Use of Multi-Resolution Wavelet Feature Pyramids for Automatic Registration of Multi-Sensor Imagery

Ilya Zavorin, Jacqueline Le Moigne, *Senior Member, IEEE*

Abstract

The problem of image registration, or alignment of two or more images representing the same scene or object, has to be addressed in various disciplines that employ digital imaging. In the area of remote sensing, just like in medical imaging or computer vision, it is necessary to design robust, fast and widely applicable algorithms that would allow automatic registration of images generated by various imaging platforms at the same or different times, and that would provide sub-pixel accuracy. One of the main issues that needs to be addressed when developing a registration algorithm is what type of information should be extracted from the images being registered, to be used in the search for the geometric transformation that best aligns them. The main objective of this paper is to evaluate several wavelet pyramids that may be used both for invariant feature extraction and for representing images at multiple spatial resolutions to accelerate registration. We find that the band-pass wavelets obtained from the Steerable Pyramid due to Simoncelli perform better than two types of low-pass pyramids when the images being registered have relatively small amount of nonlinear radiometric variations between them. Based on these findings, we propose a modification of a gradient-based registration algorithm that has recently been developed for medical data. We test the modified algorithm on several sets of real and synthetic satellite imagery.

I. INTRODUCTION

The problem of image registration, when two or more images of approximately the same scene or objects have to be geometrically aligned, arises in virtually all disciplines where digital images are used for analysis of the underlying objects or processes, including biomedical imaging, computer vision, remote sensing and microlithography [1]. In all cases, the fundamental goal is the same: it is necessary to design fast and robust algorithms that would perform automatic image registration and, in most cases, sub-pixel registration is required. However, because of the differences in how the images are acquired, what they contain, and why they need to be aligned, one cannot expect to design a perfect registration algorithm that would perform well in all cases. Nevertheless algorithms must not be too data- or application-specific to be practical. This problem

has been address to a large extent in medical imaging [2] while the area of remote sensing only now is starting to catch up.

As the amount of imaging data generated by various Earth Observing satellites grows rapidly, it becomes essential to develop reliable automatic algorithms for both on-the-ground and on-board processing of these data. However, before images generated by different sensors and/or at different times could be used for such high-level tasks as change detection or data fusion, these images have to be accurately registered. Despite the large numbers of automatic image registration methods proposed in the literature over the last 10 to 20 years, manual registration, which is often time consuming and inaccurate, remains by far the most common way that remote sensing specialists utilize to align their imaging data. Such powerful and widely used commercial packages such as ENVI, Imagine, and Matlab do not offer automatic registration. This contrasts sharply with the area of biomedical imaging where several registration packages have been successfully used in everyday operations. At the same time, automatic methods proposed by various authors are often tailored for a specific collection or type of satellite data and thus may not be widely applicable. Therefore the ultimate goal of the image registration effort at the NASA Goddard Space Flight Center is to create a toolbox-type collection of automatic and semi-automatic image registration methods that would allow an Earth Science user to quickly and reliably process as many different types of remotely sensed data as possible [3], [4]. We assume that the data have been radiometrically and systematically corrected, which usually yields registration within a few pixels, and so our goal is to develop methods that provide sub-pixel accuracy.

In our work we build upon some of the ideas developed in biomedical imaging research. For instance, the algorithms discussed in this paper were originally designed for and tested on medical data. However, the problem of remote sensing registration is different in several respects. *First*, multi-sensor satellite and aerial images often have significantly different spatial resolutions. For example, the Landsat Enchanced Thematic Mapper Plus (ETM+) sensor produces 30-meter images while the IKONOS sensor has the resolution of 4 meters. Such differences are rare in medical data. *Second*, while three-dimensional medical images may be quite large in terms of memory required to store them, remote sensing images often cover considerably larger scenes in terms of heterogenuity of their contents. *Third*, due to relative stability of imaging platforms and systematic data correction, global rigid transformations usually represent misalignment between satellite images quite well, while medical images often require local "rubber sheet" type warping to account for such phenomena as heartbeat, breathing or random movement of subjects.

A registration algorithm searches for a geometric transformation of a certain type that best aligns a given pair of images. The algorithm consists of several components that determine what information from the images (also called *images features*) is used to find the best match, how it is searched for and what metrics

measures similarity of two images. These components are discussed in Section II of this paper. When assessing performance of a given registration method, it is essential to use appropriate test data and develop a systematic testing methodology. Both of these issues are addressed in Section III.

This paper has two goals. The main objective is to evaluate and compare overall performance of several wavelet pyramids in the context of multiresolution image registration, by assessing their accuracy as function of both geometric transformation parameters and sensitivity to noise. Based on the results of this evaluation, our second objective is to develop a modification of an algorithm recently proposed for registration of medical imagery [5]. The modified method is designed to handle single- and multi-sensor satellite data with relatively small amounts of nonlinear radiometric variation.

The multiresolution approach to image registration amounts to, first, representing the two images at several spatial resolutions using some sort of filtering and decimation framework, followed by progressive alignment of the image representations by going from the coarsest one to the finest. In Section IV we study which wavelet pyramids yield image features that are best suited for registration. At this stage, we use the simplest possible search strategy, the exhaustive search, and base our tests on synthetically generated test imagery. Then, in Section V, we combine these wavelet pyramids with a more sophisticated iterative gradient-based search technique and study its performance when applied to various types of synthetic and "real-life" data. We summarize our findings in Section VI.

II. VARIOUS COMPONENTS OF IMAGE REGISTRATION

A. Definition of Muliresolution Image Registration

Let $F_R(x,y)$ and $F_I(x,y)$, $(x,y) \in \Omega \subset \mathcal{R}^2$, where Ω is a region of interest, be two grey-level images that we call *reference* and *input*, respectively. If they are both of size $A \times B$ pixels, then $\Omega = [0, A - 1] \otimes [0, B - 1]$. Let $T_p(x,y)$ belong to a certain class of geometric transformations, where p is a vector of transformation parameters. For instance, in the case of the "Rotation-Scale-Translation" (RST) transformation, $p = (t_x, t_y, \theta, k)$ and T_p has the form

$$T_p(x,y) = Q_p v, \quad \text{with} \quad Q_p \equiv \begin{pmatrix} k\cos\theta & k\sin\theta & t_x \\ -k\sin\theta & k\cos\theta & t_y \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{and} \quad v \equiv \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (1)$$

where $\{t_x, t_y\}$ are translations in x and y directions, θ is the rotation angle and k is the isometric scaling[1]. Q_p is the RST transformation matrix, and there is a one-to-one correspondence between p and Q_p. To register

---

[1]In (1), the origin is assumed to be at the upper-left corner of an image with the y axis directed downwards. When rotation or scaling is performed, the origin is shifted to the center of the image.

$F_R(x,y)$ and $F_I(x,y)$ is to find the value of p such that $F_R(T_p(x,y))$, the reference transformed by T_p, best matches the input. One way to look at the registration process is to compare manual and automatic registration. Manual registration usually consists of two steps. First, a set of matching pairs of *control points*, or *landmarks*, is selected in the reference and input. Then, these pairs are used to compute a transformation $T_p(x,y)$ between the images. Resulting accuracy is determined primarily by the quality of the control points. From this point of view, automatic registration algorithms can be divided into two groups: the first group contains methods that "mimic" manual registration in the sense that they also collect pairs of matching control points. To improve accuracy, they usually start with a large collection of candidate pairs and then discard all but a few pairs that are considered the most reliable according to a certain measure [6]–[12]. An alternative approach is to use cross-correlation or optimization and take into account the entire image [5], [13]–[16]. In this type of approach, features such as gray levels, edges, or wavelet coefficients are selected and then these sets of features in reference and input images are globally matched either in the spatial or in the spectral domain. Our work chooses an approach that corresponds to a modified global approach: while a "global" similarity metrics is utilized, portions of the images where image intensity does not vary significantly are masked out. Noise is also removed by pre-processing the data with low-pass filters.

