit of imaged surface.

assumed to be 1 meter. The value of the Reynolds number determines whether laminar or mixed flow conditions exist. Accordingly, the appropriate correlation is used. The Nusselt number is thus obtained and thence the convected heat transfer coefficient h. Equation (6) is now used to provide the estimate of convected heat flux.

Having estimated the convected heat flux, the radiated heat flux, and the absorbed irradiation as described above, the conduction heat flux is then deduced using equation (4).

5   Scene Interpretation

The estimated surface heat fluxes may be used to derive physically meaningful interpretations of the scene. Two different methods are discussed. The first approach assumes that only a single data set of the scene is available, and it consists of the thermal image, the visual image, and values of scene parameters obtained at a particular instant of time. The second method assumes that a sequence of data sets obtained at different time instants is available. The first approach is more suitable for scenes, the contents of which change frequently, e.g., one that contains automobiles. The second approach is suitable for scenes containing objects that are stationary over the sequence of data sets, e.g., scenes containing only vegetation, buildings and pavements.

5.1   Analysis of a Single Multisensory Data Set

The estimated surface heat fluxes may be used to evaluate how well an object can act as a heat source/sink. Thus a highly discriminatory feature may be extracted which is closely related to an intrinsic physical property of the imaged object. Considering a unit area on the surface of the imaged object, the equivalent thermal circuit for the surface is shown in figure 4. C_T is the lumped thermal

Object	Thermal Capacitance ($\times 10^{-6}$ Joules/Kelvin)
Asphalt Pavement	1.95
Concrete Wall	2.03
Brick Wall	1.51
Wood(Oak) Wall	1.91
Granite	2.25
Automobile	0.18

Table 1: Normalized values of lumped thermal capacitance.

capacitance of the object and is given by

$$C_T = DVc$$

where, D is the density of the object, V is the volume, and c is the specific heat. The resistances are given by:

$$R_{cv} = \frac{1}{h} \quad \text{and} \quad R_{rad} = \frac{1}{0.9\sigma(T_s^2 + T_{amb}^2)(T_s + T_{amb})}$$

From figure 4 it is clear that the conduction heat flux W_{cd} estimated in the previous section depends on the lumped thermal capacitance C_T of the object. A relatively high value for C_T implies that the object is able to sink or source relatively large amounts of heat. An estimate of W_{cd}, therefore, provides us with a relative estimate of the thermal capacitance of the object, albeit a very approximate one. Table 1 lists values of C_T of typical objects imaged in outdoor scenes. The values have been normalized for unit volume of the object.

Note that the thermal capacitance for walls and pavements is significantly greater than that for automobiles and hence W_{cd} may be expected to be higher for the former regions. Plants absorb a significant percentage of the incident solar radiation. The energy absorbed is used for photosynthesis and also for transpiration. Only a small amount of the absorbed radiation is convected into the air. Therefore, the estimate of the W_{cd} will be almost as large (typically 95%) as that of the absorbed heat flux. Thus, W_{cd} is useful in estimating the object's ability to sink/source heat radiation, a feature shown to be useful in discriminating between different classes of objects. However, in order to minimize the feature's dependence on differences in absorbed heat flux, a normalized feature was defined to be the ratio

$$R = W_{cd}/W_{abs}$$

Although the heat flux ratio, $R = W_{cd}/W_{abs}$, does capture a great deal of information about the imaged object, it is not discriminatory enough to unambiguously delineate the identity of the imaged object. Other sources of information are therefore warranted. Hence, information such as the surface reflectivity, ρ, of the region which is derived from the visual image, and average region temperature which is derived from the thermal image are also used to facilitate region labeling. Section 6 presents experimental results of using this approach on real multisensory data.

5.2 Analysing Temporal Sequence of Multisensor Data

If a temporal sequence of multisensor data consisting of thermal imagery, visual imagery and scene conditions is available, then it is possible to extract a more reliable estimate of the imaged object's

69

relative ability to sink/source heat radiation. Observe that the relationship between the conducted heat flux W_{cd} and the thermal capacitance C_T of the object is given by:

$$W_{cd} = C_T \frac{dT_s}{dt}$$

A finite (backward) difference approximation to this equation may be used for estimating C_T as

$$C_T = W_{cd} \frac{(t_2 - t_1)}{(T_s(t_2) - T_s(t_1))} \tag{7}$$

where, t_1 and t_2 are the time instants at which the data were acquired, $T_s(t_1)$ and $T_s(t_2)$ are the corresponding surface temperatures, and W_{cd} is the conducted heat flux which is assumed to be constant during the time interval. However, W_{cd} does vary and an average value of $(W_{cd}(t_1) + W_{cd}(t_2))/2$ is used in equation (7).

Section 6 presents experimental results obtained by applying the above temporal analysis method to multisensory data.

6 Experimental Results

The methods described in the previous sections were applied to real multisensory data acquired from outdoor scenes. Calibrated remote sensing data were not available along with values of ambient scene parameters such as wind speed and temperature. Hence, thermal and visual imagery were acquired from a ground based imaging setup consisting of an Inframetrics infrared imaging system and a video imaging system. An anemometer was used to measure wind speed and a digital thermometer was used to calibrate the thermal imaging system so as to allow absolute temperature estimates as discussed in section 2. The two methods discussed in the previous section were applied to several sets of data which were acquired at different times of the day and during different seasons of the year.

The results obtained using one data set are presented in figures 5 through 8. Figure 5 shows the visual image of a scene containing an automobile, buildings, asphalt pavement and vegetation. Figure 6 shows the thermal image of the same scene. The techniques described in the preceeding sections of this paper were used to estimate the surface heat fluxes, whence the ratio $R = W_{cd}/W_{abs}$ was computed at each pixel. A histogram of these values was computed for each region and the mode of the histogram was found. This value was chosen as the representative value of R for the region. Figure 7 shows the values obtained for each region. As predicted by the discussion in section 5, automobiles produce the lowest value of this feature, pavements and buildings produce intermediate values and vegetation produces the highest values.

In addition to the feature R, the average region temperature, and the surface reflectivity were also used in a decision tree classifier to label regions as building, pavement, vegetation or automobile. The classifier used heuristic rules of the form:
IF { $R \in [0.2, 0.9]$ AND $\rho \in [0.35, 1.0]$ } OR { $R \in [-.8, -.3]$ } THEN label = bldng
The resultant classification is shown in figure 8.

The method of temporal analysis of the scenes was tested on data acquired at intervals of three hours. Table 2 presents the mean and standard deviation of the value of C_T estimated for different classes of scene objects. The estimated values compare very favorably with those listed in table 1. Except for the concrete and brick walls, the estimated values for each class are of the same order of
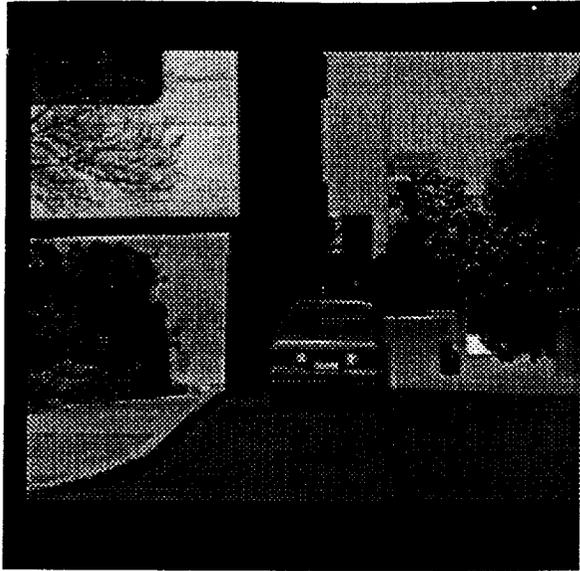
ORIGINAL PAGE
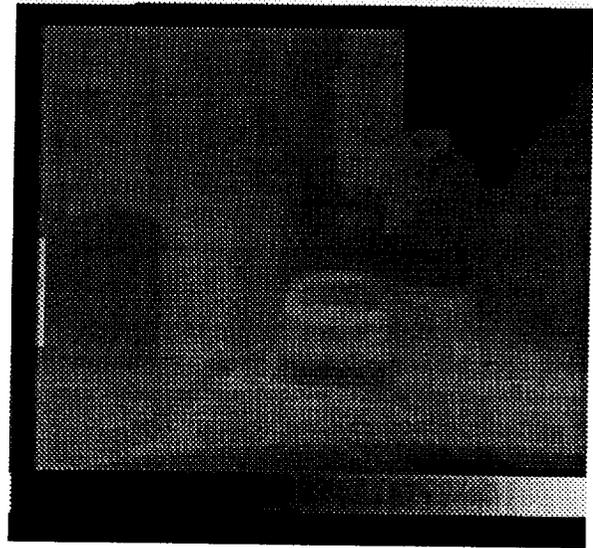BLACK AND WHITE PHOTOGRAPH

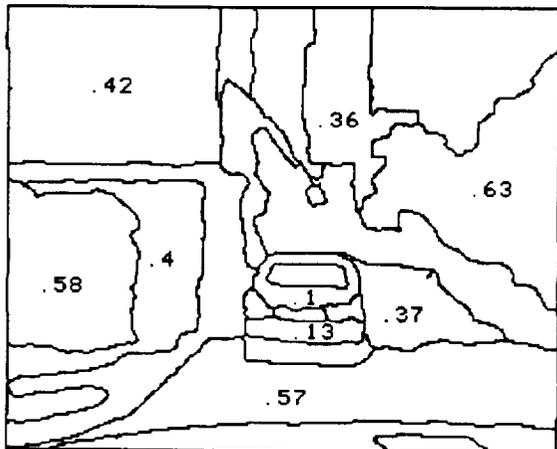

Figure 5. Visual image.

Figure 6. Thermal image.
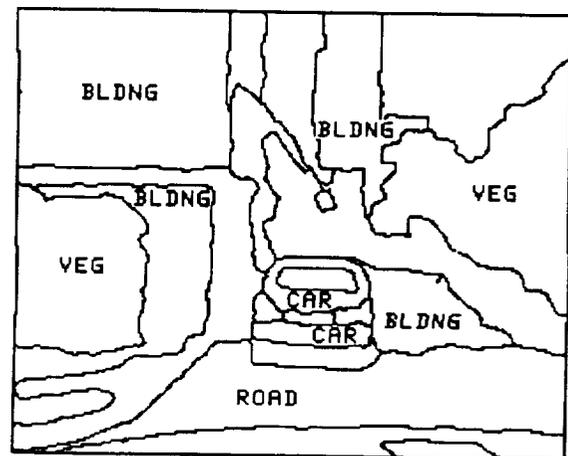

Figure 7. Mode of feature R.

Figure 8. Region classification.

71

Object	Average C_T ($\times 10^{-6}$ J/K)	Std. Devn. ($\times 10^{-6}$ J/K)
Automobile	0.08	0.08
Concrete Wall	0.22	0.37
Brick Wall	0.37	0.38
Asphalt Pavement	1.05	0.46
Vegetation	1.5	2.7

Table 2: Values of lumped thermal capacitance estimated using the method described in section 5.2.

magnitude as listed in table 1, and are ordered in a similar manner. The walls do not compare favorably possibly due to the wide variation in wall thickness that is difficult to account for and also due to the unknown thermal conditions on the interior surface of the walls. In general, a significant offset may be expected in the estimated values due to the many approximations used in the computation of the heat flux estimates. Inspite of this limitation, it is obvious that the approach described above makes available a very useful and meaningful method for the interpretation of multisensory data. Note also that the value of C_T is a deterministic value which is completely defined by a physical definition for a particular class of objects. Hence, a deterministic measure of the technique's performance is available by comparing the estimated and true values of C_T. Such a measure is not available in purely statistical interpretation techniques. This is one of the major advantages of a phenomenological approach to scene interpretation when compared to the purely statistical approach.

7 Conclusions

A new method has been described for interpreting scenes using multisensory data. The phenomenological approach combines information from the different imaging modalities to derive meaningful features. The approach is based on physical models of the energy exchange between the imaged surface and the environment. The thermal and visual images yield estimates of surface heat fluxes which in turn provide a measure of the relative ability of the imaged surface to source/sink heat energy. The interpretation thus relies on a rough estimate of the lumped thermal capacitance of the object which has been shown to vary widely for different classes of objects in outdoor scenes. The developed approach was tested on real multisensory data. Due to the unavailability of a calibrated remote sensing data and accompanying values of ambient scene conditions, the testing was performed using terrain based imaging equipment. The approach described above may be easily applied to multisensory imagery acquired by airborne or satellite-based sensors.

References

[1] B.G. Lee, R.T. Chin and D.W. Martin, "Automated Rain-Rate Classification of Satellite Images Using Statistical Pattern Recognition", *IEEE Trans. Geoscience and Remote Sensing*, Vol. GE-23, No. 3, May 1985, pp. 315-324.

[2] W.D. Rosenthal, B.J. Blanchard and A.J. Blanchard, "Visible/Infrared/Microwave Agriculture Classification, Biomass and Plant Height Algorithms", *IEEE Trans. Geoscience and Remote Sensing*, Vol. GE-23, No. 2, March 1985, pp. 84-90

[3] S. Di Zenzo, R. Bernstein, S.D. Degloria and H.G. Kolsky, "Gaussian Maximum Likelihood and Contextual Classification for Multicrop Classification", *IEEE Trans. Geoscience and Remote Sensing*, Vol. GE-25, No.

6, Nov. 1987, pp. 805-814

[4] S.W. Wharton, "A Spectral Knowledge-Based Approach for Urban Land-Cover Discrimination", *IEEE Trans. Geoscience and Remote Sensing*, Vol. GE-25, No. 3, May 1987, pp. 272-282

[5] T. Lee, J.A. Richards, and P.H. Swain, "Probabilistic and Evidential Approaches for Multisource Data Analysis", *IEEE Trans. Geoscience and Remote Sensing*, Vol. GE-25, No. 3, May 1987, pp. 283-293

[6] D.G. Goodenough, M. Goldberg, G. Plunkett and J. Zelek, "An Expert System for Remote Sensing", *IEEE Trans. Geoscience and Remote Sensing*, Vol. GE-25, No. 3, May 1987, pp. 349-359

[7] H. Asar, N. Nandhakumar and J.K. Aggarwal, "Pyramid-Based Image Segmentation Using Multisensory Data", to appear in *Pattern Recognition*.

[8] V.S. Whitehead, W.R. Johnson and J.A. Boatright, "Vegetation Assessment Using a Combination of Visible, Near-IR and Thermal-IR AVHRR Data", *IEEE Trans. Geoscience and Remote Sensing*, Vol. GE-24, No. 1, Jan. 1986, pp. 107-112.

[9] P.R. Christensen, "Martian Dust Mantling and Surface Compositions: Interpretations of Thermophysical Properties", *Journal of Geophysical Research*, Vol. 87, No. B12, 1982, pp. 9985-9998.

[10] B.M Jakosky and P.R. Christensen, "Global Duricrust on Mars: Analysis of Remote Sensing Data", *Journal of Geophysical Research*, Vol. 91, No. B3, 1986, pp. 3547-3559.

[11] N. Nandhakumar and J.K. Aggarwal, "Integrated Analysis of Thermal and Visual Images for Scene Interpretation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 4, July 1988, pp. 469-481.

[12] R. Siegel and J.R. Howell, *Thermal Radiation Heat Transfer*, 2nd Ed., Mc-Graw Hill Book Co., New York, 1981.

[13] C. Ohman, "Practical Methods for Improving Thermal Measurements", *Proc. of SPIE*, Vol. 313, 1981, pp 204 - 212.

[14] J.R. Schott and J.D. Biegel, "Comparison of Modelled and Empirical Atmospheric Propagation Data", *Proc. of SPIE*, Vol. 430, 1983, pp 45 - 52.

[15] J.M. Jarem, J.H. Pierluissi and W.W. Ng, "A Transmittance Model for Atmospheric Methane", *Proc. of SPIE*, Vol. 510, 1984, pp 94 - 99.

[16] Horn B.K.P., and Sjoberg R.W., "Calculating the Reflectance Map", *Applied Optics*, Vol. 18, No. 1, 1979, pp 1770 - 1779.

[17] Horn B.K.P., "Hill Shading and the Reflectance Map", *Proc. of the IEEE*, Vol. 19, No. 1, pp 14 - 47.

[18] Woodham R.J., "Analysing Images of Curved Surfaces", *Artificial Intelligence*, Vol. 17, 1981, pp 117 - 140.

[19] Ikeuchi K., and Horn B.K.P., "Numerical Shape from Shading and Occluding Boundaries", *Artificial Intelligence*, Vol. 17, 1981, pp 141 - 184.

[20] R.T. Frankot and R. Chellappa, "A Method for Enforcing Integrability in Shape from Shading Algorithms", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 4, July 1988, pp. 439-451.

[21] J.M. Norman, J.M. Welles and E.A. Walter, "Contrasts Among Bidirectional Reflectance of Leaves, Canopies and Soils", *IEEE Trans. Geoscience and Remote Sensing*, Vol. GE-23, No. 5, September 1985, pp. 659-667.

[22] Thepchatri T., Johnson C.P., and Matlock H., "Prediction of Temperature and Stresses in Highway Bridges by A Numerical Procedure Using Daily Weather Reports", Tech. Report 23-1, 1977, Center for Highway Research, University of Texas at Austin.

[23] Strock C., and Koral R. L., *Handbook of Air Conditioning, Heating and Ventilating*, Industrial Press, Inc., 2nd Ed., 1965.

[24] Incropera F.P., and De Witt D.P., *Fundamentals of Heat Transfer*, John Wiley & Sons, Inc., New York, 1981.

73

A Method for Classification of Multisource Data using Interval-Valued Probabilities and Its Application to HIRIS Data

H. Kim and P.H. Swain

School of Electrical Engineering
and
Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, IN 47907, U.S.A.
Tel: (317)494-0212, FAX: (317)494-0776
E-mail address: swain@ecn.purdue.edu

ABSTRACT

A method of classifying multisource data in remote sensing is presented. The proposed method considers each data source as an information source providing a body of evidence, represents statistical evidence by interval-valued probabilities, and uses Dempster's rule to integrate information based on multiple data sources.

The method is applied to the problems of ground-cover classification of multispectral data combined with digital terrain data such as elevation, slope, and aspect. Then this method is applied to simulated 201-band High Resolution Imaging Spectrometer (HIRIS) data by dividing the dimensionally huge data source into smaller and more manageable pieces based on the global statistical correlation information. It produces higher classification accuracy than the Maximum Likelihood (ML) classification method when the Hughes phenomenon is apparent.

1 INTRODUCTION

The importance of utilizing multisource data in ground-cover classification lies in the fact that it is generally correct to assume that improvements in terms of classification accuracy can be achieved at the expense of additional independent features provided by separate sensors. However, it should be recognized that information and knowledge from most available data sources in the real world are neither certain nor complete. We refer to such a body of uncertain, incomplete, and sometimes inconsistent information as "evidential information."

The objective of the current research is to develop a mathematical framework within which various applications can be made with multisource data in remote sensing and geographic information systems. The methodology described in this paper has evolved from "evidential reasoning," where each data source is considered as providing a body of evidence with a certain degree of belief. The degrees of belief based on the body of evidence are represented by "interval-valued (IV) probabilities" rather than by conventional point-valued probabilities so that uncertainty can be embedded in the measures.

There are three fundamental problems in the multisource data analysis based on IV probabilities: (1) how to represent bodies of evidence by IV probabilities, (2) how to combine IV probabilities to give an overall assessment of the combined body of evidence, and (3) how to make decisions based on IV probabilities.

The paper describes a formal method of representing statistical evidence by IV probabilities based on the Likelihood Principle. In order to integrate information obtained from individual data sources, the method presented in the paper uses Dempster's rule for combining multiple bodies of evidence. Although IV probabilities together with Dempster's rule provide an innovative means for the representation and combination of evidential information, they make the decision process rather complicated. We need more intelligent strategies for making decisions. This paper also focuses on the development of decision rules over IV probabilities.

2 AXIOMATIC DEFINITION OF IV PROBABILITY

Interval-valued probabilities can be thought as a generalization of ordinary point-valued probabilities. The endpoints of IV probabilities are called the "upper probability" and the "lower probability."

There have been various works introducing the concepts of IV probabilities in the areas of philosophy of science and statistics [1][2][3][4]. Although the mathematical rationales behind those approaches are different, there are some properties of IV probabilities which are commonly required. The axiomatic approach to IV probabilities is based on those common properties, so that it can avoid conceptual ambiguities.

DEFINITION [5] Suppose Θ is a finite set of exhaustive and mutually exclusive events. Let β denote a Boolean algebra of the subsets of Θ. The IV probability $[L, U]$ is defined by the set-theoretic functions:

$$L : \beta \to [0, 1] \qquad (2.1)$$

$$U : \beta \to [0, 1] \qquad (2.2)$$

satisfying the following properties:

I) $\quad U(A) \geq L(A) \geq 0 \quad$ for any $A \in \beta$ $\qquad (2.3)$

II) $\quad U(\Theta) = L(\Theta) = 1$ $\qquad (2.4)$

III) $\quad U(A) + L(\overline{A}) = 1$ for any $A \in \beta$ $\qquad (2.5)$

IV) $\quad$ For any $A, B \in \beta$ and $A \cap B = \varnothing$,

$$L(A) + L(B) \leq L(A \cup B) \leq L(A) + U(B)$$

$$\leq U(A \cup B) \leq U(A) + U(B) \qquad (2.6)$$

Given a system of IV probabilities over β, the actual probability measure, $P(A)$, of any subset A of Θ is assumed to lie in the interval $[L, U]$ such that

$$L(A) \leq P(A) \leq U(A) \qquad (2.7)$$

The degree of uncertainty about the actual probability of A is represented by the width, $U(A) - L(A)$, of the interval. In particular, $U(A) = L(A) = P(A)$ when there is complete knowledge of the probability of A. In this case, the IV probability becomes an ordinary additive probability. And $L(A) + L(\overline{A}) = 0$

when there is absolutely no knowledge of the probability of A.

The basic probability assignment m defined in Shafer's mathematical theory of evidence[6] has the following relations with the IV probabilities:

$$L(A) = \sum_{B \subseteq A} m(B) \qquad (2.8)$$

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} L(B) \quad \text{for all } A \subset \Theta \qquad (2.9)$$

$$U(A) = \sum_{B \cap A \neq \varnothing} m(B) \qquad (2.10)$$

3 REPRESENTATION OF STATISTICAL EVIDENCE BY IV PROBABILITY

When a body of evidence is based on the outcomes of statistical experiments known to be governed by any probability model, it is called "statistical evidence." One of the basic problems for any theory of IV probabilities is how to represent a given body of statistical evidence by IV probabilities.

DEFINITION [6] An upper probability function U is said to be "consonant" if its focal elements are nested, i.e., if for $A_i \subseteq \Theta$ $(i = 1, \ldots, r)$ such that $m(A_i) > 0$ for all i and $\sum_{i=1}^{r} m(A_i) = 1$, $A_i \subset A_j$ for any $i < j$, where m is the basic probability assignment of U.

Suppose the observed data in a statistical experiment are governed by a probability model $\{p_\theta : \theta \in \Theta\}$, where p_θ is a conditional probability density function on a sample space X given θ. Our intuitive feeling is that an observation $x \in X$ seems to more likely belong to those elements of Θ which assign the greater chance to x.

Based on the above intuition along with the consonance assumption of the upper probability function, Shafer[6] proposed the linear plausibility function defined as:

$$\mathcal{U}(A|x) = \frac{\max\limits_{\theta \in A} p_\theta(x)}{\max\limits_{\theta \in \Theta} p_\theta(x)} \quad \text{for all } A \in \beta \text{ and } A \neq \emptyset \quad (3.1)$$

The corresponding lower probability function is given as:

$$\mathcal{L}(A|x) = 1 - \frac{\max\limits_{\theta \in \overline{A}} p_\theta(x)}{\max\limits_{\theta \in \Theta} p_\theta(x)} \quad \text{for all } A \in \beta \quad (3.2)$$

In particular, when the set A is singleton, say $\{\theta'\}$, the function in Eq.(3.1) gives the relative likelihood of θ' to the most likely element in Θ.

4 DEMPSTER'S RULE FOR COMBINING IV PROBABILITIES

Dempster's rule is a generalized scheme of Bayesian inference to aggregate bodies of evidence provided by multiple information sources. Let m_1 and m_2 be the basic probability assignments associated respec-tively with the belief functions $\mathcal{B}el_1$ and $\mathcal{B}el_2$ which are inferred from two entirely distinct bodies of evidence E_1 and E_2. For all A_i, B_j, and $X_k \subset \Theta$, Dempster's rule (or Dempster's orthogonal sum) gives a new belief function denoted by

$$\mathcal{B}el = \mathcal{B}el_1 \oplus \mathcal{B}el_2 \quad (4.1)$$

The basic probability assignment associated with the new belief function is defined as:

$$m(X_k) = (1-k)^{-1} \sum_{A_i \cap B_j = X_k} m_1(A_i) \cdot m_2(B_j)$$

for any $X_k \neq \emptyset \quad (4.2)$

where k is the measure of conflict between $\mathcal{B}el_1$ and $\mathcal{B}el_2$ defined as:

$$k = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) \cdot m_2(B_j) \quad (4.3)$$

Dempster's rule computes the basic probability of X_k, $m(X_k)$, from the product of $m_1(A_i)$ and $m_2(B_j)$ by considering all A_i and B_j whose intersection is X_k. Once m is computed for every $X_k \subset \Theta$, the belief func-

tion is obtained by the sum of m's committed to X_k and its subsets. The denominator $(1-k)$ normalizes the result to compensate for the measure committed to the empty set so that the total probability mass has measure one. Consequently, Dempster's rule discards the conflict between E_1 and E_2 and carries their consensus to the new belief function.

Dempster's rule is both commutative and associative. Therefore, the order or grouping of evidence in combination does not affect the result, and a sequence of information sources can be combined either sequentially or pairwise.

5 DECISION RULES FOR IV PROBABILITIES [7]

Consider a classification problem where an arbitrary pattern $x \in X$ from an unknown class is assigned to one of n classes in Θ. Let $\lambda(\theta_i|\theta_j)$ be a measure of the "loss" incurred when the decision θ_i is made and the true pattern class is in fact θ_j, where i, j = 1, ..., n. Also, let $\hat{\theta}(x)$ denote a decision rule that tells which class to choose for every pattern x. We define the "upper expected loss" and the "lower expected loss" of making a decision $\hat{\theta}(x) = \theta_i$ as:

$$\ell_i^*(x) = \sum_{j=1}^{n} \lambda(\theta_i|\theta_j) \, \mathcal{U}_x(\theta_j) \quad (5.1)$$

$$\ell_{*i}(x) = \sum_{j=1}^{n} \lambda(\theta_i|\theta_j) \, \mathcal{L}_x(\theta_j) \quad (5.2)$$

where $\mathcal{U}_x$ and $\mathcal{L}_x$ are respectively the upper and the lower probabilities for x being actually from θ_j.

The "Bayes-like rule" is the one which minimizes both the upper and the lower expected losses, i.e.,

$$\hat{\theta}(x) = \theta_i \text{ if } \ell_i^*(x) \leq \ell_j^*(x) \text{ and } \ell_{*i}(x) \leq \ell_{*j}(x)$$

for j=1,..., n $\quad (5.3)$

A problem with the above decision rule is that there does not always exist θ which satisfies the condition in Eq.(5.3), which can lead to ambiguity. In such an ambiguous

situation, one may withhold the decision and wait for a new piece of information. Otherwise, the ambiguity may be resolved by resorting to the following rule, so-called "minimum average expected loss rule":

$$\hat{\theta}(x)=\theta_i \text{ if } \frac{\Gamma^*_i(x)+\ell_{*i}(x)}{2} \le \frac{\Gamma^*_j(x)+\ell_{*j}(x)}{2}$$

for j=1,..., n   (5.4)

As an alternative to the Bayes-like rule, there are two other rules by which a decision is made according to individual measures of the interval, that is, either the upper expected loss or the lower expected loss:

(A) minimum upper expected loss rule:

$$\hat{\theta}(x)=\theta_i \text{ if } \Gamma^*_i(x) \le \Gamma^*_j(x) \quad \text{for } j=1,..., n \quad (5.5)$$

(B) minimum lower expected loss rule:

$$\hat{\theta}(x)=\theta_i \text{ if } \ell_{*i}(x) \le \ell_{*j}(x) \quad \text{for } j=1,..., n \quad (5.6)$$

Although the above two rules always produce decisions and there is no ambiguous situation in making a decision according to the rules, they do not utilize all of the information represented by the IV probabilities. The performance of these rules are compared with the minimum average expected loss rule in the experiments by applying them to problems of ground-cover classification based on remotely sensed and geographic data.

6 EXPERIMENTAL RESULTS

The experiments have been performed over two different image data sets. In the experiments, the classification accuracies of the multisource data (MSD) classification based on the proposed method were compared with those of Maximum Likelihood (ML) classifications based on the stacked vector approach.

Table 1 describes the set of data sources for the first experiment. The image in this data set consists of 256 lines by 256 columns and covers a forestry site around the Anderson River area of British Columbia, Canada. Source 1 is 11-band Airborne Multispectral Scanner (A/B MSS) data. Sources 2 and 3 are Synthetic Aperture

Radar (SAR) imagery in Shallow mode and Steep mode, respectively. Sources 4 through 6 provide digital terrain data.

In this experiment, 6 classes were defined as listed in Table 2, and 100 pixels per class were used for training data, which is between 4% and 8% of the total pixels of the classes in the test fields. The training samples are uniformly distributed over the test

Table 1. Anderson River Data Set.

Source Index	Data Type	Spectral Region	Input Channel	Spectral Band(μm)
1	A/B MSS	Visible	1	.38 - .42
			2	.42 - .45
			3	.45 - .50
			4	.50 - .55
			5	.55 - .60
			6	.60 - .65
			7	.65 - .69
		Near IR	8	.70 - .79
			9	.80 - .89
			10	.92 - 1.10
		Thermal	11	8 - 14
2	SAR	Shallow		XHV XHH LHV LHH
3	SAR	Steep		XHV XHH LHV LHH
4	Topo-	Elevation		
5	graphic	Aspect		
6		Slope		

Table 2. Information Classes for Test of Anderson River Data Set.

Class Index	Cover Types	Tree Sizes	No. of Pixels	% of Total
1	Douglas Fir 2 (df2)	31 - 40m	2246	21.72
2	Douglas Fir 3 (df3)	21 - 30m	1501	14.52
3	DF+Other Species 2 (df+os2)	31 - 40m	1352	13.08
4	DF+Lodgepole Pine 2 (df+lp2)	21 - 30m	1589	15.37
5	Hemlock+Cedar (hc)	31 - 40m	1587	15.35
6	Forest Clearings (fc)		2064	19.96
Total			10339	100.0

78

fields so that they may be considered as good representatives of the total samples.

We have observed that some of the classes defined in Table 2 cannot be assumed to be normally distributed in the topographic data. Thus it was decided to adopt a nonparametric approach such as the "Nearest Neighbor" (NN) method [8] in computing probability measures while the optical and radar data sources were assumed to have Gaussian probability density functions.

First, the ML classification based on the stacked vector approach was carried out for various sets of the data sources, adding one source at a time to the A/B MSS data in the order Elevation, SAR-Shallow, SAR-Steep, Aspect, and Slope. Then the MSD classification based on the proposed method was performed using different decision rules. Tables 3 and 4 compare the results for the training samples and the test samples, respectively. Even though the compounded data in the ML classification were treated as having Gaussian distributions, the ML and the MSD methods produced similar results for the training samples. This is not surprising because the ML method uses conventional additive probabilities assuming that the knowledge concerning the actual unknown probabilities is complete, which is reasonable as far as the training samples are concerned.

Table 3. Results of Classifications over Training Samples of Anderson River Data.

	Decision Rule	Sources					
		1	1, 4	1, 2, 4	1 - 4	1 - 5	1 - 6
ML		82.50	88.67	91.67	92.00	92.83	93.50
MSD	MUEL	-	89.83	92.00	92.50	93.17	94.33
	MLEL	-	88.67	91.17	91.33	92.33	93.67
	MAEL	-	88.50	91.00	91.67	91.67	93.50

Comparing the performance of the three decision rules, the minimum upper expected loss (MUEL) rule was superior to the other rules, the minimum lower expected loss (MLEL) rule and the minimum average expected loss (MAEL) rule. It is not known in

general which rule is the best. Further research is needed to determine whether guidelines can be devised for selection of the decision rule.

Table 4. Results of Classifications over Test Samples of Anderson River Data.

	Decision Rule	Sources					
		1	1, 4	1, 2, 4	1 - 4	1 - 5	1 - 6
ML		74.16	77.77	79.13	78.93	79.80	81.01
MSD	MUEL	-	80.60	82.39	82.69	83.02	84.54
	MLEL	-	78.45	81.42	81.67	82.24	83.65
	MAEL	-	78.21	80.95	82.05	81.88	83.16

In the second experiment, the proposed method was applied to the classification of HIRIS data by decomposing the data into smaller pieces, i.e., subsets of spectral bands. The data set used in this experiment is simulated HIRIS data obtained by RSSIM [9]. RSSIM is a simulation tool for the study of multispectral remotely sensed images and associated system parameters. It creates realistic multispectral images based on detailed models of the ground surface, the atmosphere, and the sensor. Table 5 provides a description of the simulated HIRIS data set.

Figure 1 is a visual representation of the global statistical correlation coefficient matrix of the data. The image is produced by converting the absolute values of coefficients to gray values between 0 and 255. Based on the correlation image, the 201 bands were divided into 3 groups in such a way that intra-correlation is maximized and inter-correlation is minimized. Table 6 describes the multisource data set after division. Note that the spectral regions of the input channels in Source 3 coincide with the water absorption bands.

With 225 training samples (a third of the total samples) for each class, the ML classification and the MSD classification using the minimum upper expected loss rule were performed over the total samples for various sets of the sources, and the results are listed in Table 7.

The results of the ML method apparently show effects of the Hughes phenomenon;

the accuracy goes down as the dimensionality of the source increases while the number of training samples is fixed. In particular, the accuracy decreases by a considerable amount when all features are used. Presence of the Hughes phenomenon causes the ML method to be particularly sensitive to a bad source, Source 3 in this case. Meanwhile, the proposed MSD classification method always shows robust performance and gives consistent results.

Table 5. Description of Simulated HIRIS Data Set.

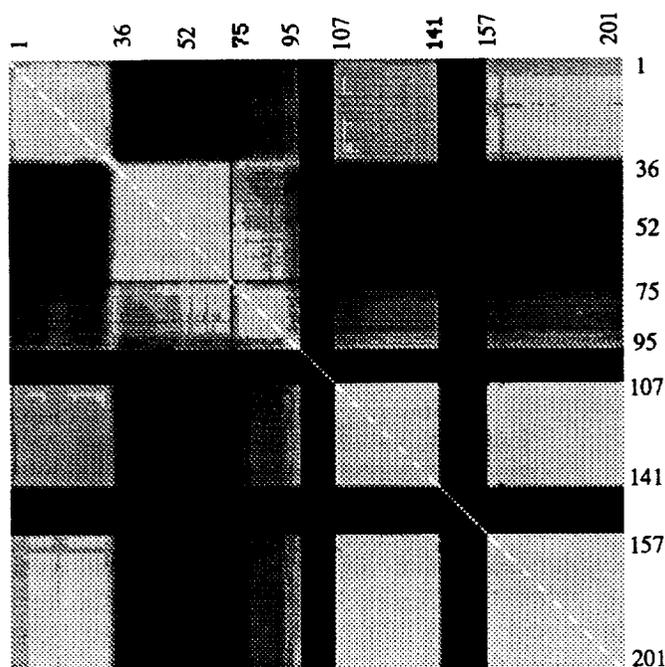| Name	Finney County Data Set
Data Type	201-band HIRIS data simulated by RSSIM
Spectral Region	0.4 - 2.4mm
Spectral Resolution	0.01mm
Image Size	45 lines × 45 columns (2025 samples)
Information Classes	Winter Wheat, Summer Fallow, Unknown

Figure 1. Global Statistical Correlation Coefficient Image of Simulated HIRIS Data.

Table 6. Divided Sources of HIRIS Data Set.

Source Index	Input Channels	No. of Features
Source 1	1- 35, 107 - 141, 157 - 201	115
Source 2	36 - 95	60
Source 3	96 - 106 (1.35 - 1.45μm) 142 - 156 (1.81 - 1.95μm)	26

Table 7. Results of Classifications over Test Samples of Simulated HIRIS Data Set.

	Sources				
	S1	S2	S3	S1, S2	All
ML	75.75	75.60	45.83	74.56	65.14
MSD	-	-	-	77.83	77.63

6 CONCLUSIONS

In this paper we have investigated how interval-valued probabilities can be used to represent and aggregate evidential information obtained from various data sources. Overall concepts of interval-valued probabilities have been employed to develop a new method of classifying multisource data in remote sensing and geographic information systems. The experiments demonstrate the ability of our method to capture uncertain information based on inexact and incomplete multiple bodies of evidence. The basic strategy of this method is to decompose the relatively large size of evidence into smaller, more manageable pieces, to assess plausibilities and supports based on each piece, and to combine the assessments by Dempster's rule. In this scheme, we are able to overcome the difficulty of precisely estimating statistical parameters, and to integrate statistical information as much as possible.

for Remote Sensing, Department of Energy, Mines and Resources, of the Government of Canada.

REFERENCES

[1] B.O. Koopman, "The axioms and algebra of intuitive probability", *The Annals of Mathematics*, vol.41, pp269-292, 1940.

[2] I.J. Good, "Subjective probability as the measure of a non-measurable set", in E.Nagel, Suppes and Tarski (eds.), *Logic, Methodology and the Philosophy of Science*, pp319-329, Stanford University Press, 1962.

[3] C.A.B. Smith, "Consistency in statistical inference and decision", *J. Roy Stat. Soc.*, *Ser.B*, vol.23, pp1-25, 1961.

[4] A.P. Dempster, "Upper and lower probabilities induced by a multivalued mapping", *Ann. Math. Stat.*, vol.38, pp325-339, 1967.

[5] P. Suppes, "The measurement of be-lief", *J. Roy. Stat. Soc.*, *Ser.B*, vol.36, pp160-191, 1974.

[6] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[7] H. Kim, "A Method of Classification for Multisource Data in Remote Sensing based on Interval-Valued Probabilities", Ph.D. Thesis, School of Electrical Engineering, Purdue University, West Lafayette, IN 47907 U.S.A., August 1990.

[8] Fukunaga, K., *Intro. to Statistical Pattern Recognition*, Academic Press, 1972.

[9] Kerekes, J.P., and D.A. Landgrebe, "Modeling, Simulation, and Analysis of Optical Remote Sensing Systems", TR-EE 89-49, School of Electrical Engineering, Purdue University, West Lafayette, IN U.S.A., 1989.

Improved Disparity Map Analysis Through the Fusion of Monocular Image Segmentations

Frederic P. Perlant and David M. McKeown

Digital Mapping Laboratory
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213[1]

Abstract

It is generally believed that the detailed analysis of remotely sensed imagery requires the extraction of a variety of partial image domain cues coupled with the use of *a priori* or contextual information. In some cases there are fundamental limits to the variety and type of information that may be extracted from a single image or stereo pair. However, in most cases a sufficient variety of cues can be extracted; the major issue is in how to utilize disparate scene cues to achieve a more complete and accurate overall scene interpretation.

The focus of this paper is to examine how estimates of three-dimensional scene structure, as encoded in a scene disparity map, can be improved by the analysis of the original monocular imagery. This paper describes the utilization of surface illumination information provided by the segmentation of the monocular image into fine surface patches of nearly homogeneous intensity to remove mismatches generated during stereo matching. These patches are used to guide a statistical analysis of the disparity map based on the assumption that such patches correspond closely with physical surfaces in the scene. Such a technique is quite independent of whether the initial disparity map was generated by automated area-based or feature-based stereo matching.

We present stereo analysis results on a complex urban scene containing various man-made and natural features. This scene contains a variety of problems including low building height with respect to the stereo baseline, buildings and roads in complex terrain, and highly textured buildings and terrain. We demonstrate the improvements due to monocular fusion with a set of different region-based image segmentations. Finally, we discuss the generality of this approach to stereo analysis and its utility in the development of general three-dimensional scene interpretation systems.

1. Introduction

One common problem for systems that interpret multiple sources of sensed data is the fusion of partial results from a variety of sources. This problem appears under many guises. For example, given a set of different scene descriptions generated from a single image using a variety of image analysis techniques, how does one intelligently combine such partial information? [8]. The introduction of additional sensor types, temporal imagery, and multiple-look imagery create dimensions along which information fusion must be performed; as such, the complexity of the problem can increase. In some cases, increased amounts of data provide improved information. This may not necessarily follow, however; complex systems having different sources of error may not reinforce correct partial interpretations nor refute incorrect ones.

Thus, the key issue is the integration of many different sources of partial information. In computer vision (and in particular, three-dimensional scene analysis), the goal is to generate an *interpretation* of the scene that is as close as possible to the actual scene imaged. Such an interpretation can include the delineations and heights of buildings, a digital elevation model, and the centerline and width of roads in a transportation network. Our belief is that no individual computer vision technique can reliably provide a complete scene reconstruction. To achieve good performance, we need to gather a variety of information, extracted by various processes from the imagery, and synthesize this disparate information into a consistent model. Figure 1-1 shows a possible structure for such a scene interpretation system.

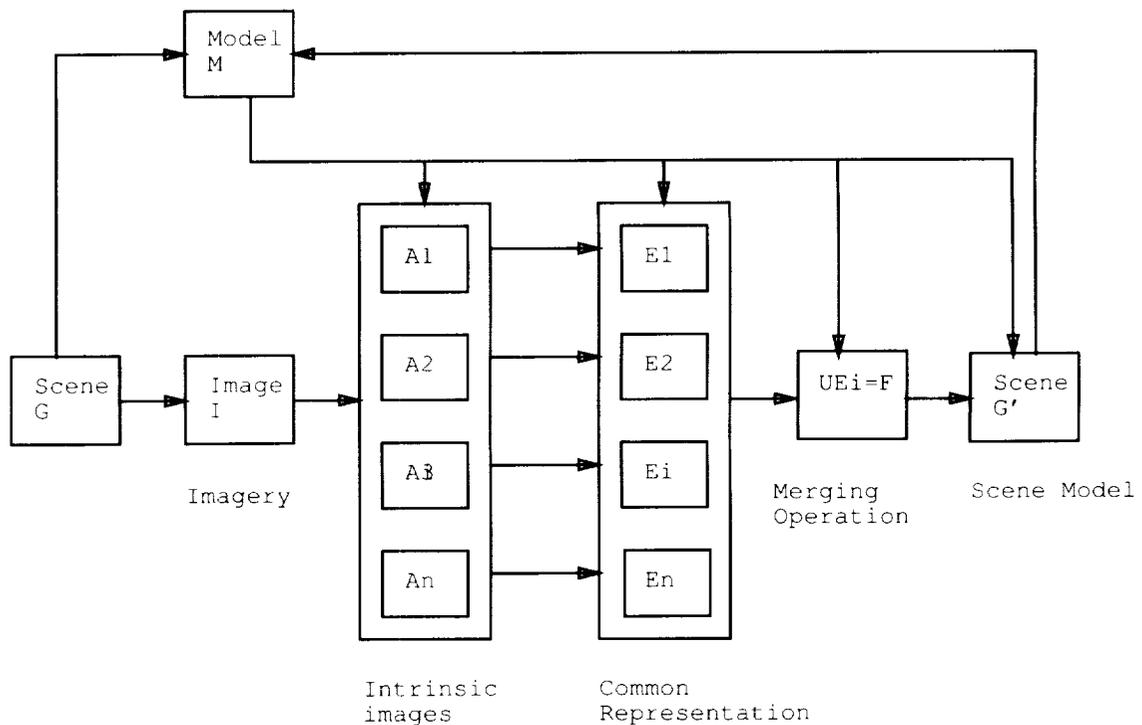**Figure 1-1:** Data fusion in image analysis

From the three-dimensional scene (G) we generally acquire two-dimensional imagery generated by a variety of different sensors. For example, a stereo pair of intensity images would represent such an imagery. As is well understood, the problem of interpreting the two-dimensional image (I) as a three-dimensional scene is underconstrained. In certain cases, we

may have access to high-level knowledge about the contents of the scene, or particular objects that can be found in the scene. Such knowledge can loosely be called a Model (M). For example, in the case of aerial imagery we may have knowledge about the sensor resolution, the general characteristics of the scene (airport, urban area, rural area), etc. From the representation (I), we try to extract features that will allow us to interpret the scene {Ai}. These features are typically segmentations, edge maps, disparity maps, intensity maps, and the like. These can be thought of as a set of intrinsic images and primitives for intermediate and high-level vision [1, 7]. In order to fuse the information embodied in these different "images", we need a common framework of representations (formed by the {Ei}). This framework needs to allow many, if not all, of the {Ai} features to be represented. The utilization of a common representation makes information fusion simpler and allows the generation of an interpretation (F), which then allows the generation of our scene model (G'). This model can be used to iterate through the fusion process again in conjunction with extra knowledge about the scene obtained from (M). This initial interpretation of the scene can help in the extraction of features {Ai}, the transformation of the features in the common representation, the merging process, and even the generation of the scene model.

Depending on the interpretation of the scene for which we are looking, we may need a varying amount of information; in most cases, more information is generally desirable. For instance, many techniques extract most of the necessary information for scene interpretation from a single intensity image; such techniques are said to apply *monocular analysis*. It is possible to take advantage of stereo disparity, however, to obtain more information that may be useful for disambiguation of monocular interpretations. Techniques utilizing stereo imagery are said to apply *binocular analysis* or *stereo analysis*. Other information such as global constraints or world models can be useful for further interpretation and disambiguation, but we believe that stereo analysis is a necessary step towards a coherent interpretation of the scene.

In this paper we describe a technique to merge information extracted from aerial imagery using a common region-based representation and show how disparate scene cues can be integrated to achieve a more complete and accurate overall scene interpretation. In Section 2 we describe techniques to improve the accuracy of a stereo disparity map using a single segmentation of the left intensity image of a stereo pair. Thus, we are able to recover from mismatches generated during stereo matching by re-utilizing the intensity image that was originally used in the matching process. In Section 3 we discuss some experimental results on disparity refinement and describe techniques that allow for the integration of additional scene segmentations to provide for a more robust refinement process. Finally, in Section 4 we give some future directions of this work in building extraction and built-up area analysis and speculate on how these techniques could be integrated into a more general three-dimensional scene interpretation system.

2. One approach to information fusion

In our research we utilize scene domain cues derived from monocular analysis and stereo analysis of left/right stereo image pairs. In the case of monocular analysis, one source of information is a region based segmentation of the left or right image. In the case of stereo analysis, our cues are primarily disparity maps derived from area-based and feature-based stereo matching algorithms. These image-based cues are different manifestations of man-made structures and terrain surfaces in the scene. In the case of three-dimensional reconstruction, we can make the assumption that the scene is composed of surfaces whose information content is primarily in terms of surface orientation and radiometry. Under these assumptions, we will see how estimates of three-dimensional scene structure (as encoded in a scene disparity map) can be

improved by the analysis of the original monocular imagery.

We have two sources of information that can be viewed as different representations of the physical surfaces found in the scene: disparity maps resulting from different stereo matchers providing the heights of the surfaces in the scene and the initial intensity images representing the radiometric properties of the surfaces in the scene. Figures 2-1 and 2-2 show an example of "initial" data used for these data fusion experiments. Figure 2-1 is a high resolution aerial image containing a variety of buildings with complex shapes, typical of an industrial area. Figure 2-2 is a disparity map derived using a feature-based stereo matching algorithm. These images are two of the many possible intrinsic images, {Ai}, in our general framework. It is important to note that, as in the intrinsic image paradigm, these two sources of information are "registered". That is, there is a pixel-by-pixel correspondence between points in the intensity image and points in the disparity map. In some many cases one issue complicating the use of multi-source information is the accurate registration or correspondence between the information sources themselves.

Figure 2-1: DC38008 industrial left intensity image
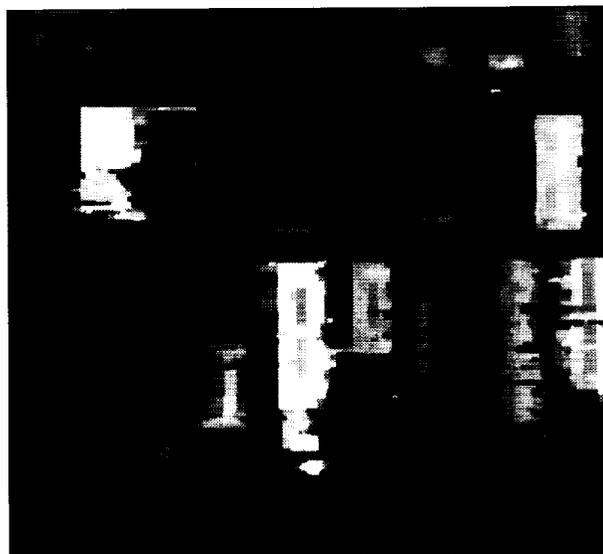
Figure 2-2: S2 left disparity map

An intensity image, subject to sampling and digitization errors, poses difficulties for monocular analysis techniques such as segmentation. On the other hand, most stereo matching algorithms are fooled by different variations in the stereo pairs, which cause mismatches in the disparity maps. The mismatches in disparity maps primarily result from geometric and radiometric differences in the left and right images, rather than local digitization or sampling errors in the intensity images. Thus, it is possible to use information from the intensity images to reduce the number of mismatches introduced by stereo matching processes.

2.1. Region based interpretation

Our approach utilizes surface illumination information, provided by the segmentation of the monocular images into fine surface patches of nearly homogeneous intensity, to remove mismatches generated during stereo matching. First, we segment the intensity image into uniform intensity regions. These regions correspond to approximately planar surfaces in the image. We assume that the orientation and surface material are the primary factors for the

radiometry of the image. Under these assumptions, uniform image radiometry is produced by a planar surface, of a certain orientation and material, in the scene.

These surfaces should have continuous linear disparity values (i.e., the disparity values of these regions are represented by continuous linear functions). Since the disparity map contains some noise, however, most of the regions segmented in the intensity image have disparity functions that are neither linear nor continuous. Ideally, we would like to approximate the actual disparity functions over the uniform intensity regions by the appropriate linear functions.

The problem of approximating a surface in three-dimensional space to a reasonable planar surface is a difficult one; we approximate such surfaces by horizontal surfaces. Then, the disparity values for each region will be the same for each pixel, and the problem is reduced to the selection of the best value for the heights of these surfaces. The general problem is that of locating of the surface which satisfies the equation

$$ax+by+cz+d=0$$

Given (x,y), we should be able to obtain

$$z = (-ax-by-d)/c$$

We assume here that $z' = -d'/c'$ only. Then the problem is to find $(-d'/c')$ that best fits the surface so that

$$ax+by+c*(-d'/c')+d\raise=0$$

or to find z' so that z-z' would have a minimal value over the region (this can be the weighted mean of the z distribution or the most 'representative' value of the z distribution). In other words, we need only select a single disparity value for each region. Since we are using an over-segmentation of the image, a piecewise planar disparity map gives a good approximation of the relief in the scene. Furthermore, since we are interested in building extraction in aerial images, this approximation will be adequate.

This region-based interpretation has been developed for two different applications. We show how this approach can support information fusion from different segmentations and well as across multiple disparity estimates based upon a local decision making evaluation. In Section 3.1 we describe how improved disparity maps may be obtained by correcting the mismatches produced by stereo matchers and by refining the disparity discontinuities. In Section 3.2 we will extract buildings from the scene using the height information in these disparity maps.