There is another way of classifying registration algorithms, according to the four criteria proposed by Brown [1]. The *search space* is the class of potential transformations $T_p(x,y)$ that establish the correspondence between the reference and the input. The *feature space* determines the type of information extracted from the images that is used to find the best transformation. The *similarity metric* gives the meaning to the term "best match". Finally, the *search strategy* describes how the features and the metric are used to find the best $T_p(x,y)$.

1) Multiresolution Approach to Registration: Since satellite images are often quite large, a multi-resolution approach is often chosen as a registration framework [5], [17]–[19]. Our earlier work also adopted such an approach to registration [20], [21]. It is illustrated in Figure 1. Starting with the reference and input images R_0 and I_0, we decompose them into respective "pyramids" $\{R_0, R_1, \ldots\}$ and $\{I_0, I_1, \ldots\}$ that contain representation of these images at reduced resolutions. For graphical reasons, Figure 1a shows the example of a pyramid generating only one sub-image at each level of decomposition. The pyramids are usually dyadic, i.e. at each level of decomposition the new sub-image is at half the spatial resolution of the sub-image at the previous level. The registration then starts with the coarsest pair of corresponding sub-images and obtains progressively refined approximations $\{T_{P1}, T_{P2}, \ldots\}$ to the final resulting transformation T_P by going up the pyramids. Note that in general the reference and input do not need to be of the same spatial resolution, but their pyramids are aligned for registration in such a way as to approximate similar spatial resolutions for
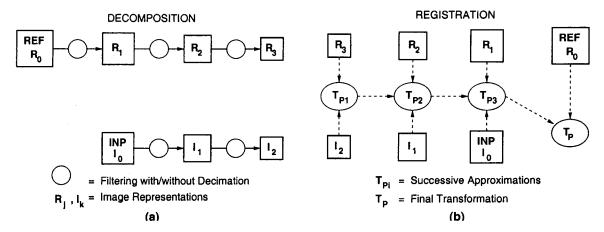
DECOMPOSITION   REGISTRATION

$$\bigcirc = \text{Filtering with/without Decimation}$$

R_j, I_k = Image Representations

T_{Pi} = Successive Approximations

T_P = Final Transformation

(a)   **(b)**

Fig. 1: Multiresolution Decomposition/Registration Framework

figure

corresponding sub-images of each pyramid level. So, in Figure 1, (sub)images R_1 and I_0 are of the same resolution, as are R_2 and I_1 etc. Although in this paper the same search method is used at all pyramid levels, in general one may select different methods that are better suited for different resolutions . For example, a wider search may be utilized at the coarsest level thus avoiding to follow the path of a local instead of a global optimum. Then, narrower searches can be pursued at finer levels when the transformation is refined. Some other authors [22] have also proposed to pursue several paths for similar robustness concerns. When rotation or translation parameters are small, independent searches may also be performed for each parameter and at each level (e.g. rotation-only, translation-only) and then be reconciled before moving to the next level [21], [23]. Overall, provided that computation of the pyramid is not too expensive and that enough information in the images is preserved in the process, using a multiresolution framework often significantly reduces computational requirements compared to registration algorithms working solely with the original images.

There is another advantage of using this approach. Since this type of image decomposition usually involves low-frequency smoothing, this regularizes the registration problem thus yielding better convergence properties of various iterative search techniques such as variations of gradient descent. This may lead to improved accuracy especially when the initial transformation is far from the solution.

In the remainder of Section II, we describe the choices we made regarding some of Brown's criteria.

B. Search Space

There are numerous types of spatial transformations including both global (e.g. rigid, affine, and perspective), where all pixels are displaced according to a single rule, and local where displacement of a pixel depends on its location in the image [24]. When accurate registration is required, the process is split into two steps. First,

a global transformation is determined that takes into account most of the warping between the images being aligned. Then, the result is refined by computing local displacements. These two stages require very different types of registration algorithms [1], and so in this paper, we only focus on the first stage. More specifically we consider two variations of the RST transform (1). One excludes isometric scaling while the other uses the full transform with the four parameters.

C. Feature Space

There are many different types of information in the images that can be used for registration, including original intensities, edges, contours, wavelet coefficients, moment invariants and higher level features [1, Section 4.1]. Some of them have been tested as part of a group effort on image registration at NASA Goddard Space Flight Center [3], [4], [19], [21], [23], [25]. When choosing a feature space, one attempts to satisfy several, sometimes contradictory, requirements. The features must preserve important information in the images while suppressing artifacts such as noise. They should also be spatially invariant to radiometric and atmospheric variations. Finally, their extraction and application to registration should be computationally optimal and should be adapted to the multiresolution approach described in Section II-A.1.

In our work, we have focused on wavelet and wavelet-like features. Their appeal was two-fold. First, wavelet pyramids provide a natural multiresolution framework. Second, band-pass or high-pass frequency filtering involved in the wavelet decomposition of an image tends to emphasize high-contrast features in satellite images, such as roads, buildings and coastlines. However, we will demonstrate in Section IV that not all wavelets or wavelet-like features provide similar results. We will show that one property of a wavelet pyramid that is particularly essential to good registration is rotation- and shift-invariance. Section IV-A describes several pyramids that we selected for comparison and results of our extensive comparative experiments are presented in Section IV-C.

D. Search Strategy

The choice of a search strategy is influenced mostly by the search space, as well as by other factors such as the purpose of the registration algorithm under development. The focus of this work is to compare different feature spaces and to evaluate the results first from an accuracy and second from a speed points of view. In addition, we do not want computational aspects of a sophisticated search technique to affect results of our study. Therefore, for the tests described in Section IV the simplest possible approach, an exhaustive search combined with cross-correlation is chosen as the search strategy. It amounts to varying each transformation parameter within its respective range, determined either manually or automatically, and selecting the combination of parameter values that yield the largest value of the similarity metric (e.g cross-correlation).

The main problems of the exhaustive search are its high computational cost and limited accuracy. Even for the four parameters of the full RST transform, the computational requirements become prohibitively expensive. In addition, different parameters (e.g. shift and scaling) have very different effect on the warping of the image, and so it is often difficult to choose appropriate discrete meshes that guarantee high accuracy. Therefore, in Section V, where the focus is to develop a practical registration algorithm, a more sophisticated search scheme is selected, based on a gradient descent approach.

In the latter experiments, an optimization algorithm developed by Thévenaz et al. [5], which we denote by the acronym TRU, is chosen for its several appealing properties. First, the algorithm is based on a modified version of the Levenberg-Marquardt algorithm which represents a hybrid optimization approach between a pure gradient-descent method and a more powerful but less robust Gauss-Newton method. Second, the method is implemented in a multiresolution fashion. In addition, it was successfully applied to various types of medical imagery, and one of the goals of our research is to evaluate its performance when applied to remotely sensed imagery.

III. PERFORMANCE ASSESSMENT

Two issues arise when evaluating any image registration algorithm: first, the type of images chosen to test the algorithm; second, the method selected to assess the registration accuracy. In this section, we address both of these issues.

A. Creating Test Images with Exact Ground Truth

Since our goal is to create an algorithm applicable to many types of satellite data, test data sets should reflect this by including imagery from different platforms, with different spatial and spectral resolutions, taken at various dates. The disadvantage of such data is that in the majority of cases, ground truth, if available at all, is approximate at best. Therefore in our experiments, we also use synthetic images created by warping a given "source" image by a predetermined transformation. We use two types of synthetic data in this paper.