2.2. Intensity Segmentation Techniques

The general scene segmentation problem is, of course, a very difficult one and has a long history in image processing and computer vision. There are no universal segmentation techniques that work well across a variety of imagery and tasks. Such low level algorithms typically differ in their approaches; they may utilize intensity-based, area-based, or edge-based techniques. Some systems combine these techniques into hybrid algorithms. We have concentrated on those segmentation methods that produce (nearly) uniform intensity regions because we wish to detect those image regions that correspond to oriented surface patches in the scene. We utilize a region segmentation algorithm based upon the histogram splitting paradigm [6] and a region growing algorithm [9] which takes into account edge strength and shape criteria [4]. Interestingly, while neither of these methods give completely satisfactory segmentation results, they provide good over-segmentations that rarely merge object/background boundaries. Both techniques will also provide different segmentations based upon modification

of a small set of parameters. In our experiments we generated three scene segmentations; two by using different parameters for histogram selection, and one by using region growing. These segmentations provided the basis for our work in intensity/disparity fusion, the goal of which was to produce an improved three-dimensional scene interpretation.

Figures 2-4 - 2-6 show examples of these segmentations on the DC38008 industrial left intensity image. We ran the experiments on smoothed images (Figure 2-3) to remove intensity noise.

2.2.1. Machineseg

One of the major difficulties with region growing techniques in complex scenes is the difficulty in determining automatic stopping conditions for the merging proceedure. MACHINESEG [4] is a region growing system that tries to preserve edges between regions and stops the growing procedure when certain shape or spectral criteria are not satisfied inside the region. It adds a decision proceedure to evaluate the effect of the next merge operation and either allows the merge to proceed or to be rejected. In the case of disparity map refinement, we want the regions to be sufficiently uniform that they could be treated as planar (or at least "soft") surfaces. We also limited the size of the generated regions so that very small regions could not be generated, as these could be considered noise or non-representative regions. As can be seen in Figure 2-4, since we are not considering the small region, our segmentation is not a complete partition of the image; it does, however, obtain most of the representative surfaces in the image.

Figure 2-3: Nagao filtered left image for DC38008

Figure 2-4: MACHINESEG segmentation on DC38008

2.2.2. Colorseg

This histogram splitting technique is based on the extraction of regions with limited intensity ranges (in other words, region of approximately uniform intensities). The technique searches for the peaks in the histogram of the image and segments the regions whose intensity values fall in windows around these peaks. The regions are then removed from the image and the process continues until all the pixels in the image have been removed. This process results in a segmentation composed of connected regions, each having an intensity range less than a certain threshold. This technique does not guarantee preservation of the edges (in particular, small edges) but it may ignore local noise with strong edges that other techniques will classify as

88

regions. As in the previous technique, we removed very small regions (less than 20 pixels) that could be considered as noise, for further processing.

In our experiments, we generated different segmentations with different segmentation techniques. For instance, using the colorseg technique we generated two segmentations of the images, one with "uniformity" defined as a maximum of 10 intensity levels inside the region (to tolerate sensor noise and allow for imperfect planar surfaces) and another with "uniformity" defined as a maximum of 20 intensity levels (to tolerate more noise). An estimation of the noise or the average intensity range for the surfaces in the image is a delicate problem, and the use of different segmentations to estimate the intensity range inside the regions does not necessarily increase the reliability of the process. It is thus important that we obtain different segmentations of the scene that are *not consistent*, such as those in Figures 2-5 and 2-6. The fusion of these data may overcome some of the inherent problems of a single segmentation since they provide different local evaluation contexts for disparity estimates in the scene. In the following sections we show how we can merge information using different intensity segmentations.
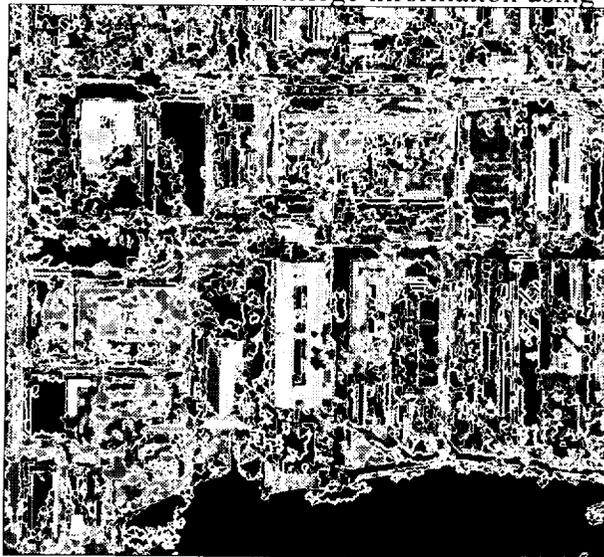
Figure 2-5: COLORSEG segmentation with 10 intensity levels sensitivity for DC38008

Figure 2-6: COLORSEG segmentation with 20 intensity levels sensitivity for DC38008

2.3. Disparity map results

Our initial height information for the industrial scene was derived using two different stereo matching algorithms. Given these sets of height information, which may or may not be reliable or unique, it becomes necessary to use a data fusion process in order to maximize the amount of useful information gained from these sets of height estimates.

We used 2 different matching techniques, one area_based (S1) and the other feature_based (S2). S1 uses the method of differences technique on neighborhoods of the image in hierarchical fashion [3, 5]. S2 performs a hierarchical matching of epipolar intensity scanlines in the left and right image [2]. The results of these stereo matching algorithms are different: S1 gives us a dense disparity map (i.e., a map containing a disparity value for each pixel in the image), while S2 gives us a sparse disparity map (i.e., a map containing a disparity value for those pixels corresponding to peaks or valleys in the intensity images).

Since we used uniform segmented regions that we assumed to be horizontal planes, a logical interpolation method for the sparse S2 disparity map is step interpolation. This produces a dense disparity map consisting of regions with uniform disparity values, which may be more easily integrated with a dense map produced by S1. Our fusion mechanism will have to correct mismatches in the S1 or S2 disparity maps and then choose the better unique disparity value for each pixel in the scene. It will have to merge very different disparity information, such as that shown in Figures 3-2 and 3-1, the two left disparity maps for the DC38008 scene.

3. Fusion Experiments

After different intensity segmentations and different disparity results were obtained, we applied a very simple fusion technique and developed a few experiments for the two applications under consideration. Most of the experiments have been performed for the disparity refinement process, but the results have been used for the building extraction process as well.

Figure 3-1: S1 left disparity result for DC38008
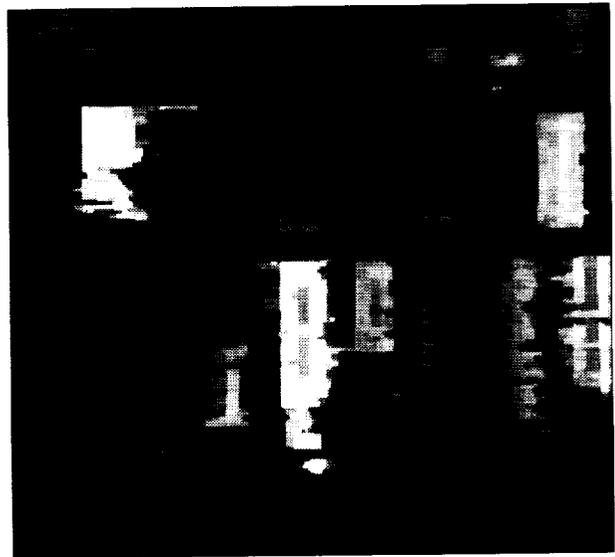
Figure 3-2: S2 left disparity result for DC38008

3.1. Disparity refinement

In order to refine the disparity maps (i.e., to remove mismatches, improve disparity discontinuities and obtain the best height estimate for each point in the scene), several approaches have been explored:

- Disparity refinement using one segmentation
- Disparity refinement using several segmentations
- Disparity refinement using one segmentation and several disparity maps
- Disparity refinement using several segmentations and several disparity maps

3.1.1. Simple disparity refinement

In this first approach, a histogram is constructed for each segmentation region. The values of each histogram are the disparity values in each region. The most representative value of each histogram is then selected. In our case, this value was simply that of the highest peak in the histogram. We chose this value for two reasons. The step-interpolated S2 disparity maps result

in disparity histograms having only a few values, which correspond to real height values or matching noise. If the matching is reasonably robust, the noise will introduce local maxima in the histogram that will be smaller in magnitude than the best height estimate. Further, a typical region histogram for an S2 disparity map exhibits one or two large peaks and a few noise peaks that influence the average value of the histogram, making it less reliable as a representative value.

For non-horizontal regions and S1 results, the average disparity may suffice for a reasonable measure of the height of the region. A confidence score can be generated for these disparity values based on the characteristics of the histograms (and, conceivably, on the type of disparity map used as well as the nature of the region histograms). Finally, this disparity value is assigned to the entire region, under the assumption that it will be a better estimate of the height for the whole region. In most cases, this removes a large number of the mismatches, but whenever our initial assumptions about scene radiometry are not valid, our height estimates may differ from the correct height value.

We implemented this approach for each segmentation and disparity map and generated new disparity maps that were based on the initial intensity regions and disparity values. The pixels that were not considered during the segmentation were removed from these new disparity maps. Figures 3-3 and 3-4 show the results of the disparity improvement process for the different segmentations using the S2 disparity map, and Figures 3-5 and 3-6 show the results of the disparity improvement process for the S1 disparity map.

Figure 3-3: S2 left disparity result for DC38008 improved using SEG10

Figure 3-4: S2 left disparity result for DC38008 improved using SEG20

It is worth noting that a common methodology is utilized among all of the approaches described in this section. A set of attributes is computed for each region in each segmentation. Among these attributes are the statistics for the disparity values inside a region, the best disparity value, and a confidence score for this value. This allows the computation to proceed at a symbolic level on a region-by-region basis.

Figure 3-5: S1 left disparity
result for DC38008
improved using SEG10

Figure 3-6: S1 left disparity
result for DC38008
improved using SEG20

Figure 3-7: S1 left disparity
result for DC38008
improved using the merging
of SEG10 and SEG20

Figure 3-8: S2 left disparity
result for DC38008
improved using the merging
of SEG10 and SEG20

3.1.2. Multi-segmentation disparity refinement

In the second approach, we can merge different height estimates, given different intensity segmentations(SEG10, SEG20) and then merging the results across the different segmentations. We refine the disparity estimate for each pixel by locating the intensity region to which it belongs, for each of the image segmentations. This list of regions can then be searched to obtain the disparity estimate attribute (computed for a given disparity map) as well as a confidence

92

score for this estimate. The confidence score is then used to select the best disparity value, which is then assigned to the pixel. Currently a simple decision is made to select the disparity value having the highest confidence score.

An attempt is made to maximize the score for each pixel in the entire image. This is done by selecting a disparity value in all of the regions resulting from the union of the segmentations. In other words, the segmentations were merged and the best height value was selected for each of these regions, by utilizing the confidence scores computed for each region. The scoring method currently in use takes into account information about the nature of the segmentation used.

In particular, higher confidences can be assigned to sufficiently large regions in a constrained segmentation such as SEG10 than to the equivalent regions in SEG20. Information of this nature must be incorporated in the confidence function for each segmentation region.

Figures 3-8 and 3-7 show the results of merging the SEG10 and the SEG20 segmentations for the S2 and the S1 disparity maps, respectively. Depending on the confidence scores of the disparity values selected for each segmentation, we were able to obtain improved disparity estimates for some of the regions. Comparing these results to Figures 3-3 and 3-4, disparity maps obtained with the simple method, we observe some of the failings of both approaches. The initial segmentations, in some cases, are under-segmented instead of over-segmented, resulting in the grouping of regions that should have been assigned different height estimates. Another factor is the confidence evaluation function for the regions of the segmentation, which only takes simple properties of the disparity histograms of each region into account.

3.1.3. Multi-Disparity Disparity Refinement

In this approach, several different disparity maps are merged using a single segmentation, looking for consistent areas across disparity maps. This approach is similar to the simple disparity improvement approach, except that we now attempt to select the best disparity value based on a set of differing confidence scores. The score established for each disparity map at each pixel should be dependent on the stereo matching algorithm used to generate the map, and should also take into account the nature of the possible mismatches resulting from each stereo matching technique.

The major problem with all of the refinement approaches discussed in this paper is the development of a reasonable confidence evaluation function for each set of data. Currently, confidence is evaluated by a scoring function that utilizes the standard deviation and the disparity range of the histogram for each region, as well as the size of the region. Ideally, this scoring function would also take into account the nature of the disparity map. As an initial experiment, we defined a similar scoring function for each disparity map and checked for disparity consistency across segmentation regions. In Figure 3-9, the areas where disparity values differ between S1 and S2 are marked in black, as we do not use any score difference information to select the most probable height value at this stage.

3.1.4. General Disparity Refinement

For the general case we can merge the results of different disparity maps and different segmentations and look for consistency across the results. The approach is similar to the multi-segmentation method; however, we should be able to add additional height hypotheses according to the different segmentations.

Again, the processes can be decomposed into two stages. The first stage will gather the

93

information and convert it into a common representation (i.e., region attributes). As an example, for each segmentation we should obtain a list of height estimates with scores associated with each of the different disparity maps we can use (S1 and S2). The second stage will attempt to merge this information by selecting the "correct" value from the available information, by comparing scores based on the nature and quality of the different pieces of information. If we can precisely evaluate the quality or confidence in the information, we should be able to maximize the amount of accurate data we merge from our different information sources.

There are still many experiments that have yet to be performed. In particular, experimentation needs to be done on merging the two different disparity values for the three different segmentations.
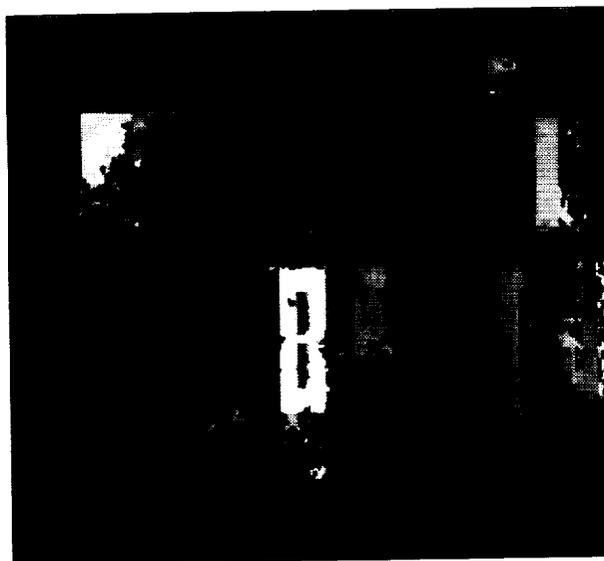
Figure 3-9: S1 left disparity
and S2 left disparity
merged using YAK

3.2. Building extraction

This second application of information fusion is an attempt to validate this region-based approach for scene interpretation. Using the previously described methods, we can obtain an estimate of the height of each of the composite regions in each segmentation. According to our representation of the scene, buildings are composed of a single intensity region or a group of intensity regions, and, in general, are higher than their surroundings. Therefore, regions representing parts of a building should be higher than their neighboring regions.

For each region, a list of its neighboring regions is constructed, and the disparity values for each of these regions are obtained. Then, a weighted histogram is computed that takes into account shared boundary length and disparity information. This weighted score is then compared with the height of the region to label the region as building structure or background terrain. This building extraction process can use either the initial disparity map or the refined disparity map.

A refinement process is used to group neighboring regions with the same height in order to obtain an intermediate segmentation containing fewer (and larger) consistent regions. This

94

grouping procedure merges connected regions having the same height to form a single region. This allows the building extraction process to use larger, and hopefully more consistent, disparity regions as a basis for the neighborhood disparity analysis. The quality of this analysis is again dependent on the accuracy of the disparity estimate, as in the previous fusion process.

Figure 3-10 shows the result of such an analysis. The white regions correspond to sections of buildings. The building extraction, as done by hand, is in Figure 3-11.

Figure 3-10: Building regions for DC38008 extracted using the merging of SEG10 and SEG20

Figure 3-11: Building regions for DC38008 extracted manually

The problem can be described as the use of "early" or "initial" information for which we do not have any confidence measures to construct a model. To perform this task, we must gather confidence about this information as computation proceeds in order to construct a three-dimensional interpretation of the scene. The building extraction process described here illustrates one facet of scene interpretation that can be performed within this framework.

4. Conclusions

We have described a set of fusion processes that allow us to improve the quality of disparity maps, and we have demonstrated the use of information fusion to improve disparity map analysis. We described a building extraction approach that utilized the fusion technique. The major feature of the information fusion technique described here is the definition of a common frame for information fusion. The representation framework (an intensity segmentation) can be used in conjunction with different types of intrinsic images. The approach developed here treats homogeneous intensity regions as surfaces, which allows three-dimensional information to be extracted readily.

Many research issues remain to be explored. The new disparity maps generated by the information fusion process contain regions which each have only one disparity value. In many cases, these unique values are not the best possible disparity estimates for the regions, and a refinement process may need to be invoked to correct these estimates. One approach might be to

use the new disparity map itself as input to a verification process which could refine disparity estimates for each pixel or for those regions with low confidence scores.

Other sources of information could be utilized at the refinement stage to further enhance the disparity map. One promising approach would be the use of left/right consistency, such as left/right matching of low confidence regions or local correlation for these regions. Again, it would be important to use as much information as possible, while conservatively adjusting or refining data based on its confidence scores. In the ideal situation, no additional information would refine the disparity estimates; it would merely verify the truth of the disparity map.

Many improvements can be obtained by the use of better segmentations and scoring functions, and by addressing the assumption that only flat horizontal surfaces are responsible for the imaged radiometry and by using a more sophisticated surface model such as non-horizontal planar surfaces or quadratic surfaces. Finally, it seems feasible that multispectral data could be integrated by similar techniques. The information fusion approaches described here provide a means for data integration that may prove useful in other aspects of scene interpretation.

5. Acknowledgments

References

[1]   Barrow, H. G. and Tenenbaum, J. M.
         Recovering Intrinsic Scene Characteristics from Images.
         In Hanson, A. R. and Riseman, E. M. (editor), *Computer Vision Systems*, pages 3-26.
             Academic Press, New York, 1978.

[2]   Hsieh, Y., Perlant, F., and McKeown, D. M.
         Recovering 3D Information from Complex Aerial Imagery.
         In *Proceedings: 10th International Conference on Pattern Recognition, Atlantic City,
             New Jersey*, pages 136-146. June, 1990.

[3]   B. D. Lucas.
         Generalized Image Matching By The Method of Differences.
         PhD thesis, Carnegie Mellon University, July, 1984.

[4]   McKeown, D.M., Denlinger, J.L.
         Map-Guided Feature Extraction from Aerial Imagery.
         In *Proceedings of Second IEEE Computer Society Workshop on Computer Vision:
             Representation and Control*, pages 205-213. May, 1984.

[5]   McKeown, D.M., McVay, C.A., and Lucas, B. D.
         Stereo Verification In Aerial Image Analysis.
         Optical Engineering 25(3):333-346, March, 1986.

[6]   Ohlander, R., Price, K., and Reddy, D. R.
         Picture Segmentation Using a Recursive Region Splitting Method.
         Computer Graphics and Image Processing 8:313-333, 1978.

[7]   Poggio, T. et. al.
        The MIT Vision Machine.
        In *Proceedings of the DARPA Image Understanding Workshop*, pages 177,198. Morgan
            Kaufmann, April, 1988.

[8]   Shufelt, J., McKeown, D. M.
        Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery.
        In *Proceedings: IAPR Workshop on Multisource Data Integration in Remote Sensing*.
            June, 1990.

[9]   Yoram Yakimovsky.
        Boundary and Object Detection in Real World Images.
        Journal of the ACM 23(4), 1976.

Fusion of Monocular Cues
to Detect Man-Made Structures in Aerial Imagery

Jefferey Shufelt and David M. McKeown

Digital Mapping Laboratory
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213[1]

Abstract

The extraction of buildings from aerial imagery is a complex problem for automated computer vision. It requires locating regions in a scene that possess properties distinguishing them as man-made objects as opposed to naturally occurring terrain features. The building extraction process requires techniques that exploit knowledge about the structure of man-made objects. Techniques do exist that take advantage of this knowledge; various methods use edge-line analysis, shadow analysis, and stereo imagery analysis to produce building hypotheses. It is reasonable, however, to assume that no single detection method will correctly delineate or verify buildings in every scene. As an example, a feature extraction system that relies on analysis of cast shadows to predict building locations is likely to fail in cases where the sun is directly above the scene.

It seems clear that a cooperative-methods paradigm is useful in approaching the building extraction problem. Using this paradigm, each extraction technique provides information which can then be added or assimilated into an overall interpretation of the scene. Thus, our research focus is to explore the development of a computer vision system that integrates the results of various scene analysis techniques into an accurate and robust interpretation of the underlying three-dimensional scene.

This paper describes preliminary research on the problem of building hypothesis fusion in aerial imagery. Building extraction techniques are briefly surveyed, including four building extraction, verification, and clustering systems that form the basis for the work described here. A method for fusing the symbolic data generated by these systems is described, and applied to monocular image and stereo image data sets. Evaluation methods for the fusion results are described, and the fusion results are analyzed using these methods.

---

1. Introduction

In the cooperative-methods paradigm it is assumed that no single method can provide a complete set of building hypotheses for a scene. However, each method may provide a subset of the information necessary to produce a more meaningful interpretation of the scene. For instance, a shadow-based method might provide unique information in situations where ground and roof intensity are similar. An intensity-based method can provide boundary information in instances where shadows were weak or nonexistent, or in situations where structure height was sufficiently low that stereo disparity analysis would not provide reliable information. The implicit assumption behind this paradigm is that the symbolic interpretations produced by each of these techniques can be integrated into a more meaningful collection of building hypotheses.

It is reasonable to expect that there will be complications in fusing real monocular data. In the best case, the building hypotheses will not only be accurate, but complementary. It is just as likely, however, that some building hypotheses may be unique. Further, it is rare that building hypotheses are always accurate, or even mutually supportive of one another. For a cooperative-methods data fusion system to be successful, it must address the problems of redundant and conflicting data.

2. Building extraction techniques

At the Digital Mapping Laboratory, we have developed several techniques for the extraction of man-made objects from aerial imagery. The goal of many of these techniques is to organize the image into manageable parts for further processing, by using external knowledge to organize these parts into regions.

For the experiments described in this paper, a set of four monocular building detection and evaluation systems were used. Three of these were shadow-based systems; the fourth was line-corner based. The shadow based systems are described more fully by Irvin and McKeown [5], and the line-corner system is described by Aviad, McKeown, and Hsieh [2]. A brief description of each of the four detection and evaluation systems follows.

BABE (Builtup Area Building Extraction) is a building detection system based on a line-corner analysis method. BABE starts with intensity edges for an image, and examines the proximity and angles between edges to produce corners. To recover the structures represented by the corners, BABE constructs chains of corners such that the direction of rotation along a chain is either clockwise or counterclockwise, but not both. Since these chains may not necessarily form closed segmentations, BABE generates building hypotheses by forming boxes out of the individual lines that comprise a chain. These boxes are then evaluated in terms of size and line intensity constraints, and the best boxes for each chain are kept, subject to shadow intensity constraints [4], [7].

SHADE (SHAdow DEtection) is a building detection system based on a shadow analysis method. SHADE uses the shadow intensity computed by BABE as a threshold for an image. Connected region extraction techniques are applied to produce segmentations of those regions with intensities below the threshold, i.e., the shadow regions. SHADE then examines the edges comprising shadow regions, and keeps those edges that are adjacent to the buildings casting the shadows. These edges are then broken into nearly straight line segments by the use of an imperfect sequence finder [1]. Those line segments that form nearly right-angled corners are joined, and the corners that are concave with respect to the sun are extended into parallelograms, SHADE's final building hypotheses.

SHAVE (SHAdow VErification) is a system for verification of building hypotheses by shadow analysis. SHAVE takes as input a set of building hypotheses, an associated image, and a shadow threshold produced by BABE. SHAVE begins by determining which sides of the hypothesized building boxes could possibly cast shadows, given the sun illumination angle, and then performs a walk away from the sun illumination angle for every pixel along a building/shadow edge to delineate the shadow. The edge is then scored based on a measure of the variance of the length of the shadow walks for that edge. These scores can then be examined to estimate the likelihood that a building hypothesis corresponds to a building, based on the extent to which it casts shadows.

GROUPER is a system designed to cluster, or group, fragmented building hypotheses, by examining their relationships to possible building/shadow edges. GROUPER starts with a set of hypotheses and the building/shadow edges produced by BABE. GROUPER back-projects the endpoints of a building/shadow edge towards the sun along the sun illumination angle, and then connects these projected endpoints to form a region of interest in which buildings might occur. GROUPER intersects each building hypothesis with these regions of interest. If the degree of overlap is sufficiently high (the criteria is currently 75% overlap), then the hypothesis is assumed to be a part of the structure which is casting the building/shadow edge. All hypotheses that intersect a single region of interest are grouped together to form a single building cluster.

There are many other interesting building detection and extraction techniques. We briefly mention some recently developed methods, to illustrate the variety of techniques that produce building hypothesis information. Although this by no means constitutes a comprehensive survey of building detection techniques, it provides some examples of the methods used to generate hypotheses for a scene, as well as examples of the types of data that may eventually be integrated into a cooperative-methods building analysis scheme.

Mohan and Nevatia [6] described a method by which simple image tokens such as lines or edges could be clustered into more complex geometric features consisting of parallelopipeds. Huertas and Nevatia [4] described a method for detecting buildings in aerial images. Their method detected lines and corners in an image and constructed chains of these to form building hypotheses which were then subject to shadow verification.

Fua and Hanson [3] described a system that used generic geometric models and noise-tolerant geometry parsing rules to allow semantic information to interact with low-level geometric information, producing segmentations of objects in the aerial image. Nicolin and Gabler [7] described a system for analysis of aerial images. The system had four components: a method-base of domain-independent processing techniques, a long-term memory containing *a priori* knowledge about the problem domain, a short-term memory containing intermediate results from the image analysis process, and a control module responsible for invocation of the various processing techniques. Gray-level analysis was applied to a resolution pyramid of imagery to suggest segmentation techniques, and structural analysis was performed after segmentation to provide geometric interpretations of the image.

3. A simple hypothesis merging technique

Building hypotheses typically take the form of geometric descriptions of objects in the context of an image. One can imagine "stacking" sets of these geometric descriptions on the image: in the process, those regions of the image that represent man-made structure in the scene should accumulate more building hypotheses than those regions of the image that represent natural features in the scene. The merging technique developed here exploits this idea.

The method takes as input an arbitrary collection of polygons. An image is created that is sufficiently large to contain all of the polygons, and each pixel in this image is initialized to zero. Each polygon is scan-converted into the image, and each pixel touched during the scan is incremented. The resulting image then has the property that the value of each pixel in the image is the number of input polygons that cover it.

Segmentations can then be generated from this "accumulator" image by applying connected region extraction techniques. If the image is thresholded at a value of 1 (i.e, all non-zero pixels are kept), the regions produced by a connected region extraction algorithm will simply be the geometric unions of the input polygons. It is the case, however, that the image could be thresholded at higher values. We motivate thresholding experiments in Section 4.4.

4. Merging multiple hypothesis sets

This section outlines the experiments performed with the scan-conversion hypothesis fusion technique. The procedure used to apply this technique to the results of four building detection and evaluation systems (BABE, SHADE, SHAVE, and GROUPER) is described. A technique for quantitative evaluation of building hypotheses is described, and applied to the hypothesis fusion results. These results are analyzed to suggest improvements to the fusion technique.

4.1. The merging technique applied to four extraction systems

There were two merging problems under consideration. The first of these was the creation of a single hypothesis out of a collection of fragmented hypotheses believed to correspond to a single man-made structure. This problem was addressed by applying the scan-conversion technique to the fragmented clusters produced by GROUPER. The technique was applied to each cluster individually, and the resulting accumulator image was thresholded at 1, and connected region extraction techniques were applied to provide the geometric union of each cluster. These clusters were then used as the building hypotheses produced by GROUPER.

The second problem was the fusion of each of these monocular hypothesis sets into a single set of hypotheses for the scene. Again, the scan-conversion technique was applied. The four hypothesis sets were scan-converted, and the resulting accumulator image was thresholded at 1. Connected region extraction techniques were applied to produce the final segmentation for the image.

Figure 4-1 shows a section of a suburban area in Washington, D.C. Figure 4-2 shows the SHADE results for this scene, Figure 4-3 shows the SHAVE results, Figure 4-4 shows the GROUPER results, and Figure 4-5 shows the BABE results. Figure 4-6 shows the fusion of these four monocular hypothesis sets.

4.2. Evaluation of the technique

To judge the correctness of an interpretation of a scene, it is desirable to have some mechanism for quantitatively evaluating that interpretation. One approach is to compare a given set of hypotheses against a set that is known to be correct, and analyze the differences between the given set of hypotheses and the correct ones. In performing evaluations of the fusion results, we use *ground-truth segmentations* as the correct detection results for a scene. Ground-truth segmentations are manually produced segmentations of the buildings in an image.

Figure 4-1: DC37 image with ground-truth segmentation

There exist two simple criteria for measuring the degree of similarity between a building hypothesis and a ground-truth building segmentation: the mutual area of overlap and the difference in orientation. A correct building hypothesis and the corresponding ground-truth segmentation region should cover roughly the same area, and should have roughly the same alignment with respect to the image. A scoring function can be developed that incorporates these criteria. A region matching scheme such as this, however, suffers from the fact that multiple buildings in the scene are segmented by a single region in the hypothesis set. In these cases, the building hypothesis will have low matching scores with each of the buildings it contains, due to the differences in overlap area.

A simpler coverage-based global evaluation method was developed. This evaluation method works in the following manner. H, a set of building hypotheses for an image, and G, a ground-truth segmentation of that image, are given. The image is then scanned, pixel by pixel. For any pixel P in the image, there are four possibilities:
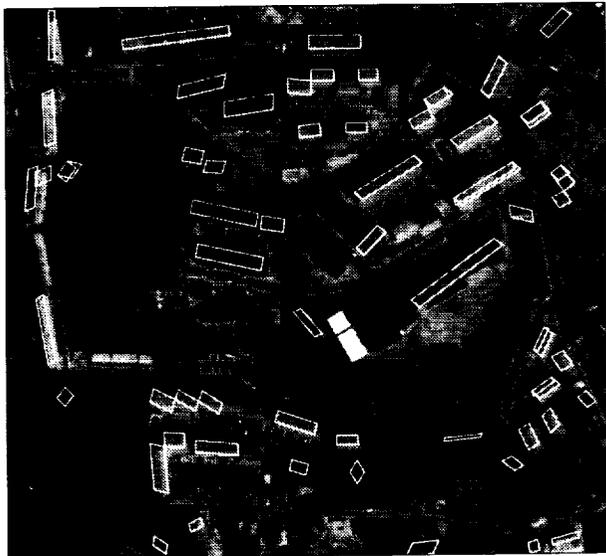
Figure 4-2: DC37 SHADE results
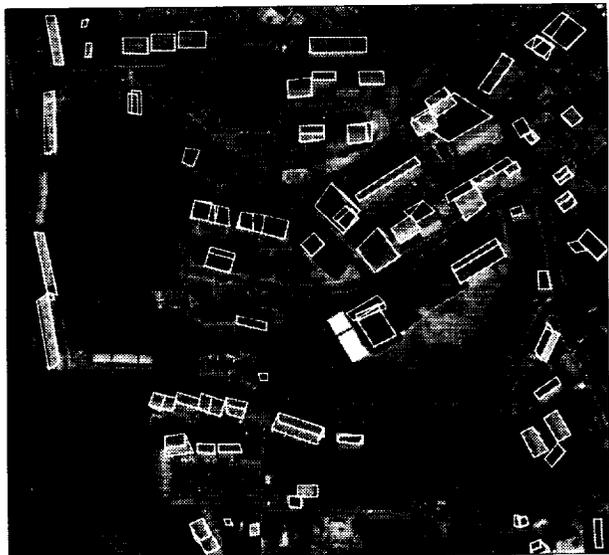
Figure 4-3: DC37 SHAVE results

Figure 4-4: DC37 GROUPER results
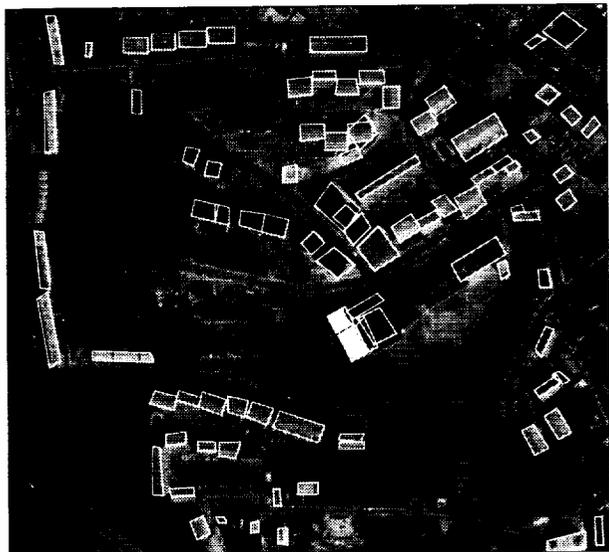
Figure 4-5: DC37 BABE results

1. Neither a region in H nor a region in G covers P. This is interpreted to mean that the system producing H correctly denoted P as being part of the background, or natural structure, of the scene.

2. No region in H covers P, but a region in G covers P. This is interpreted to mean that the system producing H did not recognize P as being part of a man-made structure in the scene. In this case, the pixel is referred to as a "false negative".

3. A region (or regions) in H cover P, but no region in G covers P. This is interpreted to mean that the system producing H incorrectly denoted P as belonging to some man-made structure, when it is in fact part of the scene's background. In this case,

Figure 4-6: Monocular hypothesis fusion for DC37

the pixel is referred to as a "false positive".

4. A region (or regions) in H and a region in G both cover P. This is interpreted to mean that the system producing H correctly denoted P as belonging to a man-made structure in the scene.

By counting the number of pixels that fall into each of these four categories, we may obtain measurements of the percentage of building hypotheses that were successful (and unsuccessful) in denoting pixels as belonging to man-made structure, and the percentage of the background of the scene that was correctly (and incorrectly) labeled as such. Further, we may use these measurements to define a *building pixel branching factor*, which will represent the degree to which a building detection system overclassifies background pixels as building pixels in the process of generating building hypotheses. The building pixel branching factor is defined as the number of false positive pixels divided by the number of correctly detected building pixels.

4.3. Results and analysis

The fusion process was run on other scenes in addition to the DC37 scene: DC36A, DC36B, and DC38, three more scenes from the Washington, D.C. area; and LAX, a scene from the Los Angeles International Airport. The coverage-based evaluation program was then applied to generate Tables 4-1 through 4-5. Each table gives the statistics for a single scene. The first column represents a building extraction system. The next two columns give the percentage of building and background terrain correctly identified as such. The fourth and fifth columns show incorrect identification percentages for buildings and terrain. The next two columns give the breakdown (in percentages) of incorrect pixels in terms of false positives and false negatives. The last column gives the building pixel branching factor.

Evaluation results for the fusion process on DC37							
System	% Bld Detected	% Bkgd Detected	% Bld Missed	% Bkgd Missed	% False Pos.	% False Neg.	Br Factor
SHADE	37.5	98.2	62.5	1.8	15.0	85.0	0.294
SHAVE	47.2	96.8	52.8	3.2	26.8	73.2	0.408
GROUPER	48.7	95.8	51.3	4.2	32.6	67.4	0.508
BABE	58.9	97.2	41.1	2.8	28.5	71.5	0.278
FUSION	77.7	92.0	22.3	8.0	68.0	32.0	0.611
99 regions in ground truth							

Table 4-1: Evaluation statistics for DC37 hypothesis fusion

Evaluation results for the fusion process on DC36A							
System	% Bld Detected	% Bkgd Detected	% Bld Missed	% Bkgd Missed	% False Pos.	% False Neg.	Br Factor
SHADE	53.8	97.0	46.2	3.0	30.7	69.3	0.381
SHAVE	63.6	96.2	36.4	3.8	41.8	58.2	0.411
GROUPER	58.0	95.8	42.0	4.2	40.6	59.4	0.495
BABE	51.0	97.9	49.0	2.1	22.1	77.9	0.273
FUSION	80.9	91.9	19.1	8.1	74.3	25.7	0.682
51 regions in ground truth							

Table 4-2: Evaluation statistics for DC36A hypothesis fusion

We note that the quantitative results generated by the new evaluation method accurately reflect the visual quality of the set of building hypotheses. Further, the building pixel branching factor provides a rough estimate of the amount of noise generated in the fusion process. Judging by these measures, we note that the final results of the hypothesis fusion process significantly improve the detection of buildings in a scene. In all of the scenes, the detection percentage for the final fusion is greater than the same percentage for any of the individual extraction system hypotheses, although the building pixel branching factor also increases due to the accumulation of delineation errors from the various input hypotheses.

106

Evaluation results for the fusion process on DC36B						
System	% Bld Detected	% Bkgd Detected	% Bld Missed	% Bkgd Missed	% False Pos.	% False Neg.	Br Factor
SHADE	29.8	93.8	70.2	6.2	46.3	53.7	2.034
SHAVE	28.4	96.7	71.6	3.3	31.3	69.7	1.146
GROUPER	10.3	96.8	89.7	3.2	25.9	74.1	3.027
BABE	9.9	98.8	90.1	1.2	11.3	88.7	1.159
FUSION	49.8	89.2	50.2	10.8	67.8	32.2	2.126
133 regions in ground truth						

Table 4-3: Evaluation statistics for DC36B hypothesis fusion

Evaluation results for the fusion process on DC38						
System	% Bld Detected	% Bkgd Detected	% Bld Missed	% Bkgd Missed	% False Pos.	% False Neg.	Br Factor
SHADE	51.3	97.4	48.7	2.6	13.2	86.8	0.144
SHAVE	43.1	95.3	56.9	4.7	19.1	80.9	0.311
GROUPER	54.6	95.8	45.4	4.2	21.0	79.0	0.221
BABE	44.7	96.0	55.3	4.0	17.3	82.7	0.260
FUSION	74.7	90.6	25.3	9.4	51.5	48.5	0.360
53 regions in ground truth						

Table 4-4: Evaluation statistics for DC38 hypothesis fusion

Evaluation results for the fusion process on LAX						
System	% Bld Detected	% Bkgd Detected	% Bld Missed	% Bkgd Missed	% False Pos.	% False Neg.	Br Factor
SHADE	34.4	99.0	65.6	1.0	10.1	89.9	0.213
SHAVE	54.1	94.9	45.9	5.1	43.6	56.4	0.655
GROUPER	46.0	98.5	54.0	1.5	16.5	83.5	0.232
BABE	63.3	98.8	36.7	1.2	18.3	81.7	0.130
FUSION	73.0	92.9	27.0	7.1	65.0	35.0	0.687
26 regions in ground truth						

Table 4-5: Evaluation statistics for LAX hypothesis fusion

It is worth noting that the results for the DC36B scene (Table 4-3) are substantially worse than those of the other scenes. This is in large part due to the fact that the DC36B scene has a low dynamic range of intensities, and the component systems used for these fusion experiments are inherently intensity-based. The building pixel branching factors reflect the poor performance of

107

the component systems; in GROUPER's case, over 3 pixels are incorrectly hypothesized as building pixels for every correct building pixel. The fusion process, however, improved the building detection percentage noticeably over the percentages of the component systems.

We also note that several difficulties are attributable to performance deficiencies in the systems producing the original building hypotheses. The shadow-based detection and evaluation systems, SHADE and SHAVE, both use a threshold to generate "shadow regions" in an image. This threshold is generated automatically by BABE, a line-corner based detection system. In some cases, the threshold is too low, and the resulting shadow regions are incomplete, which results in fewer hypothesized buildings.

GROUPER, the shadow-based hypothesis clustering system, clusters fragmented hypotheses by forming a region (based on shadow-building edges) in which building structure is expected to occur. This region is typically larger than the true building creating the shadow-building edge, and incorrect fragments sometimes fall within this region and are grouped with correct fragments. The resulting groups tend to be larger than the true buildings, and thus produce a fair number of false positive pixels.

SHAVE scores a set of hypotheses based on the extent to which they cast shadows, and then selects the top fifteen percent of these as "good" building hypotheses. In some cases, buildings whose scores fell in the top fifteen percent actually had relatively low absolute scores. This resulted in the inclusion of incorrect hypotheses in the final merger.

SHADE uses an imperfect sequence finder to locate corners in the noisy shadow-building edges produced by thresholding. The sequence finder uses a threshold value to determine the amount of noise that will be ignored when searching for corners. In some situations, the true building corners are sufficiently small that the sequence finder regards them as noise, and as a result, the final building hypotheses can either be erroneous or incomplete.

4.4. Thresholding the accumulator image

As part of the scan-conversion fusion process, an accumulator image is produced which represents the "building density" of the scene. More precisely, each pixel in the image has a value, which is the number of hypotheses that overlapped the pixel. Pixels with higher values represent areas of the image that have higher probability of being contained in a man-made structure. Theoretically, thresholding this image at higher values and then applying connected region extraction techniques would produce sets of hypotheses containing fewer false positives, and these hypotheses would only represent those areas that had a high probability of corresponding to structure in the scene.

To test this idea, the accumulator images for each of the six scenes were thresholded at values of 2, 3, and 4, since four systems were used to produce the final hypothesis fusion. Connected region extraction techniques were then applied to these thresholded images to produce new hypothesis segmentations. The new evaluation method was then applied to these new hypotheses.

In each of the scenes, increasing the threshold from its default value of 1 to a value of 2 causes a reduction of roughly 20 percent in the number of correctly detected building pixels. This suggests that a fair number of hypothesized building pixels are unique; i.e., several pixels can only be correctly identified as building pixels by one of the detection methods. Another interesting observation is that the building pixel branching factor roughly doubles every time the

108

threshold is decremented. These observations suggest that thresholding alone may eliminate unique information produced by the individual detection systems, and that more work will need to be done to limit the number of false positives (and erroneous delineations) produced by each system, and by the final fusion as a whole.

5. Conclusions

This paper has described a simple method for fusing sets of monocular building hypotheses for aerial imagery. Scan-conversion and connected region extraction techniques were applied to produce mergers of sets of building hypotheses, and the results were analyzed by the use of an evaluation technique based on pixel coverage.

The simple hypothesis fusion approach developed here appears promising; the detection rate can be improved significantly by applying it to the results of several building detection systems. Much work remains to be done, however. Analysis of the fusion results has revealed shortcomings in each of the building detection systems, and there are also a number of directions to pursue in terms of improving the intermediate and final fusions generated during the overall fusion process.

1. BABE produces two shadow thresholds, only one of which is used by SHAVE and SHADE. It may be the case that the other threshold more accurately reflects the shadow threshold for a given image, or perhaps some combination of the two may prove more effective. Experiments need to be performed in this area.

2. GROUPER is effective in clustering the fragmented hypotheses that are typically produced by BABE, but several of the grouped fragments do not correspond to building structure in the scene. Experimentation with disparity maps to refine these clusters is currently underway.

3. SHAVE's scoring system is simplistic and sometimes allows hypotheses with low shadow scores to pass as good hypotheses. Alternative scoring schemes might be explored.

4. SHADE's corner finding system can be improved. Work is currently underway on a method for iteratively approximating the location of corners in noisy lines by using an imperfect sequence finder to break lines at potential corners, and applying a gradient-based line evaluation function to score the breaks.

5. The fusion steps in the overall fusion process tend to increase the number of false positive pixels, and thresholding alone may not improve this without decreasing the number of correctly hypothesized pixels as well. The use of a refined disparity map, as well as the use of the original intensity image, may aid in eliminating false positive pixels from hypothesized regions in the final fusion. Alternatively, active contour models might be used to refine segmentations, using the fusion segmentations (possibly thresholded) as the initial seed to the process.

6. Another interesting application of this fusion technique would be on binocular imagery. One could imagine merging hypotheses from the left and right images of a stereo pair to obtain an improved interpretation of a scene, since it is likely that the left and right hypothesis sets would differ due to changes in image perspective. Experiments are underway in this area.

109

A more general question concerns the effectiveness of simple fusion approaches such as the one described here. Certainly, one can envision other approaches for combining building hypotheses that would make use of *a priori* information about the systems producing the hypotheses to produce meaningful fusions of the individual hypotheses. It is unclear, however, whether such approaches would ultimately benefit from the additional complexity required to take advantage of such knowledge. Although the results at this stage are rough, the fusion method developed here appears to be a simple and effective means for increasing the building detection rate for a scene, and may eventually provide a means for incorporating several sources of photometric information into a single interpretation of the scene.

6. Acknowledgments

References

[1]   Aviad, Z.
        Locating Corners in Noisy Curves by Delineating Imperfect Sequences.
        Technical Report CMU-CS-88-199, Carnegie-Mellon University, December, 1988.

[2]   Aviad, Z., McKeown, D. M., Hsieh, Y.
        The Generation of Building Hypotheses From Monocular Views.
        Technical Report, Carnegie-Mellon University, 1990.
        to appear.

[3]   Fua, P., Hanson, A. J.
        Resegmentation Using Generic Shape: Locating General Cultural Objects.
        Technical Report, Artificial Intelligence Center, SRI International, May, 1986.

[4]   Huertas, A. and Nevatia, R.
        Detecting Buildings in Aerial Images.
        Computer Vision, Graphics, and Image Processing 41:131-152, April, 1988.

[5]   R. B. Irvin and D. M. McKeown.
        Methods for exploiting the relationship between buildings and their shadows in aerial
            imagery.
        IEEE Transactions on Systems, Man and Cybernetics 19(6):1564-1575, November, 1989.

[6]   Mohan, R., Nevatia, R.
        Perceptual Grouping for the Detection and Description of Structures in Aerial Images.
        In *Proceedings: DARPA Image Understanding Workshop, April 1988*, pages 512-526.
            April, 1988.

[7]   Nicolin, B., and Gabler, R.
        A Knowledge-Based System for the Analysis of Aerial Images.
        IEEE Transactions on Geoscience and Remote Sensing GE-25(3):317-329, May, 1987.

110

A Computer Vision System for the Recognition of Trees in Aerial Photographs

Axel J. Pinz
Institute of Surveying and Remote Sensing
University of Agriculture
Peter Jordan Str. 82
A-1190 Wien, Austria

Abstract: Increasing problems of forest damage in Central Europe set the demand for an appropriate forest damage assessment tool. In this paper the *Vision Expert System VES* is presented. VES is capable of finding trees in color infrared aerial photographs - this is the first step towards an automatic forest damage interpretation system. Concept and architecture of VES are discussed briefly. The system is applied to a multisource test data set. The processing of this multisource data set leads to a multiple interpretation result for one scene. An integration of these results will provide a better scene description by the vision system. This is achieved by an implementation of Steven's correlation algorithm.

Key words: Aerial image understanding, image understanding, knowledge-based image analysis, frames, object representation for computer vision systems, dot pattern correlation

1 INTRODUCTION

1.1 Forest damage interpretation

During the past years research concerning the assessment of forest damage using color infrared aerial photographs was done at IVF. *IVF* stands for "*I*nstitut für *V*ermessungswesen und *F*ernerkundung" - the Institute of Surveying and Remote Sensing at the University of Agriculture in Vienna. The benefits of color infrared aerial photographs for the interpretation of vegetation are discussed in detail in [Sch89]. However, to be able to understand the method described in this paper, the reader should be familiar with a few details.

The condition of a tree is evaluated by interpreting the color of its crown in a color infrared aerial photograph. Since, compared to damaged vegetation, healthy vegetation tends to reflect more light in the infrared band and less in the red one (see Fig. 1.1), *healthy trees look red* in a color infrared photograph, while bad trees will have less red and more green color, thus appearing pale. But the color of a tree will depend on both the tree's vitality and the *tree species*. For example, a healthy pine will show a color similar to the one of a damaged spruce.

In many parts of Central Europe a very intensive and heterogeneous kind of landuse takes place. From the forest damage interpretation point of view this means, that normally many different kinds of trees will be found within one forest stand. Also, the condition of the trees in a stand

may vary significantly. In a typical Austrian forest it is quite common to find a pine by the side of a spruce and to find a healthy tree close to a very bad one. As a consequence, to get correct results of a "forest-condition-inventory", as it is called in Austria, it is necessary, to interpret *the species and the color of the single tree*. Trying to use remote sensing methods for this forest-inventory, data from satellites like LANDSAT or SPOT are not convenient, only aerial photographs will provide sufficient spatial resolution.
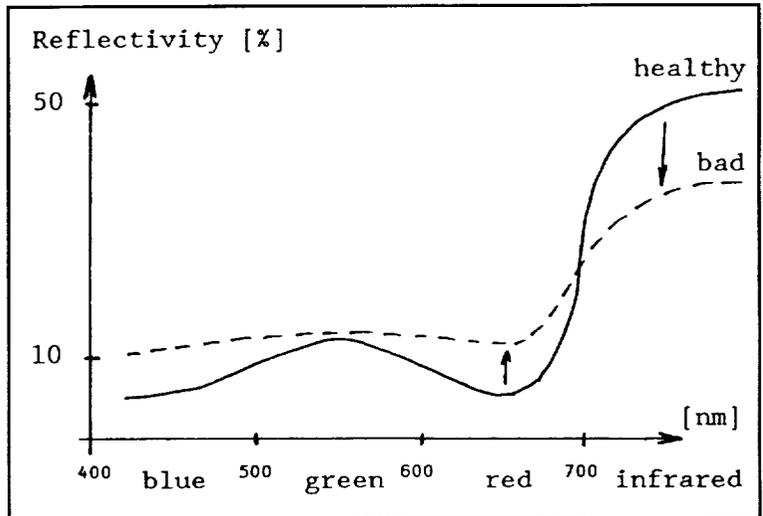

Fig. 1.1   Reflectivity of vegetation (in principle)

Interpreting color infrared aerial photographs for forest inventory purposes therefore calls for the following procedure:

1. Find a tree in the aerial photograph.
2. Determine the tree species.
3. Determine the tree vitality by interpretation of the color (and the texture) of the tree.

In this paper we discuss the problem of finding trees in aerial photographs (1.) by means of computer vision. While the color information is required for the determination of species and condition of a tree (2. and 3.), tree-finding can be done using a monochrome image. Therefore in this paper only monochrome images are shown. They were produced by averaging the three color channels of a color infrared image.

1.2 A tree finding computer vision system

In addition to the task of finding trees the application of a computer vision system will be extended to serve for several remote sensing tasks at IVF. For this purpose an image understanding system - the *V*ision *E*xpert *S*ystem *VES* - was built. The architecture of VES has already been presented in detail in [Pin88] and [Pin89]. The system therefore will be discussed very briefly in chapter 2.

Fig. 1.2   Original image   **Fig. 1.3**   VES result
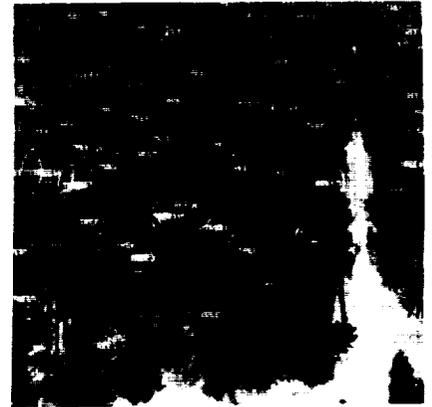
Figures 1.2 and 1.3 show the result of VES processing a typical test-image. The scale of the image was 1:4000 and it was digitized with a pixel size of 25μm. The digital image was 512x512 pixels (Fig. 1.2) and VES found 169 circular image objects from which 70 scene objects were derived (Fig. 1.3).