1) Same-Radiometry Synthetic Images: The first type of synthetic imagery is created in three steps (see the left subplot in Figure 2). First, starting with a large source image, a smaller sub-image is extracted from its center, which becomes the reference image. Second, for a given set of RST transformation parameters, the same large source image is warped. Third, a small sub-image is extracted from the warped result. This sub-image becomes the input and the warping transformation becomes the ground truth. We may also add a predetermined amount of noise to the input. This optional operation is not shown as a separate step in Figure 2. A registration algorithm being evaluated is required to recover the ground truth parameters when applied to the reference-input pair. Its performance is assessed using the accurate error measure described in
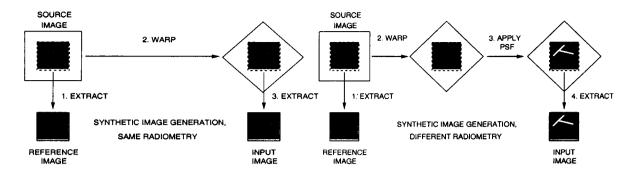
Fig. 2: Synthetic Image Generation: Same radiometry (left subplot) and Different radiometries (right subplot) figure

Section III-B. The main weakness of this approach is that the reference and input are essentially the same radiometrically, which makes them very simple data to register, while in reality images are rarely so similar.

2) Different-Radiometry Synthetic Images: In order to make synthetic test data more sophisticated, the framework described in Section III-A.1 is augmented by an extra step that makes the reference and input images radiometrically different. In real satellite imagery, such differences occur because different instruments "see" a scene differently. We would like to emulate this process by considering an image after geometric warping as an "actual scene" and by mimicing how an instrument would process this scene. To do this the warped image is convolved with a point-spread function [26]. The PSF may or may not correspond to a specific sensor, but it is very important that it does not introduce any additional geometric warping of the image. The testing framework is shown in the right subplot of Figure 2. The first two steps of the proposed procedure are the same as before. What is different is that, before extracting the sub-image that would become the input, the geometrically warped image is convolved with the chosen PSF. In this paper we focus on the case when there is a relatively small amount of radiometric variation between reference and input images. For this reason, we use a simple PSF that was constructed by convolving with itself a 512×512 image that was "black" except for the 5×5 "white" center. A similar approach for synthetic image generation was used in [27], [28] where Gaussian point-spread functions were applied. This general approach can potentially be used to synthesize multi-sensor satellite data with strong differences in radiometry, provided there exists an appropriate geometry-preserving PSF that models these differences.

B. Measuring Accuracy

When accurate ground truth is available, such as when test images are created synthetically, a standard way of assessing registration accuracy is by using the RMS error. Suppose we are given a ground truth (GT) transformation $p_{GT} = (t_{x_1}, t_{y_1}, \theta_1, \kappa_1)$ and a computed transformation $p = (t_{x_2}, t_{y_2}, \theta_2, \kappa_2)$ with the

corresponding RST matrices $Q_{p_{GT}}$ and Q_p defined by (1). To compute the error, we first need to determine the "error" transformation $p_\epsilon = (t_{x_\epsilon}, t_{y_\epsilon}, \theta_\epsilon, \kappa_\epsilon)$ that represents the discrepancy between p and p_{GT}. The corresponding RST transform matrix is $Q_{p_\epsilon} = Q_p \, Q_{p_{GT}}^{-1}$ that yields

$$\kappa_\epsilon = \frac{\kappa_2}{\kappa_1}, \quad \theta_\epsilon = \theta_2 - \theta_1, \quad t_{x_\epsilon} = t_{x_2} - \kappa_\epsilon(t_{x_1} \cos\theta_\epsilon + t_{y_1} \sin\theta_\epsilon), \quad t_{y_\epsilon} = t_{y_2} - \kappa_\epsilon(t_{y_1} \cos\theta_\epsilon - t_{x_1} \sin\theta_\epsilon). \quad (2)$$

Now let $(x, y) \in \Omega$ and let $[x', y', 1]^T = Q_{p_\epsilon} [x, y, 1]^T$. This can be equivalently rewritten as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \kappa_\epsilon \begin{pmatrix} \cos\theta_\epsilon & \sin\theta_\epsilon \\ -\sin\theta_\epsilon & \cos\theta_\epsilon \end{pmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_{x_\epsilon} \\ t_{y_\epsilon} \end{bmatrix}. \quad (3)$$

Then the RMS error may be defined as

$$E_0(p_\epsilon) \equiv \sqrt{\frac{1}{Area(\Omega)} \int_\Omega (x' - x)^2 + (y' - y)^2 \, dx \, dy} \quad (4)$$

$$= \sqrt{\frac{1}{(A-1)(B-1)} \int_0^{A-1} \int_0^{B-1} (x' - x)^2 + (y' - y)^2 \, dx \, dy}. \quad (5)$$

Substituting (3) and (2) into (5), we obtain the exact expression for $E_0(p_\epsilon)$ corresponding to the RST transform

$$E_0(p_\epsilon) = \frac{\sqrt{3}}{6} \sqrt{12(t_{x_\epsilon}^2 + t_{y_\epsilon}^2) + \alpha(\kappa_\epsilon^2 + 1 - 2\kappa_\epsilon \cos\theta_\epsilon)},$$

where $\alpha \equiv A^2 + B^2 - 2(A + B) + 2$. However, this formula needs to be modified. Suppose p_ϵ' is the RST transform inverse to p_ϵ. Individual parameters of p_ϵ' may be determined explicitly from those of p_ϵ by inverting Q_{p_ϵ}. Warping an image F with respect to an image G by p_ϵ is equivalent to warping G with respect to F by p_ϵ'. Therefore the errors produced by p_ϵ and p_ϵ' should be the same, i.e. we expect $E_0(p_\epsilon) = E_0(p_\epsilon')$. Instead, we have $E_0(p_\epsilon) = \kappa_\epsilon \, E_0(p_\epsilon')$. The reason for this is that unlike rotation or shift, scaling actually shrinks ($\kappa_\epsilon < 1$) or expands ($\kappa_\epsilon > 1$) the domain Ω by the factor κ_ϵ, and so to make the two errors identical, we would need to replace Ω with $\kappa\Omega$ in (4). We propose to modify the error formula by averaging $E_0(p_\epsilon)$ and $E_0(p_\epsilon')$ on the log scale, defining

$$E(p_\epsilon) \equiv \frac{1}{\sqrt{\kappa_\epsilon}} E_0(p_\epsilon) = \frac{1}{\sqrt{12\kappa_\epsilon}} \sqrt{12(t_{x_\epsilon}^2 + t_{y_\epsilon}^2) + \alpha(\kappa_\epsilon^2 + 1 - 2\kappa_\epsilon \cos\theta_\epsilon)}, \quad (6)$$

that yields $E(p_\epsilon) = E(p_\epsilon')$. We use (6) throughout the paper to measure registration accuracy for the cases when ground truth is known precisely. When ground truth is only known approximately, instead of measuring accuracy, consistency is used to assess the response of a given algorithm with respect to different execution parameters such as the number of pyramid levels or the initial guess.

IV. COMPARISON OF WAVELETS COMBINED WITH EXHAUSTIVE SEARCH

The main goal of this paper is to compare performance of several types of wavelet-based features spaces applied to image registration. In this section, we select four different wavelet pyramids and perform a first

round of experiments using a simple search strategy and simple test data. After a brief review, in Section IV-A, of various types of pyramids that appeared in the literature, and the pyramids selected for testing, Section IV-B describes the exhaustive search and the cross-correlation metric used in the experiments, whose results are presented in Section IV-C. Here we limit the search space to combinations of rotations and translations. The results show that one of the pyramids is clearly inferior to the others. We discard it and in Section V, we perform a second round of experiments designed to test the remaining wavelet pyramids using the full RST transform, a more practical algorithm and more sophisticated test data.

A. Wavelet Pyramids

The literature on wavelets and their applications to signal and image processing is enormous, and many different decomposition algorithms have been proposed. While selecting a few of them for comparison, we impose the following requirements to narrow down our search. *First*, our goal is to register a reference image to an input image that is geometrically transformed with respect to the reference. The multi-resolution search strategies selected in this paper assume that an image that represents the reference at a given pyramid level must be geometrically warped and compared to the corresponding representation of the input. Therefore it is essential that the order of decomposition and warping could be interchanged. More specifically, since the search space is restricted to RST transformations, it implies that the chosen wavelet pyramids be shift- and rotation-invariant. *Second*, a pyramid must be implemented efficiently, i.e. an image representation must be computationally inexpensive to generate. *Third*, the resulting representation should also be efficient when used in the registration phase. Some pyramids consist of several image representations at each resolution. The search strategies that we employ in this paper, and which will be explained later, imply computation of pixel-by-pixel difference of two images at each iterations. Therefore doubling the number of sub-images per level doubles the computational cost of a single iteration, and so we generally prefer single sub-image pyramids over multiple sub-image ones.

Previously [20], [21], [23], we have extensively experimented with separable orthogonal wavelets developed by Daubechies [29]. The pyramid is shown in Figure 3a. Given IM_n, an approximation of the original image at pyramid level n (with IM_0 being the original image), it is first processed and subsampled by columns (L_C, H_C) and then by rows (L_R, H_R) by a low-pass filter and a high-pass filter. We obtain four sub-images with half the spatial resolution of IM_n. The sub-image LL_{n+1} represents a compressed and smoothed version of IM_n, and it is used as input at the next pyramid level. The images LH_{n+1} and HL_{n+1} contain features and these are the sub-images that we use in this paper. Finally, HH_{n+1} is a high-frequency subband that contains mostly noise. Since they are orthogonal, Daubechies wavelets are computationally efficient, but for the same reason, they also have poor invariance properties. As an image is shifted, energy shifts both within

| L_C, L_R | = Low-pass with Decimation |
| H_C, H_R | = High-pass with Decimation |

(a)

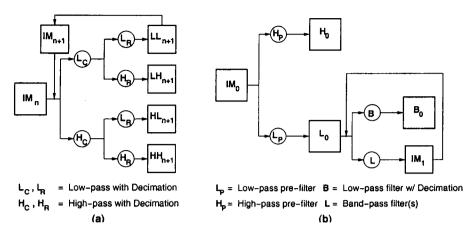| L_p = Low-pass pre-filter | B = Low-pass filter w/ Decimation |
| H_p = High-pass pre-filter | L = Band-pass filter(s) |

(b)

Fig. 3: Daubechies Orthogonal (a) and Simoncelli Steerable (b) Pyramids

figure

and across subbands [30]. Although, it was shown [23] that, when combined with correlation as similarity measure, orthogonal wavelets can still provide fairly accurate registration results, that study involved only translation. Daubechies orthogonal wavelets are included in the first part of our study to demonstrate that rotation-invariance, as well as wavelet shiftability [30] are essential for producing accurate registration results, especially in the presence of noise in the original data.

Several approaches have been proposed that attempt to overcome the deficiencies of orthogonal wavelets. In [30], the Steerable Pyramid is proposed, that enables one to build translation- and rotation-invariant filters by relaxing the critical sampling condition of the wavelet transforms. As the result, an overcomplete invertible wavelet representation is obtained. The Steerable Pyramid is shown in Figure 3b. First, the original image IM_0 is preprocessed by a high-pass pre-filter H_P and a low-pass pre-filter L_P. The resulting two images are of the same resolution as IM_0. Then, the low-pass image is further processed by a band-pass filter B and a low-pass filter L. Note that although in Figure 3b, only a single band-pass filter is shown, it is possible to have an arbitrary number of them, each emphasizing image features oriented in a specific direction. If there are k band-pass filters, then the pyramid is $4k/3$ overcomplete. Although in certain cases, it may be beneficial to use more than one oriented filter [31], our study is limited to $k = 1$ to reduce computational and memory requirements. In order to ensure shift invariance, the output of the high-pass pre-filter and of the band-pass filter(s) are not sub-sampled. As the result, the Steerable Pyramid produces a representation of an image composed of two multiresolution series of components, the low-pass $\{L_0, L_1, \ldots\}$ and the band-pass $\{B_0, B_1, \ldots\}$ series where both L_n and B_n are 2^{-n} the original size. Downsampling of the Steerable Pyramid was slightly modified to remove "shift bias" caused by filter nonsymmetries [32].

An efficient shift-invariant pyramid image representation based on polynomial splines was designed by Unser

and his colleagues [14], [33]. It has several attractive properties. First, it minimizes least-square difference between successive image representations in the pyramid. In terms of image registration, such optimality ensures that accurate approximations to the final geometric transformation can be recovered at coarse pyramid levels. Second, the pyramid is implemented efficiently using simple linear filters derived from the recursive B-splines. Starting with the original image, a pyramid $\{S_1, S_2, \ldots\}$ is produced by recursive anti-aliasing prefiltering followed by decimation by a factor of two. Third, it can be tied neatly to the cubic spline interpolation used at a given pyramid level during registration, for geometric warping and exact computation of derivatives of the objective function [5]. Fourth, it is based on the spline theory whose theoretical and practical aspects have been thoroughly developed since the original work of Schoenberg [34].

Other approaches that propose translation- and rotation-invariant wavelets are described in [35], [36]. In [35], the method involves "averaging out the translation dependence": for the de-noising application, this consists in shifting the data for a range of shifts, de-noising the shifted data and then un-shifting the data. A similar method could be devised for image registration, but it would considerably increase the computational complexity of the process and might affect the accuracy of the registration. In [36], Magarey and Kingsbury combined the efficient discrete wavelet transform with complex-valued Gabor-like filters, that have nearly optimal localization, to produce a pyramid with approximate shift-invariance. At each level of their pyramid, called CDWT, six complex-valued sub-images are produced from the original image using equally-spaced oriented filters. Although this scheme might be more reliable than most other wavelets from a translation-invariant point of view, it might also significantly increase the computation time.

For the reasons stated above, in addition to the Daubechies orthogonal wavelets, we chose to test both Simoncelli low-pass pyramid and band-pass pyramids because they appear to provide the best balance of the three requirements that are stated at the beginning of this section. They both have good invariance properties, are relatively inexpensive to compute, and both contain a single image per pyramid level. We also include the centered spline pyramid because of similar properties of invariance and computational speed. In addition, since it was included in the original gradient-search implementation [5], the spline pyramid may be used as a benchmark in the second part of our study. Throughout the rest of the paper, we use the symbols Daub, SimB, SimL and SplC to denote the Daubechies, Simoncelli band-pass, Simoncelli low-pass and centered-spline pyramids, respectively.

B. Search Strategy and Similarity Metric

1) Exhaustive Search: The principle of the exhaustive search is described first when the search space is composed of 2d rotations and then generalized to combinations of rotations and translations [21]. Assuming that the search strategy follows the multiresolution approach provided by the wavelet decomposition, the initial

search space is either defined a priori or specified by the user. At the highest level of decomposition, the search is exhaustive over the whole search space but with an accuracy equal to Δ. The first approximation of the best rotation, R_N, is chosen over this search space; then R_N becomes the center of a new search interval of length 2Δ, $[R_N - \Delta, R_N + \Delta]$, and at the next lower level, the new approximated rotation, R_{N-1}, is found within this search interval with an accuracy of $\Delta/2$. This process is repeated until the first level of wavelet decomposition, where the search interval is $[R_2 - \Delta/2^{N-2}, R_2 + \Delta/2^{N-2}]$ and the final registration rotation, R_1, is found with an accuracy equal to $\Delta/2^{N-1}$. In particular, if δ is the desired registration accuracy, Δ is chosen as $2^{N-1}\delta$, where N is the number of levels of wavelet decomposition. In some cases, for reasons of robustness, even if the final desired accuracy δ does not require Δ to be small at the lowest resolution level, smaller accuracy steps can be considered for the initial step, and then the previous process is applied starting at level $N - 1$, once the initial approximation of the transformation has been computed. This strategy avoids pursuing a false path in the search for the optimal transformation.