There were several problems encountered in the course of this first application of VES. First of all, the pixel scale was unrealistic - 1 pixel represented a square of 10cm^2 in the scene. Second, the system was very slow due to an inadequate hardware component. Third, the experience with the system led to more sophisticated ideas about representation and about the evaluation of the interpretation result.

As a consequence, a successor system of VES - the *Vision Station VS* - is currently under development at IVF. In a first step the VES functionality was ported to VS. Due to the better performance of VS most of the "VES-results" presented in this paper were done on the VS simulating a VES-behaviour.

At this point the evaluation problem should be discussed in more detail. A computer vision system starts with a given image and a problem specification (e.g. "find trees"). As the process of automatic image interpretation proceeds, a *scene description* begins to emerge. In the case of VES this is a two-stage process. At first *image objects* are found. Then some of them are put into relation to a certain *scene object*. There are several control strategies for vision systems: top-down, bottom-up and bidirectional (Fig. 1.4). The features of each of these strategies were discussed by Matsuyama [Mat87]. He and many others (e.g. [Hav83], [Keo85], [Pin89], [Nag80]) tried to avoid the problem of combinatorial explosion of the search size in a bidirectional system by using search space limiting control structures (either top-down/bottom-up or other limiting techniques in a bidirectional system). Besides these "conventional" approaches there have been more recent efforts to find other control mechanisms (e.g. Matsuyama's hyper-graph [Mat88] or Burt's pattern tree [Bur88]). However, for a conventional system it is crucial to be able to *evaluate the interpretation results*. In VES and VS we try to calculate a quality value for each object. This helps in discarding of very uncertain objects. But these quality value calcu-

Fig. 1.4   Control strategies

lations sometimes are imprecise themselves and the crucial questions still remain: Is the result correct? Is the result complete? Are there still objects missing? Can the interpretation process be terminated? As a conclusion, any additional source helping to improve the quality assessment should be used. In this paper we will investigate the *use of multisource data to gain a more robust scene-description*.

1.3 The test data set

The test data set is shown in Fig. 1.5. It consists of five aerial images taken at April 15, 1984 (images a. - d.) and August 23, 1984 (image e.). There are four different scales: 1:32000 (a.), 1:16000 (b.), 1:8000 (c.) and 1:4000 (d. and e.). These aerial images originally were taken to investigate the abilities of human interpreters. It turned out that while it is still possible to locate a tree in the 1:32000 image, the correct determination of tree species and tree vitality calls for a scale of about 1:12000 - 1:15000 (this will also depend on the selected film type and on the exposure and development conditions) [Sch89].
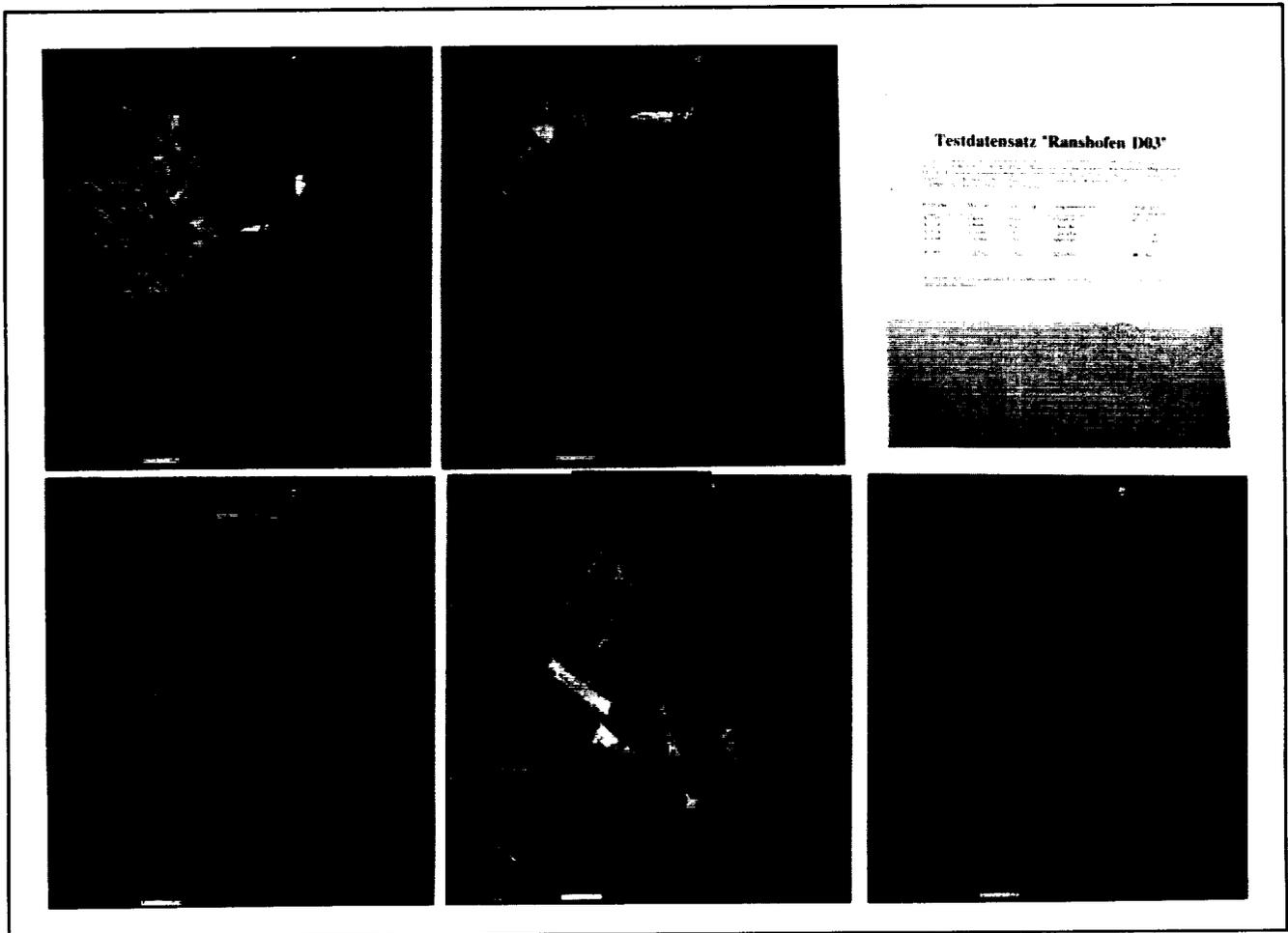
a. spring 1:32000   b. spring 1:16000
c. spring 1:8000   d. spring 1:4000   e. summer 1:4000

Fig. 1.5   The test data set "Ranshofen D03"

Small portions of these five images, *each showing the same part of the scene*, were digitized with 25μm (a. and b.), 50μm (c.) and 100μm (d. and e.) pixel size. This lead to a pixel scale of approximately 40cm in the scene (b. - e.) and 80cm in the case of a.. We plan to use this data set for several purposes. We want to investigate resolution-dependent performance variations in automatic tree detection and species interpretation [Bis89], [Pin90]. The data set also supplies *different views* (in space and time) of the same objects. It is therefore expected to get a *more robust scene description* by proper combination of results from several images.

1.4 Related work

Aerial image analysis has always been a major field of application for model based vision systems. Most of them were concerned with finding *artificial, man-made objects*. McKeown et al. present a rule-based approach in the system SPAM [Keo85]. Several systems were developed by Matsuyama (e.g. ACRONYM, SIGMA, LLVE) [Mat87]. He used frames and he examined the three "classical" control strategies bottom-up, top-down and bidirectional. VES also uses frames, which were introduced by Minsky as a proper form of representation for vision tasks [Min75]. In

the Mapsee2 system the similar concept of schemas was used for knowledge representation [Hav83]. In our Vision Station the representation of objects is based on the Common Lisp Object System CLOS [Bob88]. More recent work (e.g. Burt's pattern tree [Bur88], Matsuyama's multilayered hypergraph [Mat88]) deals with hierarchical (pyramid) control structures, trying to avoid the drawbacks of top-down, bottom-up or bidirectional. Earlier work includes the VISIONS-System [Han78a],[Han78b] and a system by Nagao and Matsuyama [Nag80].

Most computer vision systems use a kind of modeling mechanism. There are object models in the scene domain (3D) and image objects (2D). Image objects are found during the interpretation process, thus being individual (vs. generic) objects. One can distinguish between the four object classes discussed in detail below (see 2.3: scene/image, generic/individual). In comparison to other systems, where a border between two classes may be missing or implicitly defined (see e.g.: discussion of the importance of discriminating between image level and scene level information [Mat87], short vs. long term memory in VISIONS [Han78b]), there is an exact definition of all four classes in VES. This object representation scheme is in fact controlling most of the VES-processes.

A complete computational model is given by Marr [Mar82]. Viewing our results as "place-tokens" in the sense of Marr, we found a structure similar to Glass patterns [Gla69] and we tried to correlate the results from different images using Steven's algorithm [Ste78]. Several mathematical models were developed to describe the phenomenon of orientation perception in random dot patterns [Mat90].

Dealing with the problem of the *interpretation of natural (vs. man-made) scenes*, the effort is often directed towards a complete segmentation of the image (e.g. [Oht85], [Naz84]). Related work concerning the application of finding trees in aerial photographs was done by Haenel et al. [Hae87]. While he developed very specific algorithms for this task, we try to establish a more universal vision system. Supplied with proper knowledge, VES and VS will be able to solve many other perceptual tasks in remote sensing.

2 THE VISION EXPERT SYSTEM VES

There were several major goals in the development of VES. The system architecture should be open and flexible. VES should be appropriate for a broad field of applications and experiments. The resulting complicated framework was then filled with knowledge and methods for the specific problem domain of finding trees. This was the first application test of VES.

2.1 Architecture and implementation

The claimed universality of the system together with the available hard- and software at IVF led to a hybrid architecture. The system consists of a host computer and an image processing system. While under VES both the image processing software and the LISP-system is run on the same host, in the VS-environment the LISP-part is done on a seperate workstation. This is shown by the
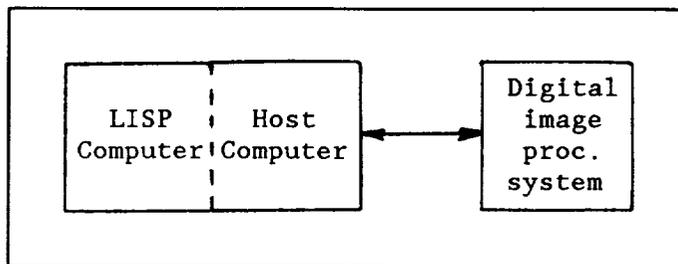

Fig. 2.1   Hardware components

115

dashed line in Fig. 2.1. The interaction between the software components is illustrated by Fig. 2.2.

VES is organized as a top-down strategy vision system with the possibility of being extended to a bidirectional system in the future. Core part of the system is the object representation in frames. VES is implemented in INTERLISP. The frame representation language FRL was used as a basis for the VES frames [Rob77]. Most of the digital image processing modules are written in PASCAL.
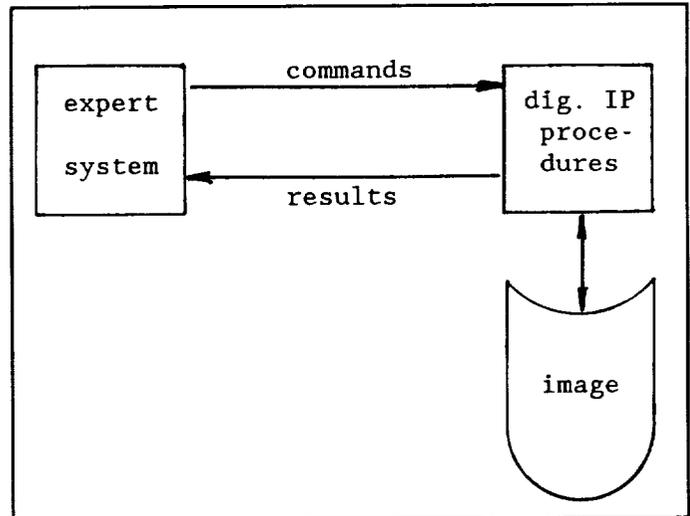
2.2 The VES frames

Fig. 2.2   Interaction of software components

With the exception of two rules all the explicit knowledge is stored in frames. There are *object-*, *method-* and *procedure-frames*. The frames are interconnected by various relations (e.g. ako/instance, part/whole, represents/rep-by) thus forming groups of several semantic networks.

If there is knowledge about how to find a certain object, then the slot METHLIST of this object's object-frame contains a list of applicable methods, each element pointing at a method-frame. When a method is selected and applied the result usually is a sequence of processes. Some of them will be LISP-functions, others are image processing modules. The interface between LISP and the image processing modules is handled by the procedure-frames. They contain information about the calling sequence, parameters and resulting effects of an image processing module.

2.3 Object representation

We distinguish between *scene objects* (OBSC) and *image objects* (OBIM) on the one hand and between *generic objects* and *individual objects* on the other hand. While the latter (CLASSIFICATION GENERIC or INDIVIDUAL) are a standard feature of FRL to separate models from instances, the distinction between scene- and image-objects is quite common for a computer vision system. In Fig. 2.3 the regions A and B represent the system's initial knowledge before an interpretation is started ("static knowledge") - the models for scene objects and models for image objects. Regions C and D constitute the "dynamic knowledge" about the interpreted scene. During the process of image interpretation, at first individual image objects are found (region D), later instances for corresponding individual scene objects are established (region C). From the VES point of view, region C is the result of a successful image interpretation: it contains all scene objects which the system has found in an image taken from a certain scene. This is a description by objects, not a segmentation of the image. Normally the objects don't cover all of the area of the image. During the course of an interpretation process, the system will try possible relations between hypotheses for scene objects and already-found image objects. It will end up with the best relation which finally constitutes the correct interpretation for the image object.

Fig. 2.4 gives an example of an interpretation situation. The world is divided into scene- and image-objects. An individual scene object (pine0) was found - pine0 is a pine, a tree and a scene

object. It is represented in the image by circle8. Circle8 is an individual circle, an area (vs. point or line) and an image object. It currently represents the scene object pine0.

A scene objects OBSC	image objects OBIM B
(AKO ($VALUE (OBSC))) (CLASSIFICATION ($VALUE (GENERIC))) **generic objects**	(AKO ($VALUE (OBIM))) (CLASSIFICATION ($VALUE (INDIVIDUAL)))
individual objects (AKO ($VALUE (OBSC))) (CLASSIFICATION ($VALUE (INDIVIDUAL))) C	(AKO ($VALUE (OBIM))) (CLASSIFICATION ($VALUE (INDIVIDUAL))) D

Fig. 2.3   The four different object classes of VES

2.4 Control of the interpretation process

The interpretation process is always invoked by the search for an object. A valid object must be represented in a generic frame. Correct search commands might be:

(FIND '(TREE))	...	find trees,
(FIND '(TREE ROAD))	...	find trees and roads,
(FIND '(CIRCLE))	...	find circles (image objects).

After an initialization phase (loading and establishing of global parameters like name of the image, scale, etc.) the system grasps the frame representing the object being searched for and the top-down search process begins. The methods found in the slot METHLIST are evaluated and the best method is chosen. While the search for image objects yields individual image objects, the search for scene objects forces the search for corresponding image objects. For example, "find tree" or "find road" might invoke "find circle" or "find line". If image objects are found they must survive object-specific tests which are also stored in the method frame. Next, a scene object is generated and the corresponding relations between scene- and image-object are set. A method may also contain tests for scene objects. If a test fails, the scene object will be removed while the image object remains. This completes a top-down process. A list of individual objects which are all instances of the generic object that had been searched for was produced.

Two rules extend this pure top-down strategy. VES is trying to improve the interpretation by applying these rules again and again, until no rule fires any more, thus finishing the complete interpretation process.

Rule 1:   If there are "tunable" parameters for an object being searched for, try to vary one parameter and repeat the search.

Rule 2:   If an object being searched for is known to have "contrary" objects, then extend the search to these objects and check if a conflict occurs.

OBSC

OBSC
```
(CLASSIFICATION
        ($VALUE (GENERIC)))
(INSTANCE
        ($VALUE (... TREE ...)))
```

TREE
```
(AKO
        ($VALUE (OBSC)))
(CLASSIFICATION
        ($VALUE (GENERIC)))
(INSTANCE
        ($VALUE (... PINE ...)))
```

PINE
```
(AKO
        ($VALUE (TREE)))
(CLASSIFICATION
        ($VALUE (GENERIC)))
(INSTANCE
        ($VALUE (... PINE0 ...)))
```

OBIM

OBIM
```
(CLASSIFICATION
        ($VALUE (GENERIC)))
(INSTANCE
($VALUE (... AREA ...)))
```

AREA
```
(AKO
        ($VALUE (OBIM)))
(CLASSIFICATION
        ($VALUE (GENERIC)))
(INSTANCE
($VALUE (... CIRCLE ...)))
```

CIRCLE
```
(AKO
        ($VALUE (AREA)))
(CLASSIFICATION
        ($VALUE (GENERIC)))
(INSTANCE
        ($VALUE (... CIRCLE8 ...)))
```

generic objects

individual objects

PINE0
```
(AKO
        ($VALUE (PINE)))
(CLASSIFICATION
        ($VALUE (INDIVIDUAL)))
(REP_BY
        ($VALUE (CIRCLE8)))
```

CIRCLE8
```
(AKO
        ($VALUE (CIRCLE)))
(CLASSIFICATION
        ($VALUE (INDIVIDUAL)))
(REPRESENTS
        ($VALUE (PINE0)))
```

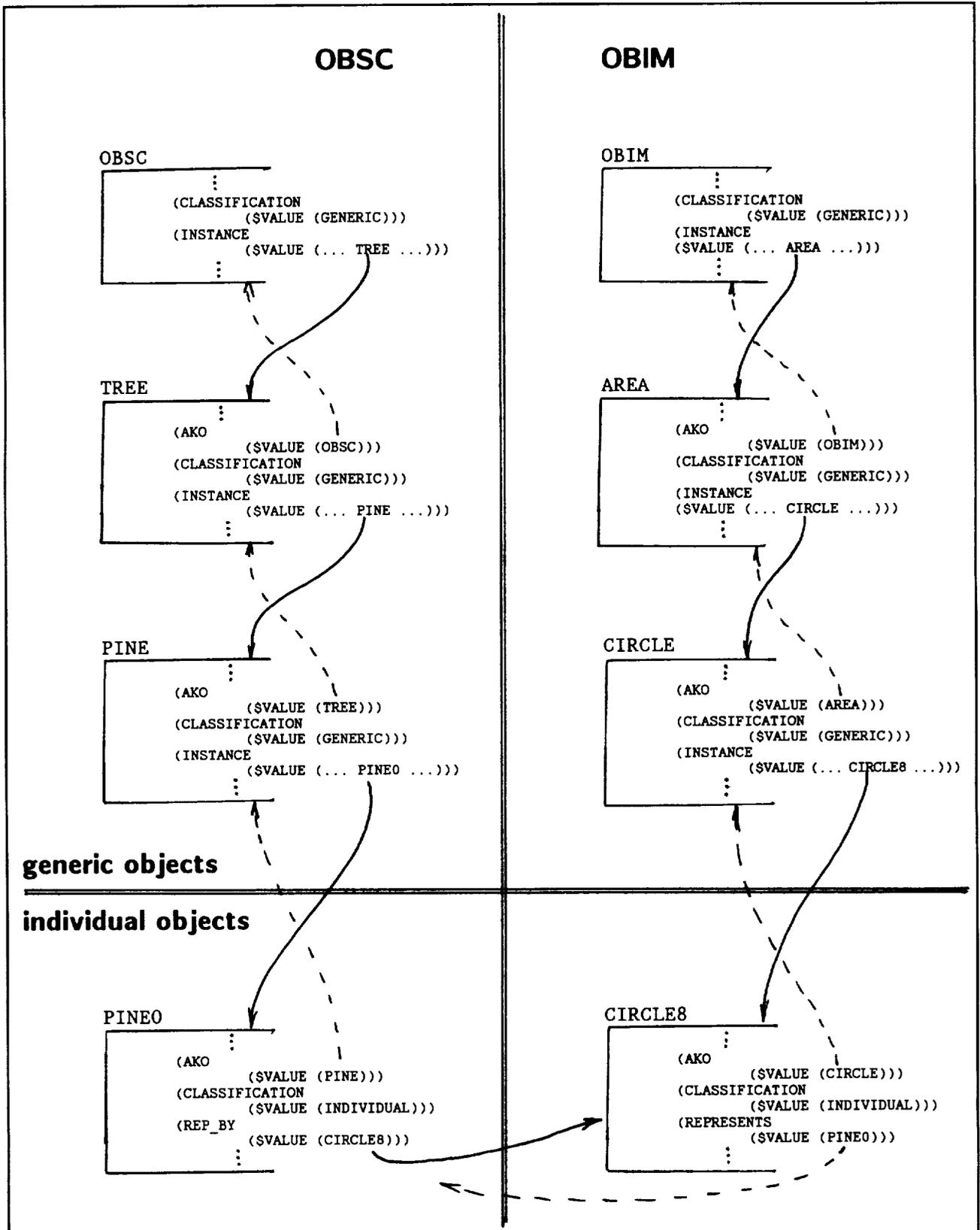**Fig. 2.4**   An example of scene and image objects

118

3 DISCUSSING VES PROCESSES

In this chapter the processes and methods which were implemented to recognize trees in aerial photos are discussed. Fig. 3.1 displays a very simplified scheme of the processes in VES. Starting with the task (usually entered by the user) of finding a certain scene object OBSC, the search for a corresponding image object OBIM is initiated. Image objects are found and connected with scene objects, thus finishing one top-down process. Application of global rules leads to several repetitions until no rule is applicable any more. The corresponding up-arrow in Fig. 3.1 is marked with a dashed line because it is also possible to request one single top-down process without application of global rules.

Search for OBSC ⟶ Search for OBIM

Parametertuning, resolving conflicts

OBIM found

OBSC found

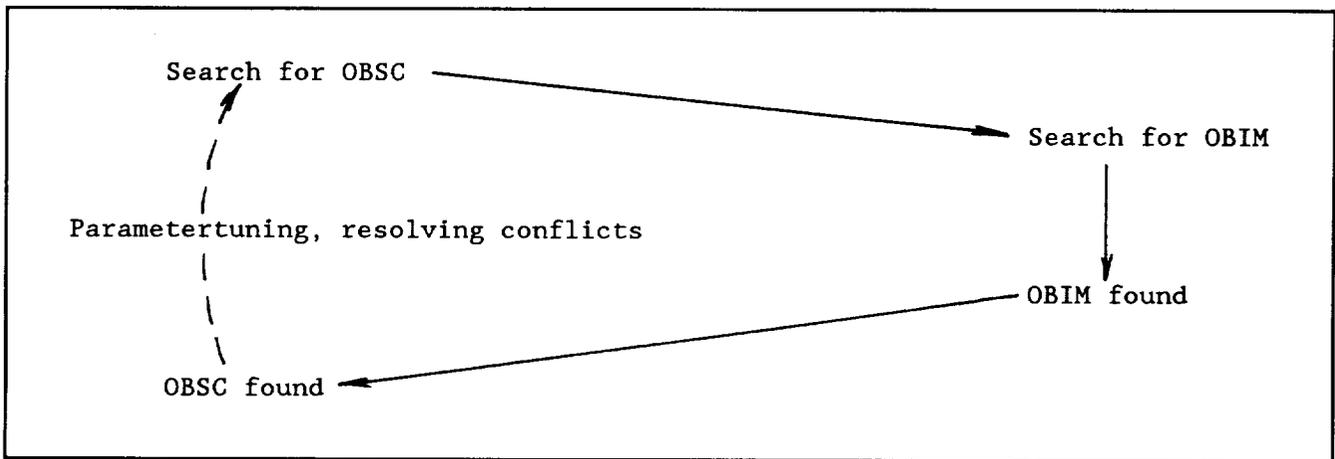**Fig. 3.1**   VES processes

After the initialization phase, VES is ready to accept search commands. One top-down process is started by

(FIND '(TREE))   ...   find trees.

A complete process, including multiple repetitions by application of the global rules, is invoked by

(START '(TREE)).

VES finds the method METH0 in the slot METHLIST of the frame TREE. METH0 assumes trees to appear as bright circularly shaped image objects. This assumption holds for trees inside a forest and is a very good assumption to make in the central parts of an aerial photo where objects are viewed from above. Towards the edges of the photograph, the direction of view is changing, e.g. a spruce appearing not circular but triangular in shape. At first METH0 is searching for bright circular image objects, next, every circle is assigned to an individual scene object "tree". This is followed by a test. If two trees are standing too close to one another, the tree with the larger radius is removed.

The application of METH0 automatically invokes the new task of

(FIND '(CIRCLE))   ...   find circles.

The structure of the frame CIRCLE is similar to the one of TREE. The method METH1, searching for bright circular image-objects in a stepwise process, is found in the slot METHLIST.

A bright circular object may be viewed as a local maximum of brightness in the image. Usually there will be a lot of texture information found within a tree's crown. This would lead to many local maxima within one crown. Therefore, a lowpass filter must be applied before the search for local maxima can take place.

The original black and white image (it was produced by averaging the 3 channels of a color infrared image) is the input to METH1. Lowpass filtering is achieved by a local window operation using the image processing system. The size of the window (the "size" of the lowpass) is calculated from the image's scale and the expected size of the searched object (radius of the tree's crown = radius of circle). Next to the lowpass filtering the local maxima are searched for. Because of the preceding lowpass filtering, a local maximum usually covers an area of pixels of equal brightness. The center of gravity of each area is taken as the exact location of the local maximum.