Similarly, in the case when both rotations and shifts are considered, the search is performed simultaneously on the parameters, rotation angle, shift in x-direction and shift in y-direction. To reduce the amount of computations when the transformations are small, the search over all parameters could be decomposed at intermediate levels into several searches in the complementary sub-spaces, sub-space of translations and sub-space of rotations. At each intermediate level of decomposition, for all combinations of parameters, all parameters but one are assumed to be known and the remaining parameter is refined. Nevertheless, in our experiments, since all values of parameters are tested, including large transformations, a full exhaustive search will be utilized.

2) Normalized Cross-Correlation: Normalized cross-correlation has been used extensively in various image processing tasks [37], especially in template matching. Given two images $F(x, y)$ and $G(x, y)$, the normalized correlation coefficient is defined as

$$C(F(x,y), G(x,y)) = \frac{\sum_{i,j}(F(x_i, y_j) - \overline{F})(G(x_i, y_j) - \overline{G})}{\sqrt{\sum_{i,j}(F(x_i, y_j) - \overline{F})^2 \ \sum_{i,j}(G(x_i, y_j) - \overline{G})^2}}, \quad (7)$$

where (x_i, y_j) are (integer) pixel coordinates and $\overline{F}$ and $\overline{G}$ are the mean intensities of the two images. The coefficient always satisfies $-1 \leq C(F, G) \leq 1$ with its values close to 1 and -1 corresponding to, respectively, strong positive and negative correlation between the images, while values close to 0 implying weak correlation. In our work, however, we use the absolute value of the expression (7). The reason is that we register multisensor images that may have very different radiometric characteristics. Consider, for instance, registering an image to its copy with inverted intensities. Therefore using the absolute value allows us to register two images with strong negative correlation that are actually very well aligned.

C. Numerical Experiments

In this section, only synthetically generated imagery was used for testing. The experiments with synthetic data consisted of two parts. First, sensitivity of the four pyramids was studied relative to large rotations and shifts, in the absence of noise. Second, their noise sensitivity was assessed.

In the first study, same-radiometry synthetic imagery is generated as described in Section III-A.1. The source is a 1024×1024 image extracted from Band 4 of a Landsat Thematic Mapper (TM) scene of a Pacific NorthWest region. Warping is performed using nearest neighbor interpolation with θ varying between 0 and 90 degrees and t_y between 0 and 20 pixels while keeping t_x at zero. Reference and input images are of size 512×512. They are decomposed using the four pyramids. In order to compare images of similar sizes, the top levels of either SimB or SimL are not used (see Table IV-C). The initial shift and rotation range for the exhaustive search is determined at the coarsest pyramid level from the "ground truth" values as follows: rotation is varied within -5 to 5 degrees of the true value while shift is varied within -2 to 2 coarse pixels of the scaled and rounded true value. Thus, if the GT transformation is $(t_x, t_y, \theta) = (0, 9, 4)$ then at the coarsest pyramid level, θ varies between -1 and 9, t_x between -2 and 2 and t_y between $round(9/8) - 2 = -1$ and 3. Performance is assessed using the RMS error defined by (6). Its surfaces corresponding to the four pyramids are plotted in Figure 4. Examining the results, we observe that when rotation is small or average, all pyramids produce no error. However, performance of the Daubechies pyramid eventually deteriorates as rotation is increased while that of the other pyramids does not.

In the second study, θ is varied between 0 and 9 degrees and t_y between 0 and 20 pixels. In order to produce input images, we again use the same-radiometry framework. This time, however, after warping, we also add a controlled amount of Gaussian white noise to the warped image. This amount is determined by the signal-to-noise ratio between $-30dB$ and $20dB$. The SNR of β dB is defined as $\beta = 10\log_{10}(\mathrm{Var(Image)}/\mathrm{Var(Noise)})$. The decomposition and registration steps are identical to those for the noiseless data. In this case, however, in order to produce three-dimensional error plots, we average the RMS error over all ground truth rotations. The resulting surfaces are shown in Figures 5 and 6. As in the previous study, Daub performed significantly worse than the other three pyramids. It yielded correct results only for SNR values of about $16dB$ and above while the "breakdown" noise values in the other cases were below $-10dB$. SimL and SplC performed almost identically and were better than SimB, which is expected since low-pass filtering generally has better noise-suppressing capabilities than band-pass filtering.

V. COMPARISON OF WAVELETS COMBINED WITH GRADIENT-BASED SEARCH

We now continue our comparison of wavelet pyramids. Based on the results of the previous section, we exclude the Daubechies pyramid from further tests as, for image registration purposes, it is clearly inferior to

	Pyramids			
	SimB	**SimL**	**Daub**	**SpIC**
Decomposition Levels	{B1,B2,B3}	{L1,L2,L3}	{LH1,LH2,LH3} {HL1,HL2,HL3}	{S1,S2,S3}

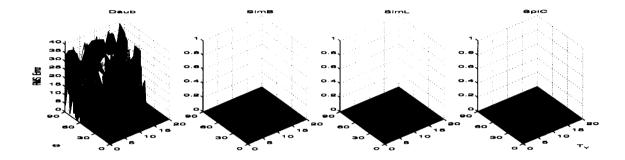TABLE I: Pyramid Levels (Decimation of {2,4,8}) Used in Section IV-C

table

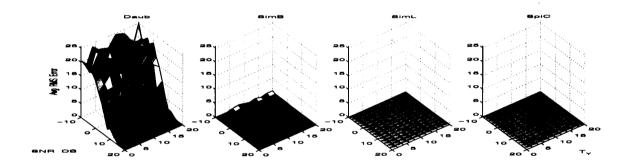Fig. 4: *Shift/Rotation Sensitivity:* Surfaces of $E(p_t)$ (see (6)) for Noiseless Data

figure

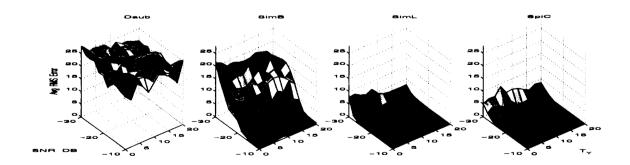Fig. 5: *Noise Sensitivity, Small Noise:* Surfaces of $E(p_t)$ (see (6)), Averaged over Rotations

figure

Fig. 6: *Noise Sensitivity, Large Noise:* Surfaces of $E(p_t)$ (see (6)), Averaged over Rotations

figure

the spline and Simoncelli pyramids . At the same time, we consider a more general search space, combine the pyramids with a more sophisticated and practical search strategy and apply them to more complex test data, both synthetic and "real-life", than those in Section IV.

A. Description of the TRU algorithm

In this section we give an overview of the TRU algorithm. As we pointed out in Section II-D, it is a robust gradient-based optimization algorithm implemented in multiresolution fashion and successfully applied to 2d and 3d medical imagery. Below we provide some additional details about TRU.

1) Modified LM Optimization: Its main component is a modified version of the Levenberg-Marquardt (LM) method [38] for nonlinear unconstrained least-squares optimization, that the authors call ML*. It minimizes

$$\chi^2(p) = \frac{1}{N} \sum_{i=1}^{N} (F_R(T_p(x_i, y_i)) - F_I(x_i, y_i))^2,$$

where N is the number of relevant pixels and (x_i, y_i) is the (integer) coordinate of each pixel. Since the search space contains affine transformations and its simpler variations the standard normal equations can be rewritten in such a way as to avoid computation of the first and second derivatives at each iteration. This is the basic difference between the standard LM method and ML*. Assuming the cubic spline model of the image, that represents an image $I(x, y)$ as smooth when (x, y) is off the standard pixel lattice, all derivatives involved in ML* are computed exactly. This is essential because while being significantly faster than standard LM, ML* requires higher accuracy of the objective function and its derivatives to achieve convergence. Its radius of convergence may be smaller than that of LM.

2) Multiresolution Pyramids: The original TRU algorithm is implemented in a coarse-to-fine fashion provided by the spline pyramid that we described in Section IV-A. Throughout the rest of the paper, we denote this version by TRU-SplC. We also combine ML* with the two Simoncelli pyramids and denote these versions of TRU by TRU-SimL and TRU-SimB, respectively. The goal of this part of our study is to compare their performance on different types of satellite data.