In the final step METH1 checks the found object for circular shape by inspecting the "radial brightness distribution". This distribution is obtained by drawing concentric circles around the maximum's position, summing up all pixels lying on a circle and taking the average (see Fig. 3.2). For a circular object the resulting diagram (mean brightness / radius) should show a distribution as in Fig. 3.2. The module which is computing the radial brightness distribution to decide whether the object is circular needs the following three input parameters: smallest radius, largest radius and minimum brightness decrease (the mean brightness has to be n% lower at the edge of the object than at its center). It turned out, that the necessary brightness decrease n is scale-dependent. In images of a scale of 1:4000 a good value for n was 35 - 40 %, while n had to be reduced to 30 % for scales of about 1:8000. The module returns either the radius of the found circular object at which this minimum decrease is reached or NIL, if any of the above three conditions do not hold.
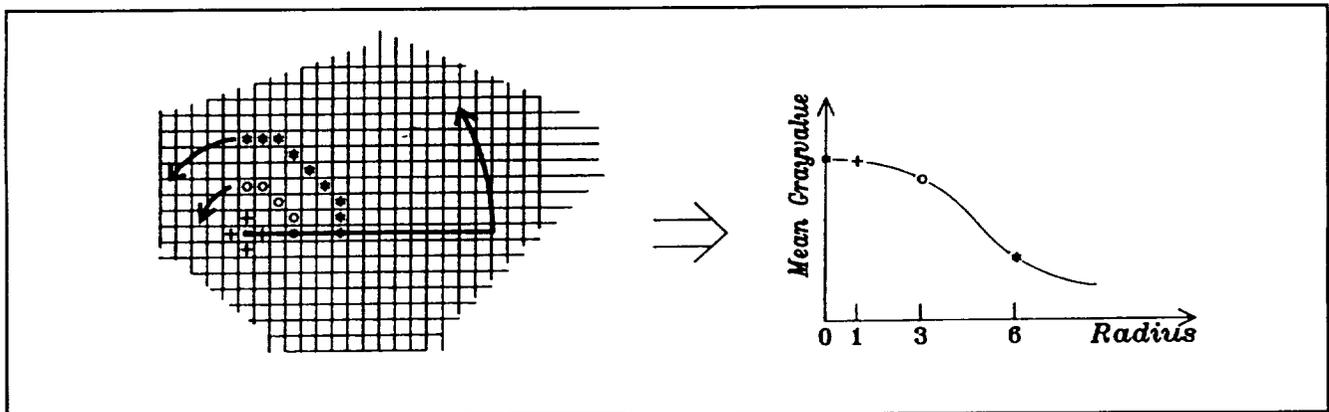

Fig. 3.2   Radial brightness distribution

This completes one top-down shot. The two main stages are shown in Fig. 3.3 and Fig. 3.4 (the original image is Fig. 1.2). Fig. 3.3 shows the lowpass filter (in this case a 25x25 window lowpass was selected by VES) together with the local maxima. Fig. 3.4 shows the corresponding circles that survived the "radial brightness distribution" test. Each of these circles is assigned to a scene object (tree). Some of the trees are removed by the final test in METH0 (if standing too close).

If the interpretation process is started by (START '(TREE)), the global rules will be applied. The parameter variation will produce two more lowpass filters and this will result in new local

maxima, circles and trees. The search for contrary objects (in this test case a road was entered manually) leads to the elimination of trees that would grow in the middle of a road. The final result shown in Fig. 1.3 was obtained after two parameter variations (19x19 and 31x31 lowpass window).
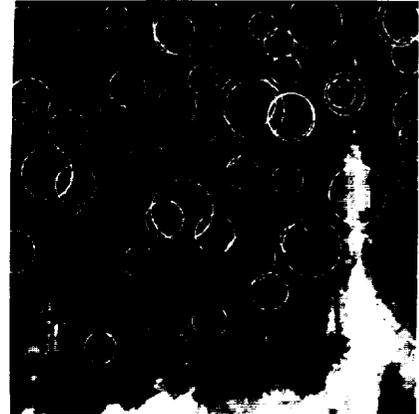
Fig. 3.3   Local maxima   Fig. 3.4   Circles from Fig. 3.3

4 PROCESSING THE TEST DATA SET

We took a small portion of each of the five images Fig. 1.5 a. - e. each showing approximately the same part of the scene. The size of these portions is 512x512 pixels (b. - e.) and 256x256 pixels (a.). All five images were processed with the standard VES tree-search (search for a default crown radius of 2,5m followed by two parameter variations (1,25m and 5m)). The original 512x512

Fig 4.1 512^2 portion of d.   Fig. 4.2 512^2 portion of c.   Fig. 4.3   Circles from Fig. 4.2

Fig. 4.4 128^2 portion of d.   Fig. 4.5 128^2 portion of c.   Fig 4.6   A correlation result

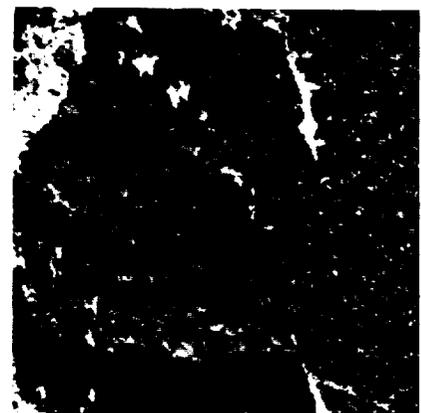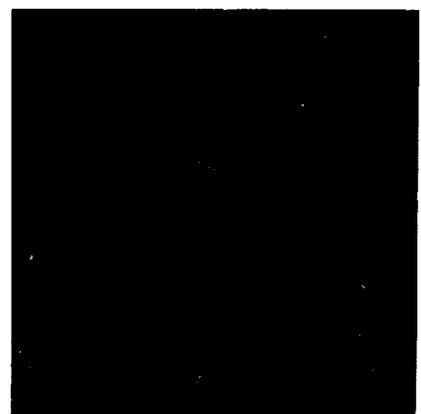images are shown for d. (Fig. 4.1) and c. (Fig. 4.2). Fig. 4.3 shows the circles found in Fig. 4.2 after the first top-down process. The final results (trees found) are shown in detail for 128x128 portions of d. (Fig. 4.4) and c. (Fig. 4.5).

The results of this experiment were very interesting: While even in the worst case (1:32000, image a.) many of the large crowns were detected, there was no "perfect" interpretation in any of the five cases a. - e.. Of course, the best results were obtained for the larger scales (c. - e.). But in each result there were several trees missing that were found in another case. The same is true for erroneous artifacts, which don't show up in more than one result at the same location. As a conclusion - the desired result of the interpretation of the whole data set (a. - e.) would be a careful *combination of the several results*. And, working with "intelligent" vision systems, we would favour a robust solution that doesn't require too precise and detailed instructions, similar to the ability of a person to identify the same tree in two different images.

As a first step towards this goal we tried the following procedure. We generated dot images of the five results. For each tree a dot mark located at the center of the circle representing the tree was produced. When two different results were overlayed and displayed in different colors, the resulting image was very similar to the dot patterns described by Glass [Gla69] and Stevens [Ste78]. In our case the patterns of one result may be converted to another one by assuming a superimposition of translation, rotation and a small change of scale. The remaining "noise" is caused by the individual height of each tree, and by the different position of the sun and viewing position for each image. In addition, due to the imperfect interpretation, some points are missing or added in the other image. Stevens called his patterns "Glass patterns" and he developed a *local algorithm* for the correct correlation of associated points. We implemented Steven's algorithm and tested it on the dot images generated from the interpretation results of a. - e.. One result of a correlation between the two images shown in Fig. 4.4 and Fig. 4.5 is shown in Fig. 4.6.

The results of this experiment were imperfect but very promising. Taken alone, Steven's algorithm is not effective enough for our patterns. This is due to the noise effects discussed above and due to the occurence of rather large point displacements. The algorithm will have to be adapted for our purposes - there are already several ideas for improvements. When viewed as one component of a larger vision system, even the actual performance of the algorithm is valuable. The correlation results will be processed by VES. Several heuristics may be applied, e.g. the fact that correlated trees should be of similar size. The correlation should also hold for more than two of the results (a. - e.). If there is a component in the system, that is able to determine the tree species [Bis89,Pin90], then correlated trees must have the same species. Current research at IVF is addressing these topics.

5 CONCLUDING REMARKS

It has been shown, that the use of multisource data can improve the quality and robustness of the interpretation result of a computer vision system. While synergic effects of this kind are well known, the proposed approach is also robust from another point of view. We do not need the geometric rectification of our multisource data to compare them. We also don't need complete or very accurate correlation results. The system is able of comparing two objects from two scenes just like a human interpreter looking at the two images. In a way the knowledge of a system like VES may be viewed as an alternate data source itself.

Many problems were discussed only very briefly or not at all. The ideas about the representation of objects, processes and methods in VES are improved in the VS environment. This representation problem is closely coupled with the problem of control of the interpretation process. Methods like the one described above can help in getting a better assessment of the current interpretation result. Dealing with multisource data, the representation problem becomes even more difficult: While there is one individual object, there can be several scenes (several scene objects) and many images (many image objects). Furthermore we believe that a good approach for a vision system in a natural environment should be rather different from the one in a man-made environment. Fuzziness in shape and morphology of natural objects has to be reflected in fuzzy and robust models and methods.

ACKNOWLEDGEMENTS

REFERENCES

[Bob88]   Bobrow D., et al., "Common Lisp Object System Specification", Draft submitted to X3J13, 1988

[Bis89]   Bischof, H. and Pinz, A., "The Use of Neural Networks for the Recognition of Tree Species in Digital Images", in Pinz, A. (ed.): "Knowledge-based Pattern Recognition", OCG Schriftenreihe 49, Oldenbourg 1989, (in German)

[Bur88]   Burt, P.J., "Attention Mechanisms for Vision in a Dynamic World", Proc. 9.ICPR, Rome, 1988, pp.977-987

[Gla69]   Glass, L., "Moiré Effect from Random Dots", Nature, Vol.223, pp.578-580, Aug. 1969

[Han78a] Hanson A.R. and Riseman E.M. (eds.), "Computer Vision Systems", Academic Press, Orlando, 1978

[Han78b] Hanson A.R. and Riseman E.M., "VISIONS: A Computer System for Interpreting Scenes", in Hanson, Riseman (eds.), "Computer Vision Systems", Academic Press, Orlando, 1978

[Hae87]   Haenel S., Tränkner H. and Eckstein W., "Automatic Detection of Tree Crowns in Aerial Photographs: The Path through the Bottleneck", in 2.DFVLR-Statusseminar, Oberpfaffenhofen, 1987, (in German)

[Hav83]   Havens W. and Macksworth A., "Representing Knowledge of the Visual World", IEEE Computer, Aug. 1983, pp.90-96

[Mar82] Marr, D., "VISION", Freeman, 1982

[Mat87] Matsuyama, T., "Knowledge-Based Aerial Image Understanding Systems and Expert Systems for Image Processing", IEEE Trans. Geoscience and Remote Sensing, Vol. GE-25, No.3, May 1987, pp.305-316

[Mat88] Matsuyama, T., "Expert Systems for Image Processing - Knowledge-Based Composition of Image Analysis Processes -", Proc. 9.ICPR, Rome, 1988, pp.125-133

[Mat90] Mates J., Lánsky P. and Yakimoff N., "Models for the Perception of Orientation in Random Dot Patterns", Biol. Cybern. 63, pp.71-80, Springer 1990

[Keo85] McKeown D.M., Wilson A.H. and McDermott J., "Rule-Based Interpretation of Aerial Imagery", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-7, No.5, Sept. 1985, pp.570-585

[Min75] Minsky, M., "A framework for representing knowledge", in Winston, P.H. (ed), "The Psychology of Computer Vision", Mc Graw Hill, 1975

[Nag80] Nagao M. and Matsuyama T., "A Structural Analysis of Complex Aerial Photographs", Plenum Press, New York, 1980

[Naz84] Nazif A.M. and Levine M.D., "Low Level Image Segmentation: An Expert System", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-6, No.5, Sept. 1984, pp.555-577

[Oht85] Ohta Y., "Knowledge-Based Interpretation of Outdoor Natural Color Scenes", Pitman publishing, London, 1985

[Pin88] Pinz, A., "An Image-Understanding Expert System for the Recognition of Trees in Color-Infrared Aerial Photographs", Ph.D. Thesis, TU Wien, Okt.1988, (in German)

[Pin89] Pinz, A., "Final Results of the Vision Expert System VES", in Pinz, A. (ed.): "Knowledge-Based Pattern Recognition", OCG Schriftenreihe, Oldenbourg 1989

[Pin90] Pinz A. and Bischof H., "Constructing a Neural Network for the Interpretation of the Species of Trees in Aerial Photographs", Proc. 10.ICPR, Atlantic City, IEEE 1990

[Rob77] Roberts, Goldstein, "The FRL Primer, The FRL Manual", MIT Memos 408 and 409, MIT, 1977

[Sch89] Schneider, W., "Using Remote Sensing for Forest-Inventory; Methods, Possibilities and Limits", FBVA-Berichte, Sonderheft, Wien, 1989, (in German)

[Ste78] Stevens, K.A., "Computation of Locally Parallel Structure", Biol. Cybern. 29, pp.19-28, Spiringer 1978

Visualizing Characteristics of Ocean Data Collected During the Shuttle Imaging Radar-B Experiment

David G. Tilley
The Johns Hopkins University
Applied Physics Laboratory
Laurel, Maryland 20723-6099

Abstract

Topographic measurements of sea surface elevation collected by the Surface Contour Radar (SCR) during NASA's Shuttle Imaging Radar (SIR-B) experiment are plotted as three dimensional surface plots to observe wave height variance along the track of a P-3 aircraft. Ocean wave spectra were computed from rotating altimeter measurements acquired by the Radar Ocean Wave Spectrometer (ROWS) aboard the same NASA aircraft as it was flown under the space shuttle Challenger. Fourier power spectra computed from SIR-B synthetic aperture radar (SAR) images of the ocean are compared to ROWS surface wave spectra. Fourier inversion of SAR spectra, after subtraction of spectral noise and modeling of wave height modulations, yields topography similar to direct measurements made by the SCR. Visual perspectives on the SCR and SAR ocean data are compared, although for surface tracks differing somewhat in space and time, for wind generated wave fields observed off the coast of Chile in October of 1984. Threshold distinctions between surface elevation and texture modulations of SAR data are considered within the context of a dynamic statistical model of rough surface scattering. The result of these endeavors is insight as to the physical mechanisms governing the imaging of ocean waves with synthetic aperture radar.

Keywords: Doppler radar, ocean waves, image processing, computer graphics

Introduction

Remotely sensed earth science data offer the potential for monitoring global change in our environment. Large data sets now exist and much more information on the spectral properties of our oceans and atmospheres will be forthcoming in the next decade. Visualization is emerging as a scientific tool for investigating theoretical computer models of physical processes in relation to empirical data from a variety of sensors operating over a wide range of spatial and temporal scales. As an example, radar measurements of ocean wave height and slope along the ground track of airborne and spaceborne remote sensors are viewed as shaded surface perspectives to appreciate correlations in short-scale texture and long-scale sea state during the Shuttle Imaging Radar (SIR-B) experiment.

The NASA P-3 aircraft conducted underflights of the space shuttle Challenger as it approached the southwestern coast of Chile (55°S, 80°W) on each of 4 days, 9 October to 12 October 1984. As a result, directional surface wave spectra have been computed from data acquired by the Surface Contour Radar [SCR, Walsh et al., 1985] and Radar Ocean Wave Spectrometer [ROWS, Jackson et al.,

1985] for comparison with Fourier wave power spectra computed from Synthetic Aperture Radar [SAR, Beal et al., 1986] image data. The four day period was characterized by a significant wave height that varied from 1.7 m to 4.6 m. The lowest sea state occurred on 10 October when an actively growing wind driven system with a wavelength of about 80 m appeared from the northeast propagating approximately -30° from the look direction of the SAR. This data set is typical of a fresh steeply sloped sea state in which non-homogeneous and transient hydrodynamic modulations of backscatter influence the SAR Doppler imaging technique. The highest sea state occurred the following day of 11 October when the three radar remote sensors reached a consensus in measuring a well developed swell with a wavelength of 270 m from the northwest. On 12 October this wave field was observed with a wavelength of 380 m having diminished to a significant wave height of about 3.5 m and turned so as to propagate along the shuttle track about 90° from the SAR look direction. The radars also detected an apparently new wave system with a wavelength of about 140 m developing again from the northeast on 12 October, the last day of the aircraft underflights. This data set is of particular interest in modeling the along track and across track imaging properties of SAR as it responds to waveheight modulations of surface velocity and texture.

Homogeneous ocean wave fields induced by distant storms and imaged with SAR may be fast Fourier transformed to estimate directional wave power spectra for oceanographic applications. However, individual wave groups are not necessarily well characterized by the normal statistics of the spectral approach and might be better examined in speckle reduced SAR images with restored wave height significance. Hence, Fourier domain restoration and enhancement filters have been developed [Tilley, 1987] to apply what has been learned about the SAR ocean-imaging modulation transfer function in the spectral domain and derive estimates of surface elevation in the image domain. Speckle reduction is based on empirical methods for determining a broadband spectral noise floor that can be subtracted as the random influence of transient surface facets tilted toward the radar as it looks down at a 23° incidence angle upon a homogeneous rough surface. Such a spatially stationary distribution of backscattering facets has been used to estimate the SAR wavenumber response for the SIR-B remote sensor [Tilley, 1986] using data collected over Baie Missisquoi near Montreal, Canada on October 7, 1984. This stationary response function can be used in an inverse Fourier filtering operation to improve the broadband spectral response of the SAR data obtained off the coast of Chile a few days later.

Non-homogeneous rough surface scattering may well be deterministically related to waveheight via a hydrodynamic modulation theory that is not well understood at present. However, it is apparent that the SAR along track wavenumber response is limited by ocean dynamics and can be partially restored [Tilley, 1987] with an empirical model of surface motion blurring based on a stochastic distribution of backscattering events in the time required for Doppler image synthesis. After the empirically estimated stationary and dynamic response functions are applied to SAR spectra, a broadband power threshold is applied to separate wave signal from random noise. A theoretical model of surface tilt and velocity modulation [Monaldo, 1987] is then applied to restore wave height significance to the Fourier image power. Advances in SAR spectral processing techniques are required to improve remote sensor estimates of ocean wave height variance, including distributions over wavelength and propagation direction, for the variety of sea states that are of interest to oceanographers, ship captains and coastal authorities. The object of ocean research with SAR is to develop

theoretical descriptions of radar cross section modulations that can be parameterized by empirical analysis of Fourier statistical data. Fourier spectra can then be compared with ocean wave spectra, or inverse transformed and compared with ocean surface topography, to evaluate SAR methodology using non-Doppler radars (e.g., the ROWS and SCR) that make more direct measurements of geophysical surface statistics.

Synergy of Aircraft and Spacecraft Ocean Observations

The SAR, SCR and ROW data collected during the SIR-B experiment at the Chilean site have all been processed to yield directional ocean wave height variance spectra in common units of m^4. Intercomparisons in this format have been reported [Beal, 1987] for the purpose of designing future SAR systems and assessing their potential value to computer wave models. In general, the three radar remote sensors were able to reach a consensus for all the sea states encountered, although the fresh wind-driven sea with low wave height on 10 October appeared to be somewhat misrepresented by the SAR. Therefore, this SAR scene is the subject of continuing investigation to develop assessments of the various algorithms applied for signal detection, clutter suppression and restoration of wave height significance. Comparisons with ROWS spectral estimates are considered in terms of action variance, in units of m^2. Once the ROWS data have served to guide the SAR to its best spectral estimates of the wave field, an inverse Fourier transform is applied to recreate SAR scenes of the surface elevation. Comparison with SCR measurements of surface topography are made by computing wave height statistics over similar ocean sites and by computing three dimensional surface visualizations of non-homogeneous wave grouping at these sites.

On 12 October at the Chilean site, the ROWS spectra depicted in Figure 1a indicates a 380 m wavelength system, propagating nearly along its eastern flight direction, that appears spread at low power to a more southerly heading. A weaker 140 m wavelength system, propagating across the ROWS flight direction, is apparently detected near the instrument's signal-to-noise limit and may be confused with or the cause of the broadening observed for the dominant swell. Both of these wave systems are also detected by the SAR, as shown in Figure 1b, when only the empirical instrument response functions are applied to estimate the wave action variance spectrum. The space shuttle was also travelling along an eastern flight direction, so it is not surprising that the SAR, as well as the ROWS, is able to detect the weak wind driven wave system via surface tilt modulations of backscatter occurring in the across track direction. The SAR also observed the dominant swell wave system travelling along its flight direction. An image of surface wave height variance can be computed by an inverse Fourier transform of the SAR spectrum after a theoretical ocean imaging transfer function is applied to account for both surface tilt and velocity modulations of the backscattered field. The interaction of the long 380 m wavelength swell and the 140 m wavelength wind driven sea are represented in Figure 2a as a computer generated visualization of the surface elevation. A similar visualization is depicted in Figure 2b where the direct ranging measurement of surface topography is depicted over 3 SCR aircraft tracks, each 400 m wide, to simulate the same coverage as the spacecraft SAR.

On 10 October at the Chilean site, the ROWS aircraft and the space shuttle carrying the SAR were even more closely aligned along eastern flight directions. Both remote sensors detected an 80 m wavelength wind driven system propagating

at 60° and a 200 m wavelength swell propagating at 130° from their flight direction. The spectral amplitude of the wind driven sea dominated that of the longer swell for both the ROWS and SAR remote sensors after correction for their respective instrument response functions. The surface tilt and velocity modulations of radar cross section may not suffice to describe SAR ocean imaging when the non-homogeneous and transient seas violate the ergodic and stationary assumption of two-scale scattering models. Hence, the SAR image spectrum was inverse transformed both before and after the application of the ocean imaging transfer function traditionally used to restore wave height significance. The statistics of the two Fourier filtered SAR images are compared to SCR surface elevation statistics in Figure 3 over 6 km² ocean segments differing spatially by about 20 kilometers and temporally by about 2 hours. This data set also presents a unique opportunity to compare surface height and texture acting in hydrodynamic modulation of radar resonant wavelengths (i.e., 23 cm surface waves for the L-band SAR) by longer wind generated waves (i.e., the 80 m sea) with periods comparable to the scene integration time. The correlation properties of the surface elevation and texture are visualized as a shaded surface plot in Figure 4, assuming that the wave action spectrum is proportional to height variance without tilt and velocity bunching modulation.

Computing and Display Technology

The radar data presented herein were collected in 1984 and have been processed and displayed using image processing and computer graphics workstations that have evolved in several different departmental facilities. Initial development of the SAR Fourier filtering algorithms was accomplished with a PDP-11/70 minicomputer system purchased from Digital Equipment Corporation at the beginning of the decade. Ocean images scaled to 32-bits in intensity over 512 x 512 arrays of picture elements (pixels) could be fast Fourier transformed in about 13 minutes using an optimized mass storage algorithm to coordinate data transfers between the 64K word memory partition of the 16-bit computer and large magnetic disk peripherals. Fortran program code was developed to apply filtering algorithms to the complex spectral database prior to inverse Fourier transformation. About 1 hour was typically required to restore wave height significance to a SAR ocean scene.

A Comtal Vision One/20 image processor was interfaced to the PDP-11/70 in 1981 allowing DMA transfers of the 512 x 512 x 8 bit pixel scenes over a UNIBUS in about 1 second. This system was equipped with 2 Mbytes image memory and an LSI-11 microprocessor controlling a pipeline delivering up to 30 ocean scenes a second to a 512 x 512 x 24 bit color monitor. A Matrix camera, Model 4007, was also acquired and interfaced to R,G,B outputs from the color monitor. This unit can be used to expose 35 mm roll film or format from 1 to 25 images on 8" x 10" sheet film. Figures found herein were photographed with this image processing and display system which now stands alone receiving its inputs from 9-track magnetic tape.

The SAR and SCR surface plots in the figures were computed using the PV-WAVE software package developed and supported by Precision Visuals, Inc. Version 2.2 of PV-WAVE running on a DECstation 3100 workstation offers the algorithms for using SAR texture information to shade a surface plot of ocean wave height visualized from a number of elevation and rotation angle perspectives. PV-WAVE, Version 1.0, is also installed on a VAXstation 3500 interfaced via a Q-bus to a QUEN-16 wavefront array processor. This desktop processor is being developed

jointly by Interstate Electronic Corporation and The Johns Hopkins University Applied Physics Laboratory. Initial experimentation with the QUEN-16 have shown that a 512 x 512 x 32 bit two dimensional fast Fourier transform computes in about 20 seconds. Basic spectral filtering algorithms are now being programmed and tested on the QUEN/VAX system. Wave height perspectives will be computed from SAR image data in minutes, rather than hours, allowing experimentation with new Fourier filtering algorithms. Larger ocean scenes, from SAR processors producing up to 8192 x 8192 x 64 bits of complex pixel data, could be addressed with future improvements in this workstation. Such a capability will accelerate development of hydrodynamic imaging models that will improve our understanding of microwave radars like the SCR, SAR, and ROWS.

Summary

Oceanographic remote sensors operated from aircraft and spacecraft as part of NASA's SIR-B experiment have yielded surface data at comparable resolution, but over ocean regions of different size. The spacecraft SAR images were Fourier filtered to obtain topographic information using a linear model of the SIR-B system response and modulation transfer function. The SAR instrument response functions were parameterized to yield Fourier spectra similar to those obtained by the ROWS instrument. A spectral power threshold was applied to segment the SAR image data to representation of surface elevation and texture.

The filtered SAR data were plotted as three dimensional surfaces to visually compare their estimate of wave height variance with that of the SCR. The SAR surface plots were also shaded with their texture information (generally referred to as speckle noise) to visually correlate short scale backscatter modulations with long wave height. For the wind driven sea observed on 10 October 1984 off the southwest coast of Chile, the synergistic study of SAR, SCR and ROWS data indicates that the speckled texture of SAR imagery may contain useful information and that a hydrodynamic theory of backscatter modulation is needed to supplement velocity bunching and tilt modulation theories.

The Johns Hopkins University Applied Physics Laboratory is developing the QUEN wavefront array processor [Dolecek, 1989] to serve as a rapid prototyping tool that can be applied to general purpose visualization in a desktop or personal computing environment. SAR processing algorithms are now being transferred to a QUEN-16 unit hosted by a VAXstation 3500 workstation to implement a menu driven user interface. Fourier filtering experiments applied quickly over larger ocean fields will accelerate research directed towards developing a hydrodynamic model [Tilley, 1990] of ocean imaging with spaceborne SAR remote sensors. High speed graphics visualization and flight simulations combined with simultaneous comparisons with scanning altimeter data will facilitate the communication of theoretical hypotheses and assist in their evaluation.