There are certain advantages and disadvantages of using a Simoncelli pyramid as opposed to the spline pyramid. On one hand, the band-pass pyramid may provide more accurate results because it combines noise-removing properties of a low-pass filter with high-frequency extraction of important edge- and contour-like features. In addition, both Simoncelli pyramids provide an option of using something other than the original image at the finest resolution level for the final adjustment of the answer. On the other hand, they are rather expensive to compute, especially in the case of the band-pass pyramid, for which the low-pass filtering must also be done.

In this section, in addition to rotation and shift, we also optimize for isometric scaling. One of the features that distinguishes multisensor satellite data from medical data, for which TRU was originally designed, is the large difference in scale between the reference and the input. For instance, Landsat Enchanced Thematic Mapper Plus (ETM+) images have a spatial resolution of 30 meters compared to 4 meters for IKONOS. Since convergence of TRU is local, we use pyramid downsampling to approximate scale with its subsequent fine-tuning using TRU. In other words, if $F_R(x, y)$ and $F_I(x, y)$ have spatial resolutions r_R and r_I with $r_R/r_I = 2^L \kappa_1$, where L is a positive integer and $\kappa_1 \approx 1$, we can downsample the reference by a factor of 2^L and then apply TRU to the new image pair trying to recover κ_1. If the transformation between the downsampled reference and the original input is found to be $(t_x, t_y, \theta, \kappa_2)$ then the transformation between the original pair is simply $(t_x, t_y, \theta, 2^{-L} \kappa_2)$.

3) Implementation Details: We use a C implementation of TRU [39] that allows one to process large images which would be impossible if the algorithm were implemented in a higher-level language such as Matlab. At the same time, it is easily customizable. In particular, individual transformation parameters can be easily excluded from the optimization process, therefore transformations as general as affine or as simple as translation can be tested.

B. Numerical Experiments

To test the algorithms, we used two types of imagery, synthetic and real. Synthetic data, described in Section V-B.1, consisted of same-radiometry imagery with added noise, just like in Section IV-C, and different-radiometry imagery without noise. The first real test dataset, described in Section V-B.2, represents images acquired by four different satellite imaging platforms over a single EOS Land Validation Core Site used for calibration and validation of the MODIS instruments [40], while the second dataset, described in Section V-B.3, was acquired by only two platforms, but over four different Core Sites representing different types of terrain.

1) Synthetic data: The purpose of the synthetic data tests was to evaluate sensitivity of TRU to noise, radiometric variations and initial guess. As the source (ref. Figure 2) we used the same Landsat-TM image as in Section IV-C. The reference and inputs were made of size 256×256. We *first* describe the same-radiometry experiments. To make the test more statistically sound, we wanted to vary as many transformation parameters as possible while keeping the number of experiments reasonable as well as being able to present results pictorially. Therefore we decided to vary two parameters, α that would represent both shift and rotation, and n that would represent the signal-to-noise ratio. We varied α between 0 and 8 and n between $-30dB$ and $20dB$. We fixed the value of $\kappa_{GT} = 0.95$, and for each value α_i, we warped the source according to the "ground truth" vector $p_{GT} = (t_x, t_y, \theta, \kappa) = (\alpha_i, \alpha_i, \alpha_i, \kappa_{GT})$, where rotation was measured in degrees. Various input

images were then created by adding to the warped image different amount of white Gaussian noise determined by values n_j of the SNR parameter. The three variations of TRU were tested on each reference-input pair using the trivial initial guess $p_0 = (t_{x_0}, t_{y_0}, \theta_0, \kappa_0) = (0, 0, 0, 1)$ and three pyramid levels, including the finest resolution. A final transformation p computed by a TRU run was used together with p_{GT} to compute the corresponding error transformation p_ϵ, and the error $E(p_\epsilon)$ was computed by (6) to assess performance.

The test results are shown in the left subplot of Figure 7. The shaded regions in each subplot correspond to those pairs $\{\alpha_i, n_j\}$ for which the resulting error exceeded a threshold value of 1.0. We observe that TRU-SimB was the most consistent among the three variations of TRU in the sense that for small displacements, its performance improved as less noise was added, while TRU-SplC and TRU-SimL actually failed on some small-noise cases while converging on other, larger-noise, cases with the same amount of geometric displacement. When the algorithms did converge to a sub-pixel level, they all produced comparable final $E(p_\epsilon)$.

The *second* step is to describe the different-radiometry experiments without noise. We used a similar way of generating input images. This time, however, both parameters that we varied were geometrical, namely, α that represented both shift and rotation, and κ that represented isometric scale. We varied α between 0 and 20, and κ between 0.8 and 1.25. For each pair $\{\alpha_i, \kappa_j\}$, we created a ground truth vector $p_{GT} = (\alpha_i, \alpha_i, \alpha_i, \kappa_j)$. An input image was then generated in two steps. First, the source image was warped according to p_{GT}. Then, the warped image was convolved with a PSF (see Section III-A.2). No noise was added to the inputs. The 256×256 center of the resulting image became the input. As before, the three variations of TRU were then applied, using the trivial initial guess, to the given reference-input image pair and the resulting RMS errors were recorded. The results are shown in the right subplot of Figure 7. Again, the threshold value of 1.0 was used for $E(p_\epsilon)$. We observe that convergence regions of TRU-SplC and TRU-SimL are considerably larger than that of TRU-SimB, so they may be more appropriate to find a reasonably accurate results when misregistration between two images is large. However, when we inspected the average final errors produced by the three algorithms when they converged (see Table II), we found that TRU-SimB generated significantly more accurate results. It appears from these tests, as well as others presented in [41], that TRU-SimB is less sensitive to the type of radiometric variations introduced by a simple PSF, than TRU-SplC or TRU-SimL.

	Number of converged	Median converged $E(p)$	Mean converged $E(p)$
TRU-SplC	1715 / 9801 $\approx$ 17.5%	0.326289	0.108127
TRU-SimB	718 / 9801 $\approx$ 7.3%	0.0656085	0.0325503
TRU-SimL	1659 / 9801 $\approx$ 16.9%	0.403861	0.115943

TABLE II: Different-radiometry SIG: Average $E(p_\epsilon)$ for Converged (Unshaded) Region in Figure 7
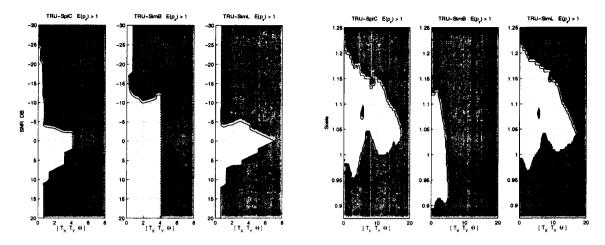
table

Fig. 7: $E(p_\epsilon) > 1$ Regions (Shaded) for Same-Radiometry Noisy Data (left) and Different-Radiometry Noiseless Data (right)

figure

	Imaging Platform			
	IKONOS	ETM+	MODIS	SeaWiFS
Bands	3 (RED). 4 (NIR)	3 (RED). 4 (NIR)	1 (RED). 2 (NIR)	6 (RED). 8 (NIR)
Resolution	4m (3.91m)	30m (31.25m)	500m	1000m
Size (row × col)	2048 × 2048	5760 × 7552	392 × 776	196 × 388

TABLE III: CORE1 Images

table

2) EOS Land Validation Core Site Data, Multiple Platforms: The purpose of the next two tests was to compare the algorithms on a "real-life" dataset. Our primary goal was to investigate sensitivity of convergence to the initial guess. For this dataset, which we denote by CORE1 and whose properties are summarized in Table III, only approximate ground truth was known. The images were taken by four satellites over the Konza Prairie located in the state of Kansas. The IKONOS and ETM+ images were resampled to respective resolutions of 3.91 and 31.25 meters. After some preprocessing, we assembled a set of eight images, that are shown in Figure 8.