Applications of remote sensor technology for earth science include documentation of coastal erosion, surveillance of oil spills and prediction of hurricane tracks. Monitoring global change in our environment will require that remotely sensed data, collected at different scales in time and space, be reviewed, reduced and assimilated into physical models. High speed data distribution networks, desktop workstations, and advanced computing technology [Jenkins, 1989] are now being developed at The Johns Hopkins University Applied Physics Laboratory. Computer visualization is emerging as a combination of

129

three-dimensional graphic concepts with two-dimensional image processing methods as a scientific tool for investigating relationships between theoretical models and empirical data. It is planned to apply these resources to models of microwave scattering from rough surfaces that can be investigated with radar data from oceanographic remote sensors. The result of this exemplary endeavor will be insight as to the physical processes governing the wind generation of ocean waves.

Acknowledgement

References

Beal, R. C., F. M. Monaldo, D. G. Tilley, D. E. Irvine, E. J. Walsh, F. C. Jackson, D. W. Hancock III, D. E. Hines, R. N. Swift, F. I. Gonzalez, D. R. Lyzenga and L. F. Zambresky, "A Comparison of SIR-B Directional Ocean Wave Spectra with Aircraft Scanning Radar Spectra and Global Spectral Ocean Wave Model Prediction", Science, 232, 1531-1535, 1986.

Beal, R. C., "Spectrasat: A Hybrid ROWS/SAR Approach to Monitor Ocean Waves from Space", Johns Hopkins APL Tech. Dig., 8, No 1, 107-115, 1987.

Dolecek, Q. E., "QUEN: The APL Wavefront Array Processor", Johns Hopkins APL Tech. Dig., 10, No. 3, 198-207, 1989.

Jackson, F. C., W. Glazier, D. Hines, and C. Y. Peng, "ROWS Estimates of Wave Height, Wind Speed, and Directional Wave Spectra for SIR-B Underflights of Chile", 1985 International Geoscience and Remote Sensing Symposium Digest, 2, 804-811, 1985.

Jenkins, R. E., "Computing at APL Moves Into the Nineties", Johns Hopkins APL Tech. Dig., 10, No. 3, 184-185, 1989.

Monaldo, F. M., "A Practical Methodology for Estimating Wave Spectra from the SIR-B", Johns Hopkins APL Tech. Dig., 8, No. 1, 82-86, 1987.

Tilley, D. G., "Use of Speckle for Determining the Response Characteristics of Doppler Imaging Radars", Optical Engineering, 25, No. 6, 772-779, 1986.

Tilley, D. G., "SIR-B Ocean Wave Enhancement with Fast-Fourier Transform Techniques", Johns Hopkins APL Tech. Dig., 8, No. 1, 87-93, 1987.

Tilley, D. G., "Anomalous Synthetic Aperture Radar Modulations for Ocean Scenes", 1990 International Geoscience and Remote Sensing Symposium Digest, 2, 1083-1086, 1990.

Walsh, E. J., D. W. Hancock, D. E. Hines, R. N. Swift, and J. F. Scott, "Spectral Measurements in Support of SIR-B Using the Surface Contour Radar", 1985 International Geoscience and Remote Sensing Symposium Digest, 2, 798-803, 1985.
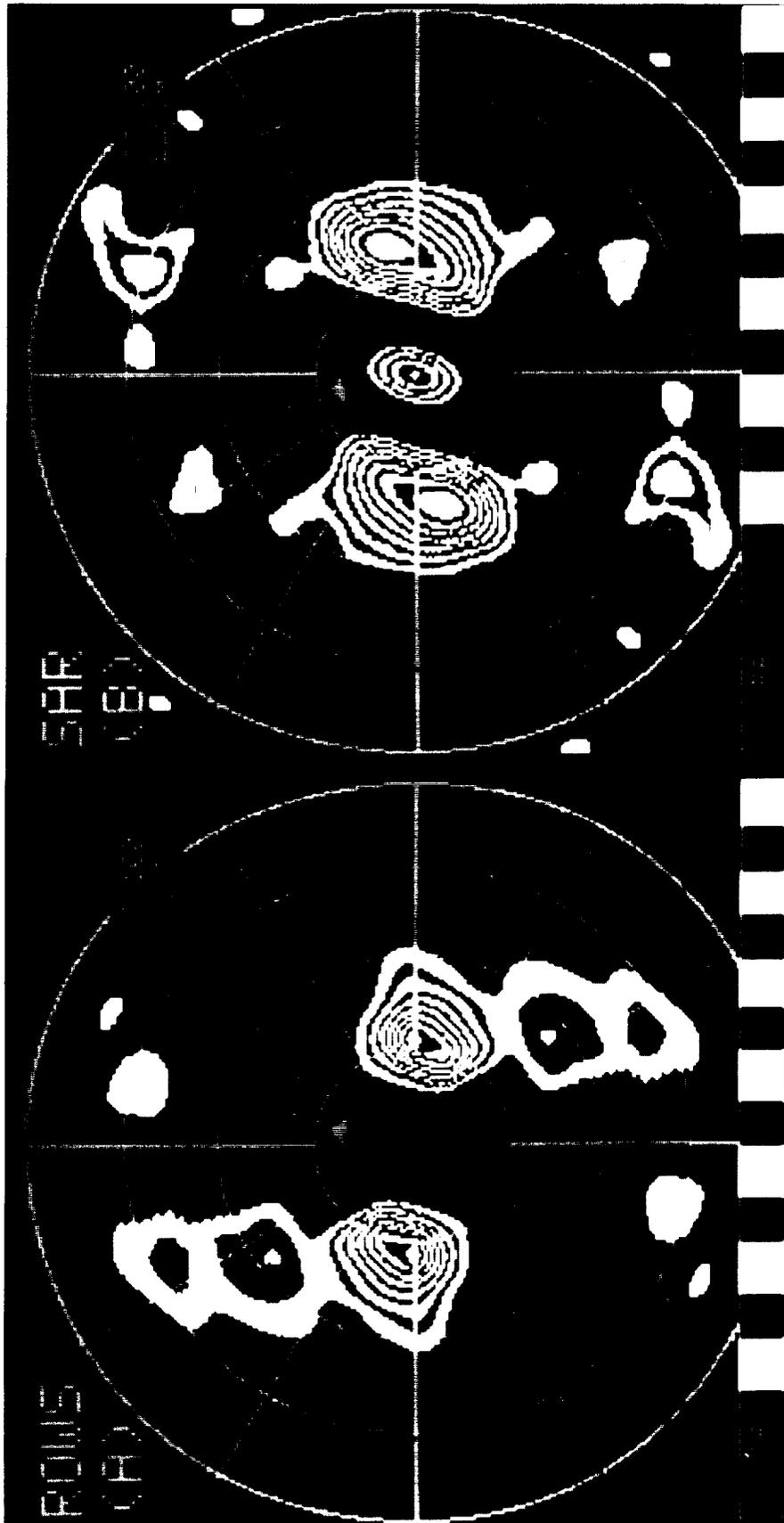
Figure 1   ROWS (a) and SAR (b) directional ocean wave spectra for wavelengths between 400 meters (inner circle) and 100 meters (outer circle). Flight directions are from left to right, horizontally.
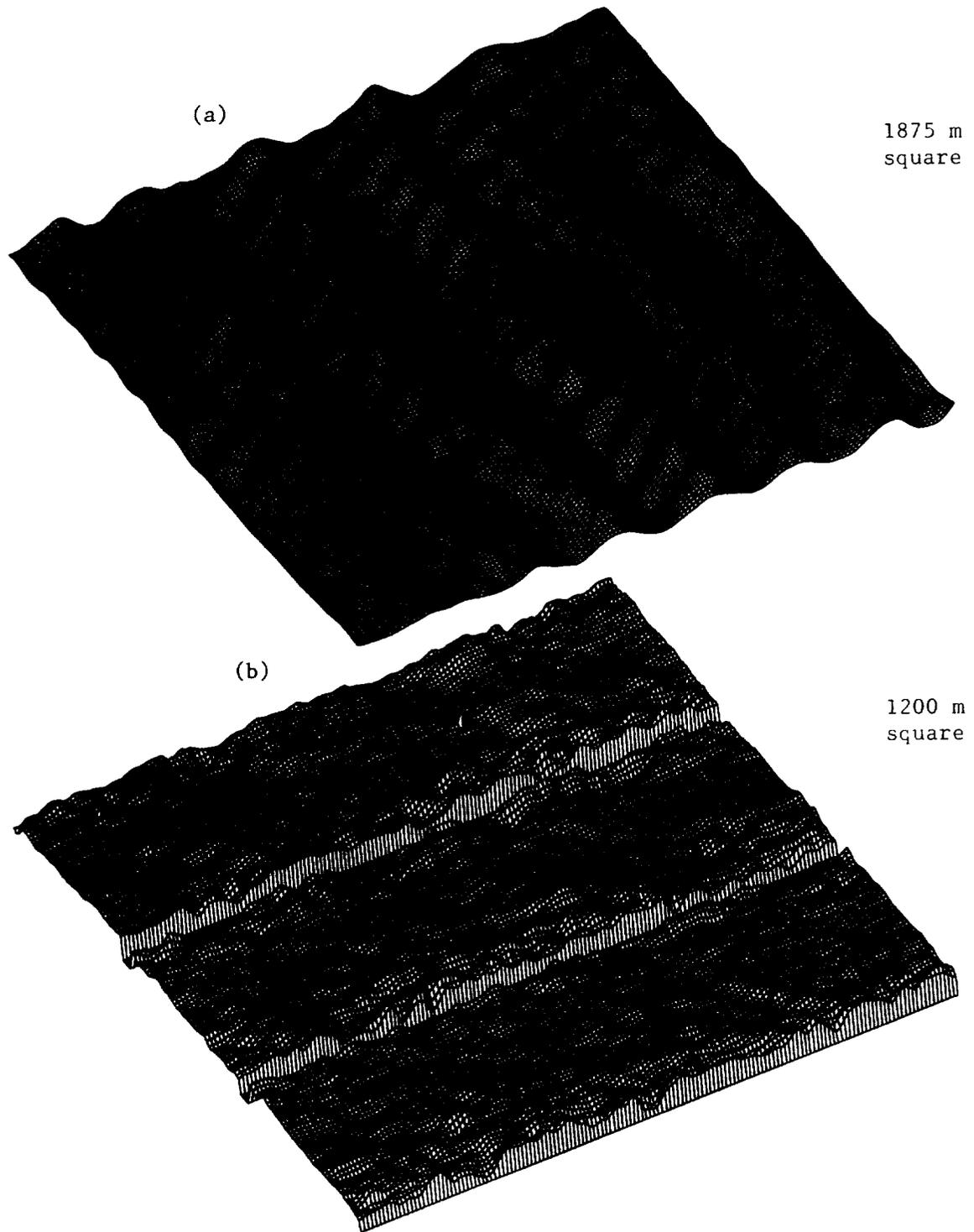
Figure 2   SAR (a) and SCR (b) surface elevation data are plotted as
three dimensional graphs depicting wave height variance for a 380
meter swell and a 140 meter sea, respectively, propagating along and
across the swaths of the remote sensors.   The SCR swath was only 400
meters wide so that 3 data sets approximate the contiguous SAR seg-
ment, although at a different place and time.
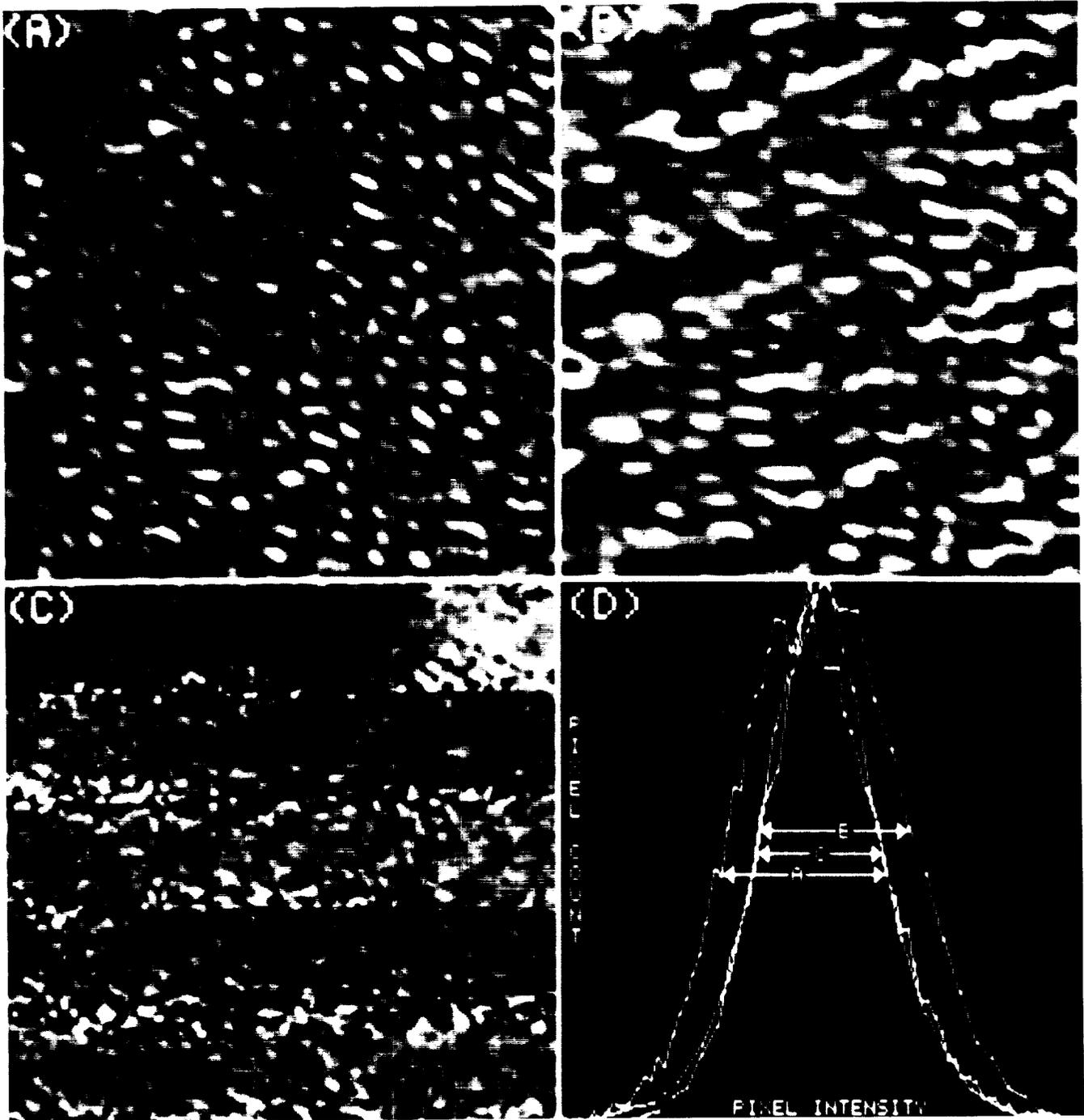
Figure 3 A 200x200 pixel segment of the Fourier filtered SAR image (a) is depicted as a two-dimensional distribution of the wave action intensity. Tilt and velocity bunching modulations of the SAR cross section can be included in the Fourier filter to simulate the wave height distribution (b), which can be compared with the SCR topography (c) along 5 separate aircraft tracks. Histograms (d) of pixel intensity counts are computed for the three data sets.

Figure 4   Long scale and short scale ocean wave correlations are
evident in a three-dimensional surface plot of the SAR wave action
intensity that has been shaded with speckled data values derived
from the unfiltered SAR image.

A PROPOSAL TO EXTEND OUR UNDERSTANDING
OF
THE GLOBAL ECONOMY

by

Robbin R. Hough
Oakland University
Rochester, Michigan 48063

Manfred Ehlers
University of Maine
Orono, ME 04469

ABSTRACT

To date, global models have been just that. They have identified problems common to the peoples of the globe, without setting forth a basis for the specific actions carried out by specific peoples that could step toward a resolution of the identified problems. There is another way.

Satellites acquire information on a global and repetitive basis. They are thus ideal tools for use when global scale and analysis over time is required. Data from satellites comes in digital format which means that it is ideally suited for incorporation in digital databases and that it can be evaluated using automated techniques.

The paper proposes the development of a global multi-source data set which integrates digital information regarding some 15000 major industrial sites worldwide with remotely sensed images of the sites. The resulting data set would provide the basis for a wide variety of studies of the global economy.

The preliminary results obtained to date give promise of a new class of global policy model which is far more detailed and helpful to local policy makers than its predecessors. It is the central thesis of this proposal that major industrial sites can be identified and their utilization can be tracked with the aid of satellite images.

The Problem

The focus of this paper is the role of policy, resources and the environment in economic and social development. The world view which it attempts to sketch for the reader will be alien to many who perceive economics as a policy science increasingly concerned with abstract "macro" entities. To be of use to humankind, policies for humans must be on a human scale. Moreover, it is of central importance that peoples be moved to perceive the objectives of the policies as important to them. To date, global models have been just that. They have identified problems common to the peoples of the globe, without setting forth a basis for the specific actions carried out by specific peoples that could step toward a resolution of the identified problems. There is

another way. Humans must have stories, images, habits and rules of thumb to live in complex environments and they just don't have them as they relate to our world as a whole. Simply browsing an atlas or even a world globe is an exercise in looking at more information than an individual can handle at once. The importance of manufacturing activity in the total market economy makes it imperative that manufacturing images and understandings be developed along with those which have to do with agricultural lands, forest lands, air quality, water quality, etc. The purpose of this proposal is to seek support for a project which will contribute to manufacturing images and understandings.

Satellites acquire information on a global and repetitive basis. They are thus ideal tools for use when global scale and analysis over time is required. Data from satellites comes in digital format which means that it is ideally suited for incorporation in digital databases and that it can be evaluated using automated techniques.

Background

In order to approach the problems discussed above, a number of preliminary efforts have been undertaken. The first effort was a study of the relationship between economic activity in the 494 Rand-McNally basic trading areas and the sites of major manufacturing facilities. The 494 basic trading areas have 1361 sites of major economic activity. The 378 trading areas with manufacturing sites have a total employment in manufacturing of 20529027. The 116 trading areas without manufacturing sites have a total manufacturing employment of 1385727. The observation that 80 percent of the trading areas have 95 percent of the manufacturing employment indicates that an appropriate definition of "major" was utilized in the identification of the manufacturing sites. A simple linear regression of sites against manufacturing employment suggests that there are 12804.5 employees per site. The results are statistically significant at the one percent level. The scale of the coefficient (1/4 of the total regional mfg. employment) again points to an appropriate definition of "major". Moreover, it was possible by further work to account for nearly 85% of manufacturing employment on an industry by industry basis.

The second effort undertaken was to construct a global database of regions which include as their centers, 738 cities chosen for their scale, their size in relation to their neighbors or their role as largest city in a nation bounded by geographic barriers or other regions. Based on the size given for the cities used as regional centers:

Average size of the largest city	959173
Number of cities per country	3.67052

A profile of manufacturing employment for the regions centering on each city was computed based on the U.S. estimates. That is, each of approximately 13000 plant sites was assumed to provide the same level of employment provided by its counterpart in the United States. Assuming that a local labor-force is trained to work in manufacturing (L) and that fuel (F) and machines (K) are imported a production function would appear as:

$$Output = f(L,F,K)$$

When estimates are aggregated by country the following result is obtained for 50 countries on which World Bank data is available:

$$National\ income = .45 * L\hat{}.54 * F\hat{}.24 * K\hat{}.18$$

$$R\hat{}2 = .912$$
$$F = 158.915$$

In essence, over 90 percent of the variation in national income appears to be accounted for by the site based employment estimates, fuel imports and machine imports. The close fit and plausible magnitude of the coefficients are taken as preliminary evidence that the U.S. based estimates will serve as reasonable substitutes for the actual levels of manufacturing employment in the regions.

A third effort sought to understand the scale of cities around the globe on the basis of the economic activity which they sustain. On a global basis, commercial farms, trading facilities and the manufacturing facilities discussed above can be shown to determine the scale of the cities supported by a given region.

A fourth effort has been directed at the development of a multi-level hierarchical flat-file manipulation language which allows relational operations on digital geographic information systems as well as the storage of links to image collections. Language M is the result of that effort. It has been in commercial use for nearly two years.

In a fifth project, DIRIGO - an image processing system, for remote sensing data has been developed. The DIRIGO system adheres strictly to the Macintosh interface guidelines and hence enables users to become quickly familiar with the system. The system supports four file formats (1) image data, (2) ASCII text files of control point coordinates or training area statistics, (3) classified images and (4) training areas for classification. An intelligent interface insures that only files with the appropriate format can be opened for the selected application tasks. At present DIRIGO supports (1) point operations such as linear and Gaussian contrast stretch and histogram equalization; (2) spatial filtering; (3) geometric correction such as image-to-map rectification and image-to-image registration using first-degree polynomials and standard re-sampling techniques; and (4)

classification techniques such as parallelpiped, minimum
distance, and maximum likelihood. Execution times for a 512 by
512 image with three bands range from seconds to approximately
five minutes for a maximum likelihood classification with six
classes.

The results reported above give promise of a new class of
global policy model which is far more detailed and helpful to
local policy makers than its predecessors. It is the central
thesis of this proposal that major industrial sites can be
identified and their utilization can be tracked with the aid of
satellite images. The concept that a major industrial facility
gives rise to employment and income which cannot be ignored is
both simple to grasp and open to verification. Policy makers
are already highly sensitive to plant openings and closings on
the scale of the plants studied here. Indeed these "local"
matters are matters of focus for legislatures, the press, trade
unions, chambers of commerce and others. By global
interdependence is meant that those actors must become aware of
the kinds of changes occurring elsewhere on the globe that will
influence the plants with which they are concerned. The paper
industry in Green Bay should be vitally interested in changes in
Finland.

Project Objective

The objective of the project being proposed is to produce a
remotely sensed image for each of the 13,000 major industrial
sites worldwide. A latitude and longitude as well as an
industrial classification can be identified for each site. Given
this information and a scene which centers on the information it
should be possible to identify the major site and a variety of
the key facilities which are physically associated with it by
interpretation. It would be expected that the working image so
identified would be but a small fraction of the total image
from which it would be extracted. That is, the industrial
plant image would usually occupy less than one tenth of one
percent of the total image. Thus each plant image would be
expected to require only about 36k of storage. The complete
global collection of images covering all sites of economic
interest would be reduced to an easily manageable 1/2
gigabyte. A full working system for global analysis and
modelling including detailed digital demographics for each of
the regions and a variety of digital maps could be handled in
less than 600 megabytes.

Products of the Project

With the images identified by site a variety of analyses are
enabled. In principle, interpretation could be supplemented by
automated analysis. A few of the many possibilities are:

1.   A study of the requirements of the industry
      specific sites based on the analysis of the

images of the site. Buildings, parking lots, water sources, retention basins, storage space, special facilities, etc would be open to examination.

2. A comparison of image features to known characteristics of the plants from other sources. Such analysis would, for example, make it possible to distinguish integrated manufacturing facilities from assembly plants, consider power requirements and siting, the relationship to common public facilities, residential spaces, transport networks and the like.

3. A third form of analysis would proceed with the tracking of these sites over time and the relating of the characteristics of the tracked images to the known economic and operating characteristics of the plants.

4. The most straightforward product of the proposed effort is a digitized industrial atlas of the world. The resulting product would have an exceedingly wide range of potential uses. It is possible to compare potential industry sites based on the characteristic parameters of other existing industrial sites. If criteria can be established combining image and non-image information of the global database, it will be possible to rate potential sites based on this information. Earlier work in this area by the project principals includes the development linear associative retrieval systems for use in comparative evaluations.

5. It would be possible to include other (future and current) satellite images with other spectral information (e.g., thermal, mid-infrared, microwave etc.) and tie this together with SPOT's high spatial resolution. This would allow multispectral, multitemporal, and multispatial site assessments.

References

Ehlers, Manfred. 1990, *Their Own Mac-Based Remote Sensing System*, Advanced Imaging, May, 1990, pp. 65-67.

Folland, Sherman T. and Hough, Robbin R. *Some Evidence of a Relationship Between Nuclear Power Plants and the Value of Agricultural Land*, Land Economics, forthcoming February 1991.

Hough, Robbin R. and Hough, Tor H., *Database Tools for the Manipulation of Hierarchic Relations*, Advances in Support systems Research, Baden-Baden, 1990.

139

Hough, Robbin R., *Natural Language and Geographic Information Systems*, Proceedings of a NCGIA Workshop on Geographic Information Systems, University of Maine, 1990.

_____. *Specialization and the Ecology of Microcomputers*, Advances in Systems Research and Cybernetics, ed. George E. Lasker, University of Windsor and Oakland University, 1989.

_____. *Evolutionary Process in Hierarchical Systems*, Perspectives on Adaptation, Environment and Population, ed. John C. Calhoun, New York: Frederick Praeger, 1986.

SUMMARIES

Summary of Data Sources Utilized in Workshop Papers

James C. Tilton

Mail Code 936
Information Systems Development Facility
NASA Goddard Space Flight Center
Greenbelt, MD 20771
Phone: (301) 286-9510
E-Mail: tilton@chrpisis.gsfc.nasa.gov
E-Mail: tilton@dftnic.gsfc.nasa.gov

The papers discussed at the IAPR TC7 Workshop utilized a wide range of data sources. These data sources are summarized in the following table:

DATA SOURCES SUMMARY

Presenter	Data Sources	Data Type	Data Resolution
Robinson	Landsat Thematic Mapper (TM)	Digital Imagery	30 meter
	Digitized Aerial Photography	Digital Label Map	60 meter
Sauer	35mm Color Slides	Analog Imagery	1mm to 1cm
Kropatsch	Synthetic Aperture Radar	Digital Imagery	(25 meter?)
	Digital Elevation Model	Elevation Map	?
Rignot	Landsat Thematic Mapper (TM)	Digital Imagery	28.5 meter
	SEASAT	Digital Imagery	25 meter
	SIR-B Synthetic Aperture Radar	Digital Imagery	25 meter
	Thermal Infrared Multispectral Scanner (TIMS)	Digital Imagery	30 meter
	SPOT	Digital Imagery	20 meter

DATA SOURCES SUMMARY (continued)

Presenter	Data Sources	Data Type	Data Resolution
Banninger	Landsat Multispectral Scanner (MSS)	Digital Imagery	80 meter
	Landsat Thematic Mapper (TM)	Digital Imagery	30 meter
	NS001 Thematic Mapper Simulator (TMS)	Digital Imagery	30 meter
	Thermal Infrared Multispectral Scanner (TIMS)	Digital Imagery	30 meter
	Airborne Imaging Spectrometer (AIS-2)	Digital Imagery	?
	Fluorescence Line Imager/		
	Programmable Multispectral Imager (FLS/PMI)	Digital Imagery	?
	Laboratory Spectral and Chemical Data	Point Data	50 meter grid intervals
Nandhakumar	Thermal Wavelength Imagery	Digitized Photography	< 1 meter
	Visible Wavelength Imagery	Digitized Photography	< 1 meter
Tilton (Kim & Swain)	Airborne Multispectral Scanner	Digital Imagery	50 meters
	Synthetic Aperture Radar	Digital Imagery	50 meters
	Simulated HIRIS data	Digital Imagery	approx. 30 meters
Perlant	Aerial Photography (Stereo Pairs)	Digitized Photography	1.3 meter
Shufelt	Aerial Photography	Digitized Photography	1.2 meter
Pinz	Color Infrared Aerial Photographs	Digitized Photography	10 to 80 cm
Tilley	SIR-B SAR	Digital Imagery	25 meters
	Surface Counter Rader	Digital Imagery	<= 25 meters
	Radar Ocean Wave Spectrometer	Digital Imagery	<= 25 meters
Hough	Digital Database	Point Data	--

144

Summary of Types of Data Fusion Methods Utilized in Workshop Papers

Sylvia S. Shen

Lockheed Palo Alto Research Lab.
O/96-30, B/251
3251 Hanover St., Palo Alto, CA 94304
Phone: (415) 354-5019
E-Mail: dipl!shen@apple.com

ABSTRACT

This paper gives a brief overview of multisource data fusion techniques. Three types of data integration methods, namely pixel, feature, and decision-level fusion are discussed. The paper also gives a summary of all the papers presented at the IAPR Workshop on Multisource Data Integration in Remote Sensing, categorizing the type of data fusion methods used and detailing the utilization of multisource data sets in earth science applications.

1. OVERVIEW OF MULTISOURCE DATA FUSION TECHNIQUES

With the recent advances in sensor technology, the number of different sensor platforms that carry imaging payloads has increased tremendously. These sensors produce data covering different portions of a broad range of the electromagnetic spectrum at different spectral and spatial resolutions, furnishing users with enormous amount of useful information for terrestrial, oceanic, geophysical, meteorological, reconnaissance, surveillance, and planetary studies. In addition to these spaceborne and airborne sensory data, digital conversion of analog data such as aerial photographs, topographical and geophysical data are also available for researchers. These data are heterogeneous in their format, radiometric characteristics, geometric properties, and temporal sampling. To fully exploit these increasingly sophisticated multisource data, advanced analytical or numerical data fusion techniques must be developed.

Data fusion techniques can be categorized into three types according to the stage at which fusion takes place. The three types are pixel, feature, and decision-level fusion techniques. Pixel-level fusion techniques form 'new' pixels with a pre-selected spatial resolution common to all data sources involved. Sensory and/or ground reference measurements for each of the 'new' pixels are derived from the original data sources. These measurements are then accumulated for each 'new' pixel. Image registration is a typically example of data fusion performed at the pixel-level. In this case, pixel-by-pixel comparison of multiple images of a scene obtained from different sensors or taken from the same sensor at different time are formed. It is accomplished by spatially registering the images, namely correcting for relative translational shifts, magnitude differences, rotational shifts, as well as geometrical and intensity distortions of each image. Pixel-level fusion

145

allows different sensory or ground reference information be accumulated for each 'new' pixel. Feature-level fusion techniques generally start with applying image analysis techniques to extract some features from each data source independently. Certain measurements of the extracted features are derived from each data source and these feature measurements are then integrated. For instance, when both visible and thermal infrared image data are available, regions of nearly homogeneous intensity can be extracted using image segmentation techniques. Average intensity value and average surface temperature of each region can be calculated from the visible and thermal infrared images respectively. The intensity and temperature information are fused at the region-level in this example. Decision-level or interpretation-level fusion deals with integrating the various versions of interpretation obtained from the different data sources to arrive at a consensus interpretation. For example, in the case of image classification, the interpretation or classification results can be represented in terms of probability assignments. Bayes rule and Dempster-Shafer's rule of combination are often used as a means to reinforce common interpretation and to resolve differences in order to arrive at a more complete and accurate interpretation. They are typical examples of decision-level fusion techniques.

A summary of all the papers given at the IAPR Workshop on Multisource Data Integration in Remote Sensing is presented in the next section. The data types and data fusion methods used in each paper are discussed. Fusion techniques are categorized into either pixel, feature, or decision-level fusion. The utilization of multisource data sets in earth science applications by each paper is also briefly described.

2. SUMMARY OF WORKSHOP PAPERS

The paper entitled "Refinement of Ground Reference Data with Segmented Image Data" by Robinson and Tilton discusses an approach to refining and adding detail to the ground reference data through the use of satellite image data. More specifically, the approach segments the Landsat TM images, finds the edges from the segmentation, and uses the edges from the ground reference data to mask out those segmented edge points that are beyond a certain distance from the reference data edge points. This approach is categorized as a feature level integration method where the edge information extracted from ground reference data and Landsat TM image data are fused. This fusion allows reference edge error be corrected and detailed edge information be added to the ground reference data.

The paper entitled "Near Ground Level Sensing for Spatial analysis of Vegetation" by Rasure, Sauer and Gage describes a workstation based image processing software system called Chorus. This system consists of a low level segmentation process, a supervised and unsupervised clustering process, an object feature extraction process, a classification process, and a region labeling process. The system has been used to process near ground level image data to distinguish living biomass from non-living biomass. No multisource data fusion is discussed in this paper. However, correlation between changes detected from Landsat TM image data and those observed from near ground level image data will be studied in the future.

The paper "Integration of SAR and DEM Data -- Geometrical Considerations" by Kropatsch describes an approach to fusing information from SAR images and Digital Elevation Model

146

(DEM) data based on establishing some geometrical relations between the data sets. Layover and shadow regions are detected in SAR images. Such regions are also found independently using DEM data. Information from both data sets are then accumulated through geometrical matching of these layover and shadow regions. This fusion technique is considered to be at the feature level where the features are the layover and shadow regions.

The paper entitled "Towards Operational Multisensor Data Registration" by Rignot, Kwok and Curlander addresses the problem of automated precision registration of multisource image data acquired from a number of different remote sensors. Two types of registration techniques are presented. The first technique uses the Digital Elevation Model (DEM) to simulate the sensory data such as Landsat TM, Seasat and SPOT, therefore the various types of optical data are registered. The second technique first extracts features such as edges from the different images and then matches the extracted features between the different images. This feature matching technique replaces the manual selection of tie points for transformation parameter estimation. Both techniques allow us to accumulate information from different sensors for each pixel at a pre-selected spatial resolution common to all sensors involved. Registration is, in general, considered to be a typical example of data fusion at the pixel level.

The paper "Combined Fluorescence, Reflectance, and Ground Measurements of Stressed Norway Spruce Forest for Forest Damage Assessment" by Banninger discusses the problem of using remote sensing data, laboratory data, and ground measurements to monitor and differentiate stress or damage in forested areas. Discrete data sets such as foliar chlorophyll-a and nitrogen content, leaf area indices, and foliage reflectance and fluorescence measurements are obtained from samples collected at 50-m intervals over the test sites. These discrete data sets are resampled to facilitate information integration with remotely sensed data such as Landsat MSS and TM, NS001 TMS, TIMS, AIS-2 and FLS/PMI. Therefore, the data fusion is performed at the pixel level.

The paper entitled "A phenomenological Approach to Multisource Data Integration : Analyzing Infrared and Visible Data" by Nandhakumar describes a region classification approach based on integrated analysis of thermal IR and visible data of remotely sensed scenes. In his approach, the visible image is first segmented into regions of nearly homogeneous intensity. The thermal IR image and a phenomenological model describing the exchange of energy between the imaged surface and the environment allow the estimation of surface heat flux for each pixel. Here, the surface temperature information is fused with the intensity values from the visible image at the pixel level. When the information is used to classify regions into distinct class categories, data fusion is performed at the feature level where the features are the segmented regions. Aggregated intensity and surface temperature values are computed for each region and used as discriminants in a rule-based classification scheme.

The paper entitled "A Method for Classification of Multisource Data Using Interval-Valued Probabilities and Its Application to HIRIS Data" by Kim and Swain discusses a method of classifying multisource data in remote sensing. The authors defines what is called the interval-valued probability which is a generalization of the point-valued probability designed to represent the pos-

sibility of small deviations from the unknown true probability. Their approach treats each data source as a body of evidence, represents the evidence provided by each data source in terms of interval-valued probabilities, and uses the Dempster-Shafer's rule of combination to integrate the interval-valued probabilities form different data sources. This integration is therefore categorized as to be at the decision or interpretation level. This method can combine parametric as well as nonparametric information.

The paper "Improved Disparity Map Analysis Through the Fusion of Monocular Image Segmentation" by Perlant and McKeown first addresses the issue of using the segmentation of the monocular imagery to improve the estimates of three-dimensional scene structure, namely the scene disparity map. Their approach generates the initial disparity map from a stereo pair of images using area-based and/or feature-based matching techniques. Then the surface illumination information provided by a monocular image is used to segment the image into regions of nearly homogeneous intensity. Based on the assumption that such regions correspond closely to physical surfaces in the scene, the region information is used to remove mismatches. Here, the disparity information extracted from stereo pair is fused with segmentation results obtained form monocular image at the feature level where the features are the segmented regions. The authors later address the problem of using a number of different region segmentation methods and stereo matching techniques to improve building extraction accuracy. Each pair of region segmentation and stereo matching techniques produces a version of building extraction interpretation. The various versions of interpretation are then used in a voting process to arrive at a consensus interpretation. The information integration used here to achieve a more complete and accurate building extraction is considered to be at the decision or interpretation level.

The paper entitled "Use of Information Fusion to Improve the Detection of Man-Made Structure in Aerial Imagery" by Shufelt and McKeown addresses the problem of building information fusion in aerial imagery. Since it is assumed that no single detection method will accurately delineate buildings in every scene, a cooperative method is proposed. Various building extraction techniques such as the ones based on edge analysis, shadow analysis, and stereo analysis are used to produce building delineations. The different delineation results are superimposed to generate a more accurate building interpretation. Quantitative measures such as building percentage, background percentage and branch factor are utilized as a means to evaluate building extraction results. In this study, only one type of sensory data, namely, the aerial image is used. Therefore, fusion is not performed using different sensory data, but rather using the different delineation results on the same image data. The information fused is the edge information. It is therefore categorized as a feature level integration method.

The paper "A Computer Vision System for the Recognition of Trees in Aerial Photographs" by Pinz describes the concept of a computer vision system capable of finding trees in infrared aerial photographs. Generic description of a type of an object is integrated with the image object where an object is, in this case, a tree. The tree recognition results from aerial photographs taken at different resolution scales and under various conditions are also integrated at the feature, namely tree, level.

The paper "Visualizing Characteristics of Ocean Data Collected During the Shuttle Imaging Radar-B Experiment" by Tilley discusses the computation of ocean wave spectra (power spectra of the Fourier transform) from SIR-B Surface Contour Radar (SCR) data and the comparison of these clusters of energy with those computed from the Radar Ocean Wave Spectrometer (ROWS) data and the SIR-B Synthetic Aperture Radar (SAR) data to parametrize models of ocean imaging. No data fusion of these three types of sensory data is used in this study. Fusion will, however, be considered in the future.

The paper "A proposal to Extend Our Understanding of the Global Economy" by Hough and Ehlers proposes the development of a global multisource data set which integrates digital information including remotely sensed images regarding 15000 major industrial sites worldwide to track their utilization as a basis for a variety of global economic studies. No specific data types or fusion techniques are mentioned in this conceptual paper.

Summary of Types of Data Output Produced
by Studies Described in Workshop Papers

Axel J. Pinz

Institute of Surveying and Remote Sensing
University of Agriculture
Peter Jordan Str. 82
A-1190 Wien, Austria
Phone: (43-222) 342500-546

There were a total of 12 presentations at the IAPR TC7 Workshop. Most of them dealt with the description of systems. While the two previous summaries discussed the input and processing aspects of these systems, this summary discussed the *output* aspect.

Many contributions were concerned with an *improvement* or a kind of *refinement of results* by use of multisource data. The improvements are either in overall accuracy or in robustness. The output may be categorized as either a better segmentation, an improved symbolic description of the scene, or a more robust classification. These *"improvement-papers"* are:

- Jon Robinson: The output is a refined data set of edges (which can be translated into a refined ground cover label map).

- Tom Sauer: The analysis process deals with classification, region analysis and spatial information. In its current state of development the system is separating vegetation from background. The final output is the percentage of ground coverage by specific vegetation classes

- N. Nandakumar: In this combination of thermal and visual information, the thermal capacity is calculated and used to improve a segmentation, and to label various classes like pavement, vegetation, car, etc.

- H. Kim and P. H. Swain (presented by J. C. Tilton): The integration of multispectral and DEM data leads to a more robust classification.

- Frederick Perlant: These two papers from David McKeown's group at Carnegie-
- Jeffrey Shufelt: Mellon University discuss information fusion. There are several different processes run on the same image or on different images (stereo pair). The output of these processes in than fused yielding an improved result. In Perlant's work the output is a segmentation. Shufelt ends up with a set of boundaries delineating the objects of interest (e. g. houses).

- Axel Pinz: Having several images of the same scene, the first step of processing leads to several symbolic description of this scene. The integration (fusion) of theses results yields an improved and more robust symbolic description of the scene objects (in this case, trees).

The following contributions cannot be categorized so easily (as "improvement-papers"). It seems that the desired output of these systems wouldn't be possible without multisource data integration:

- Walter Kropatsch: The integration of SAR, DEM and knowledge results in a geocoded SAR image. Furthermore, Dr. Kropatsch can map additional information (e.g., slope) into the original SAR image, providing means for a better interpretation of the original image.

- Eric Rignot: This is going to be a very huge system. The output will be a pack of registered and resampled data at uniform scale.

- C. Banninger: The integration of a few Landsat pixels with a very detailed ground data set is used for an optimal positioning of the pixel grid. The outputs of this process are correlation results and models.

- David Tilley: Since the original radar image of the ocean appear very uniform to an human interpreter, the first task to solve is the proper visualization of the data. Furthermore, Dr. TIlley hopes to gain more insight into imaging mechanisms of SAR, and to uncover information in the speckles.

- Robbin Hough: A database systems of this kind would be able of producing manyfold outputs depending on the query. It could be used as a tool for economic forecasting, planning and analysis.

APPENDIX

Workshop Participants

PARTICIPANTS

Richard Arens
GTE Government Systems
1700 Research Blvd
Rockville, MD 20850
Tel: (301) 294-8430

Cliff Banninger
Institute for Image Proc. & Computer
Graphics
Joanneum Research
Wastiangasse 6, 1-8010
Graz, AUSTRIA
Tel: 011-43-316-8021-0
Fax: 011-43-316-8021-20
E-Mail: banninger%rzj.fgi-graz.ada.at@
cunyvm.cuny.edu

David W. Case
Applied Research Corp.
8201 Corporate Drive
Landover, MD 20706
Tel: (301) 805-0379 or (301) 459-8442

Long S. Chiu
General Sciences Corporation
6100 Chevy Chase Drive
Laurel, MD 20707
Tel: (301) 953-2700

Edward A. Dunlop
Center for Remote Sensing
College of Marine Studies
University of Delaware
Newark, DE 19716
Tel: (215) 430-6549
E-Mail: dunlop@freezer.it.udel.edu

Marcus E. Glenn
MITRE Corp.
7525 Coleshire Dr.
McLean, VA 22102
Tel: (703) 883-5386
E-Mail: glenn@mitre.org

Dr. Robbin R. Hough
Oakland University
Rochester, MI 48063
Tel: (313) 651-0820
Fax: (313) 651-4071
E-Mail: hough@unix.secs.oakland.edu

Walter G. Kropatsch
Institute for Technical Informatics
Technical University of Vienna
Treitlstr. 3
A.1040 Vienna, AUSTRIA
Tel: 011-43-222-58801-8181
Fax: 011-43-222-569149
E-Mail: geli@vmars.uucp

Robert Losa
USDA/NASS
Rm 4168 - South Bldg
Washington, D. C. 20250
Tel: (202) 447-9295

David M. McKeown
Digital Mapping Laboratroy
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
U.S.A.
Tel: (412) 268-2626
E-Mail: dmm@cs.cmu.edu

N. Nandhakumar
Dept. of Electrical Engineering
Thornton Hall, Rm E-210
University of Virginia
Charlottesville, VA 22903-2242
Tel: (804) 924-6108
Fax: (804) 924-8818
E-Mail: nandhu@virginia.edu

Mishh Pavel
Bldg 420
Stanford University
Stanford, CA 94305
Tel: (415) 725-2430
mis@psych.stanford.edu

Frederic P. Perlant
Digital Mapping Laboratory
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213-3890
Tel: (412) 268-7552
E-Mail: fpp@cs.cmu.edu

Axel J. Pinz
Inst. of Geodesy and Remote Sensing
Univ. of Agriculture
Peter Jordan Str. 82
A-1190 Wien, AUSTRIA
Tel: (43-222) 342500-546

Eric Rignot
Jet Propulsion Laboratory
4800 Oak Grove Drive
California Institute of Technology
Pasadena, CA 91109
Tel: (818) 354-1640
telex: 675-429
Fax: (818) 354-3437
E-Mail: eric@malibu.jpl.nasa.gov

Jon Robinson
ST Systems Corporation
4400 Forbes Boulevard
Lanham, MD 20706
U. S. A.
Tel: (301) 794-5200
E-Mail: mscc@dftnic.gsfc.nasa.gov

Tom Sauer
University of New Mexico
Dept. of Electrical and Computer
Engineering
Albuquerque, NM 87131
Tel: (505) 277-6563
Fax: (505) 277-1439
E-Mail: sauer@bullwinkle.unm.edu

Andrew Segal
University of Illinois
Dept. of EE CS
M/C 154, Box 4348
Chicago, IL 60680
U.S.A.
Tel: (312) 996-5486
E-Mail: segal@uicbert.eecs.uic.edu

Robert Series
RSRE
St. Andrews Road, Malvern
Worcs WR14 3PS
UNITED KINGDOM
Tel: (+44) 684-5784
E-Mail: series@uk.mod.rsre

Sylvia Shen
Lockheed Palo Alto Research Lab.
O/96-30, B/251
3251 Hanover Street
Palo Alto, CA 94304
Tel: (415) 354-5019
E-Mail: dipl!shen@apple.com

Jeffrey Shufelt
Digital Mapping Laboratory
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
Tel: (412) 268-3884
E-Mail: js@maps.cs.cum.edu

Stan Stopyra
GTE Government Systems
1700 Research Blvd
Rockville, MD 20850
Tel: (301) 294-8430

David G. Tilley
The Johns Hopkins University
Applied Physics Laboratory
MS 23-346
Johns Hopkins Road
Laurel, MD 20723-6099
Tel: (301) 953-5000 x8644
Fax: (301) 953-1093
Telex: 511303
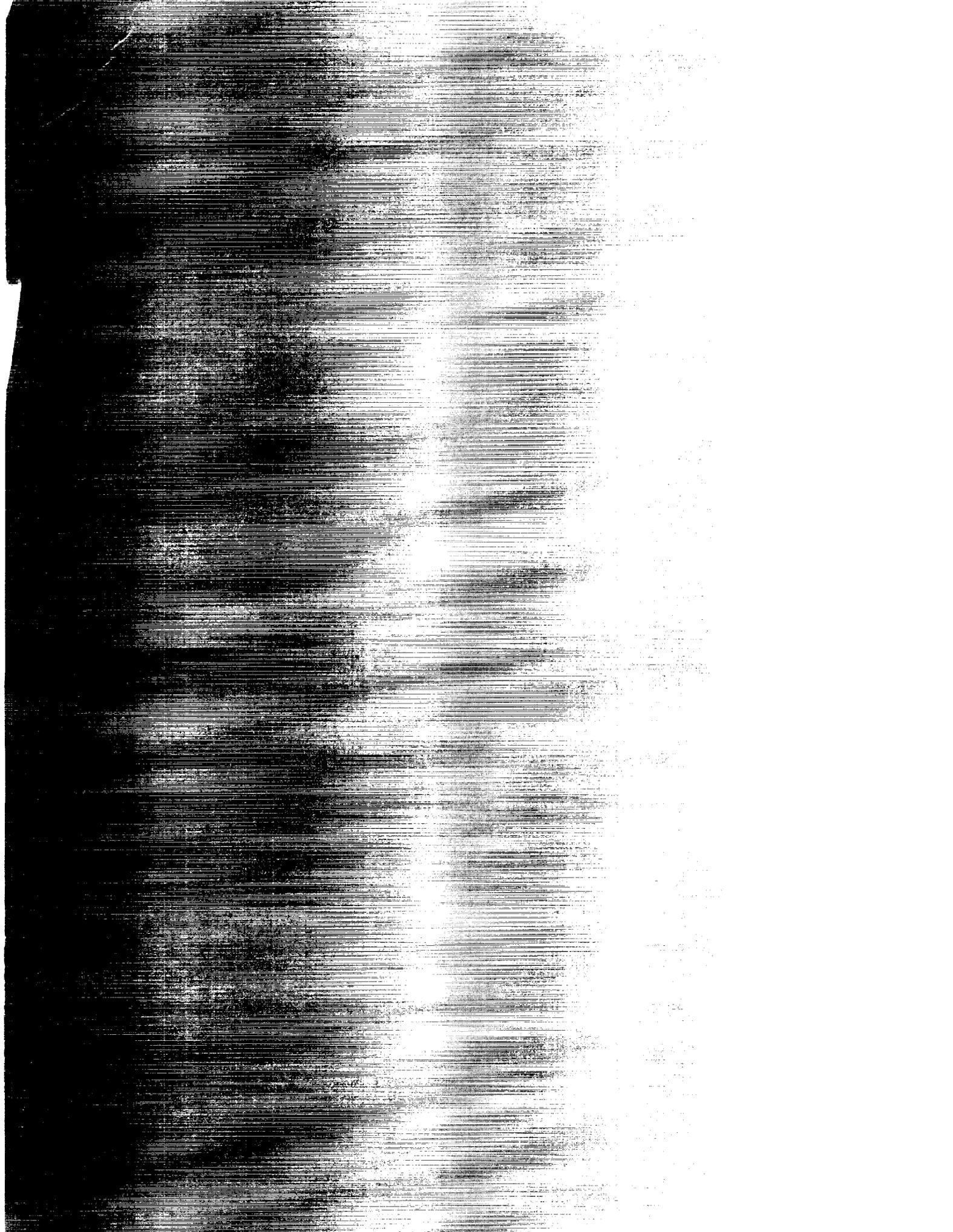E-Mail: davet@fermi.apljhu.edu

James C. Tilton
Mail Code 936
NASA GSFC
Greenbelt, MD 20771
Tel: (301) 286-9510
E-Mail: tilton@chrpisis.gsfc.nasa.gov or
tilton@dftnic.gsfc.nasa.gov

E. D. Zeisler
8242 West Buckspark Lane
Potomac, MD 20854
Tel: (703) 883-5768

NASA
National Aeronautics and
Space Administration

Report Documentation Page

1. Report No. NASA CP-3099	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Multisource Data Integration in Remote Sensing		5. Report Date **January 1991**
		6. Performing Organization Code 930
7. Author(s) James C. Tilton, Editor		8. Performing Organization Report No. 90B00122
		10. Work Unit No.
9. Performing Organization Name and Address NASA Goddard Space Flight Center Greenbelt, MD 20771		11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546-0001		13. Type of Report and Period Covered Conference Publication
		14. Sponsoring Agency Code

15. Supplementary Notes

This workshop was organized by Technical Committee 7 of the International Association for Pattern Recognition and co-sponsored by the Space Data and Computing Division (Code 930) of the Goddard Space Flight Center, and by the Washington/Northern Virginia Chapter of the IEEE Geoscience and Remote Sensing Society.

16. Abstract

Multisource Data Integration in Remote Sensing will become a real challenge in the near future. New instruments and new sensors will provide us with a large variety of new views of the "real world." This huge amount of data has to be combined and integrated in a (computer-) model of this world. But also, the knowledge of how these data are gathered and what their characteristic properties are is among the useful sources of information that contribute to a meaningful interpretation. Multiple sources may give us complementary views of the world--consistent observations from different (and independent) data sources support each other and increase their credibility, while contradictions may be caused by noise, errors during processing, or misinterpretations, and can be identified as such. As a consequence, integration results are very reliable and they represent a valid source of information for any geographical information system (GIS).

This workshop consisted of three sessions of three to five individual presentations. All papers were discussed both individually and in the general context of the session. The full text of these papers is included in this Conference Publication, along with three workshop summary papers.

17. Key Words (Suggested by Author(s)) Multisource Data Integration Remote Sensing Geographic Information Systems Ground Reference Data Image Registration	18. Distribution Statement Unclassified - Unlimited Subject Category 43
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of pages 164	22. Price A08

NASA FORM 1626 OCT 86

National Aeronautics and
Space Administration

Goddard

Wallops Flight
Zone 23001

NASA