From the Earth Science standpoint, it would not make much sense to register images with extremely large differences in spatial resolutions, e.g. IKONOS to SeaWiFS. Therefore we adopted a "cascaded" approach to testing by registering IKONOS to ETM+, ETM+ to MODIS, and MODIS to SeaWiFS. In addition, Red/NIR bands from one imaging platform were registered to the corresponding Red/NIR bands from another platform. We thus obtained six test image pairs. In [42] we used these images to evaluate several registration algorithms, and results obtained during this evaluation were used as approximate ground truth in our study (see Table IV).
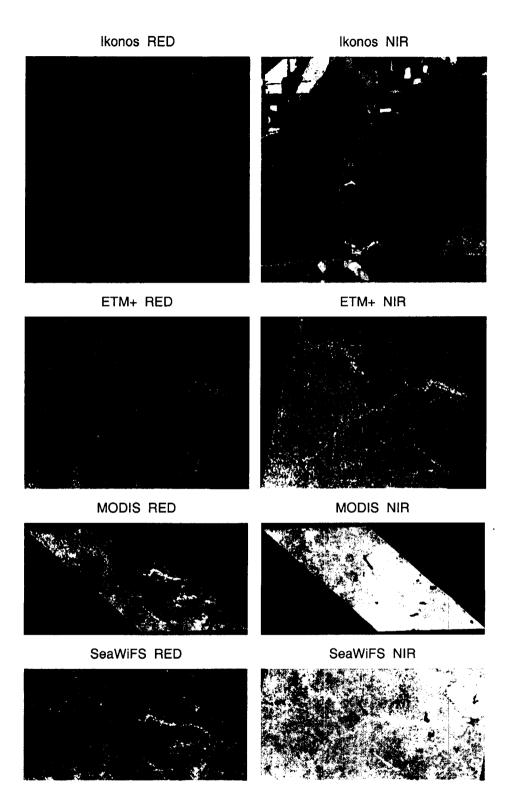
Ikonos RED

Ikonos NIR

ETM+ RED

ETM+ NIR

MODIS RED

MODIS NIR

SeaWiFS RED

SeaWiFS NIR

Fig. 8: CORE1 Images

figure

	Ref = IKONOS RED/NIR Inp = ETM+ RED/NIR	Ref = ETM+ RED/NIR Inp = MODIS RED/NIR	Ref = MODIS RED/NIR Inp = SeaWiFS RED/NIR
$p_{GT} \approx [t_x, t_y, \theta, \kappa]$	$[2, 1, 0, 0.125]$	$[-2, -4, 0, 0.0625]$	$[-8, 0, 0, 0.5]$

TABLE IV: CORE1: Approximate Ground Truth

table

We applied the three variations of TRU to the same six image pairs as follows. First, using one of the pyramids, a reference image, which is always of higher resolution than the corresponding input, was downsampled by a factor 2^L to approximately equalize resolutions. For instance, in IKONOS-ETM tests $2^L = 8$ while in ETM-MODIS tests $2^L = 16$. Then TRU combined with the same pyramid was applied to the new reference-input pair, using three levels of decomposition. The search space consisted of rotation, shift and scale. The resulting scale was then multiplied by 2^{-L} to obtain the final value. In all cases, we varied the shift values in the initial guess from zero to the values given in Table IV.

In addition to optimizing for shift, rotation and scaling, the TRU algorithm can adjust a grey-level intensity factor which can account for certain types of radiometric differences between images. Ironically, however, including this parameter in optimization often leads to failures even on simple data. The reason for this is that a reduction in the value of the objective function, which measures energy between misaligned images, can be achieved by pushing intensities of one of the images to zero [43]. Since the intensity parameter appears in the optimizer in exponential form, this can cause TRU to endlessly iterate decreasing this parameter while not adjusting the geometry at all. Another reason for switching this parameter off during optimization was that we wanted to see if the three pyramids could actually be used to handle radiometric differences.

Tables V and VI show results of IKONOS-ETM and ETM-MODIS tests, respectively, using Red bands. Note that in these tables, both the initial scale value κ_0 and the final κ are scaled by the appropriate reference-downsampling factor 2^{-L}. These tables show that the three variants of TRU can behave quite differently depending on the given test data. In the case of IKONOS-ETM registration (Table V), all three algorithms gave consistent, although slightly different, results regardless of the initiall guess. In all three cases, the results were quite close to the approximate GT. In the case of ETM-MODIS registrations (Table VI), TRU-SimB also gave consistent and accurate results essentially invariant to the starting point. Results of TRU-SplC varied noticeably more and appeared to deteriorate as the initial guess moved farther away from the ground truth. TRU-SimL gave acceptable results when started close to GT but completely failed in the other cases. Since presenting all registration results for the CORE1 dataset in the format of Tables V and VI would take too much space and would be difficult to follow, we summarize our findings in a more compact form, shown in Table VII. For each image pair and each initial guess p_0, we determined approximate RMS registration

error by using transformations given in Table IV as exact ground truth and applying (6). We then calculated the mean, median and standard deviation of the approximate error over all initial guesses. The mean/median values are used to detect failure of an algorithm, as in the case of TRU-SimL applied to the ETM-MODIS NIR image pair. Both closeness of mean to median and small deviation indicate consistency of an algorithm and its sensitivity to p_0. We observe that TRU-SimB gives the best overall performance while TRU-SimL is the worst completely breaking down in several cases. TRU-SimB converged well even from the trivial starting point. This is important since when registering real data, the trivial guess is used most often, especially when little ground truth information is available. The TRU-SplC algorithm also performed well in many cases. However, its results were less consistent than those of TRU-SimB and tended to break down closer to the solution than TRU-SimB.

We also observed that, when converged, TRU-SplC tend to outperform the other two algorithms in terms of the number of iterations, especially at the finer resolutions. This may be caused by the fact that B-spline interpolation used to transform the images is matched by the spline pyramid filters, but not matched by the Simoncelli filters.

Out of the six image pairs only the two IKONOS-ETM pairs did not require masking of the images and their representations at different pyramid levels. While eliminating "border effects" caused by filtering of edges of the original mask, at each decomposition level, progressive masking also removed regions which may have contained strong features. Since filter size is independent of decomposition level, at coarse levels rather significant portions of images are masked out. We believe this to be the primary reason for large variations in results of ETM-MODIS and MODIS-SeaWiFS tests, as well as for the breakdown of TRU-SimL while registering ETM-MODIS NIR bands. Again, overall TRU-SimB seems to be the most resistant to this phenomenon.

Finally, for all reference-input pairs, the final transformations are different among the three algorithms, with the differences of less than a half of a pixel, and it remains an open problem to determine which of the three results is the closest to the actual ground truth.

3) EOS Land Validation Core Site Data, Multiple Terrain Types: The second real dataset, which we denote by CORE2, represents images acquired by IKONOS and Landsat-7/ETM+ sensors over four different Core Sites, which contain different types of terrain:

- The Cascade Mountains, data acquired in September of 2000.
- An agricultural area in the Konza Prairie, data acquired between July and August of 2001.
- An urban area around the United States Department of Agriculture (USDA) center in Greenbelt, Maryland, data acquired in May of 2001.

Init Guess $p_0 = [t_{x_0}, t_{y_0}]$	Final Transform $p = [t_x, t_y, \theta, \kappa]$ (Approx GT $p_{GT} = [2, 1, 0, 0.125]$)		
	TRU-SplC	TRU-SimB	TRU-SimL
[0, 0]	[2.304, 0.788, −0.069, 0.125]	[2.275, 0.796, −0.069, 0.125]	[2.244, 0.798, −0.062, 0.125]
[0, 1]	[2.304, 0.788, −0.069, 0.125]	[2.275, 0.796, −0.069, 0.125]	[2.244, 0.798, −0.062, 0.125]
[1, 0]	[2.304, 0.788, −0.069, 0.125]	[2.275, 0.796, −0.069, 0.125]	[2.244, 0.798, −0.062, 0.125]
[1, 1]	[2.304, 0.788, −0.069, 0.125]	[2.275, 0.796, −0.069, 0.125]	[2.244, 0.798, −0.062, 0.125]
[2, 0]	[2.304, 0.788, −0.069, 0.125]	[2.275, 0.796, −0.069, 0.125]	[2.243, 0.798, −0.062, 0.125]
[2, 1]	[2.304, 0.788, −0.069, 0.125]	[2.275, 0.796, −0.069, 0.125]	[2.244, 0.798, −0.062, 0.125]

TABLE V: CORE1: Final Transforms of IKONOS-ETM RED-RED Tests ($[\theta_0, \kappa_0] = [0, 1/8]$)

table

Init Guess $p_0 = [t_{x_0}, t_{y_0}]$	Final Transform $p = [t_x, t_y, \theta, \kappa]$ (Approx GT $p_{GT} = [-2, -4, 0, 0.0625]$)		
	TRU-SplC	TRU-SimB	TRU-SimL
[0, 0]	[−0.948, −1.181, −0.02, 0.063]	[−1.604, −3.448, 0.08, 0.062]	[−12.1, 3.166, 0.06, 0.070]
[0, −1]	[−1.163, −2.087, 0.05, 0.063]	[−1.605, −3.448, 0.08, 0.062]	[−12.0, 3.121, 0.05, 0.070]
[0, −2]	[−1.675, −1.716, 0.06, 0.064]	[−1.604, −3.448, 0.08, 0.062]	[−11.9, 3.082, 0.04, 0.070]
[0, −3]	[−1.316, −2.780, 0.10, 0.063]	[−1.604, −3.448, 0.08, 0.062]	[−11.9, 3.047, 0.03, 0.070]
[0, −4]	[−1.345, −2.934, 0.11, 0.063]	[−1.604, −3.448, 0.08, 0.062]	[−1.635, −3.182, 0.09, 0.063]
[−1, 0]	[−1.788, −1.400, 0.03, 0.064]	[−1.605, −3.448, 0.08, 0.062]	[−12.2, 3.194, 0.07, 0.070]
[−1, −1]	[−1.420, −2.133, 0.02, 0.063]	[−1.605, −3.448, 0.08, 0.062]	[−12.119, 3.146, 0.06, 0.070]
[−1, −2]	[−1.859, −1.616, 0.05, 0.064]	[−1.605, −3.448, 0.08, 0.062]	[−12.031, 3.104, 0.05, 0.070]
[−1, −3]	[−1.827, −2.828, 0.10, 0.063]	[−1.605, −3.448, 0.08, 0.062]	[−11.958, 3.067, 0.04, 0.070]
[−1, −4]	[−1.827, −1.655, 0.06, 0.064]	[−1.605, −3.450, 0.08, 0.062]	[−11.897, 3.033, 0.03, 0.070]
[−2, 0]	[−2.310, −0.316, −0.07, 0.063]	[−1.607, −3.451, 0.08, 0.062]	[−1.962, −2.645, 0.08, 0.063]
[−2, −1]	[−2.080, −2.124, −0.00, 0.063]	[−1.607, −3.452, 0.08, 0.062]	[−1.881, −2.793, 0.06, 0.063]
[−2, −2]	[−1.899, −2.510, 0.02, 0.063]	[−1.607, −3.452, 0.08, 0.062]	[−1.756, −3.043, 0.07, 0.063]
[−2, −3]	[−1.739, −2.948, 0.05, 0.063]	[−1.606, −3.449, 0.08, 0.062]	[−1.711, −3.149, 0.07, 0.063]
[−2, −4]	[−1.719, −3.055, 0.07, 0.063]	[−1.607, −3.452, 0.08, 0.062]	[−1.756, −3.378, 0.04, 0.063]

TABLE VI: CORE1: Final Transforms of ETM-MODIS RED-RED Tests ($[\theta_0, \kappa_0] = [0, 1/16]$)

table

Platform	Band	# inits	Approximate $E(p_\epsilon)$ [Mean Median Stdev]		
			TRU-SplC	TRU-SimB	TRU-SimL
IKONOS-ETM	RED-RED	6	[0.409, 0.409, 0.000]	[0.384, 0.384, 0.000]	[0.356, 0.356, 0.000]
	NIR-NIR	6	[0.599, 0.601, 0.048]	[0.399, 0.399, 0.000]	[0.379, 0.379, 0.000]
ETM-MODIS	RED-RED	15	[2.714, 2.431, 1.212]	[0.790, 0.790, 0.002]	[14.388, 23.052, 11.198]
	NIR-NIR	15	[6.579, 3.737, 7.464]	[0.824, 0.824, 0.000]	[294.887, 294.819, 1.250]
MODIS-SeaWiFS	RED-RED	9	[3.179, 2.429, 1.222]	[3.402, 3.374, 0.040]	[3.119, 3.214, 0.550]
	NIR-NIR	9	[3.673, 3.357, 2.393]	[2.946, 2.413, 1.560]	[2.982, 2.711, 1.969]

TABLE VII: CORE1: Means, Medians and Standard Deviations of Approximate RMS Errors

table

	Ref = IKONOS RED/NIR	Inp = ETM+ RED/NIR		
	Cascades (Mountains)	Konza (Agricultural)	USDA (Urban)	VA Coast (Coastal)
$p_{GT} \approx [t_x, t_y, \theta, \kappa]$	$[8.7, 10.2, 0, 0.1333]$	$[13.2, 12.3, 0, 0.1333]$	$[10.0, 14.2, 0, 0.1333]$	$[13.0, 13.2, 0, 0.1333]$

TABLE VIII: CORE2: Approximate Ground Truth

table

- The Virginia Coastal Reserve, data acquired in October of 2000.

We again used Red and NIR bands from both sensors. The images were not resampled, preserving the original resolutions of 4 and 30 meters, respectively, for IKONOS and ETM+. The NIR band images are shown in Figure 9. Note a considerable amount of clouds in the USDA ETM+ image, which had to be masked out. Approximate ground truth was obtained manually and is summarized in Table VIII. We performed the same type of test as in Section V-B.2 and studied sensitivity of the three algorithms to the initial guess. The results are summarized in Table IX where we again present average approximate RMS errors and their standard deviations. We observe that the Cascades scene was processed equally well by all three algorithms. In the case of Konza and USDA data, TRU-SimB and TRU-SimL outperformed TRU-SplC, although TRU-SimL failed on the USDA NIR images, possibly due to the significant cloud masking. Interestingly, in this case both median and mean were quite large (≈ 295) while the deviation was small (1.250). The cause of this phenomenon was that for all initial guesses, TRU-SimL consistently converged to a transformation $p = [t_x, t_y, \theta, \kappa] \approx [-75, 36, -1, 0.166]$. This consistency suggests that the computed p, while clearly being very far from the correct solution, is nevertheless a local minimum point with a large region of attraction. The VA Coast scene was also processed well by the three algorithms with TRU-SimB being somewhat inferior to the other two. Note that while the mean and deviation were large for TRU-SimB, the median was small. The reason for this was that TRU-SimB failed in a few cases while giving good results for other initial guesses (for details see Table X).

VI. DISCUSSION AND CONCLUSIONS

In this paper we compared performance of several wavelet pyramids in the framework of multiresolution sub-pixel image registration. These pyramids were combined with both correlation-based exhaustive search and L_2-based gradient least-squares optimization. Our first observation was that invariant wavelets pyramids were clearly superior to orthogonal wavelets. We also found that in terms of sensitivity to noise, distance of the initial guess from the ground truth, as well as to small variations in radiometry, a gradient based algorithm combined with the Simoncelli steerable band-pass pyramid (TRU-SimB) outperformed the same
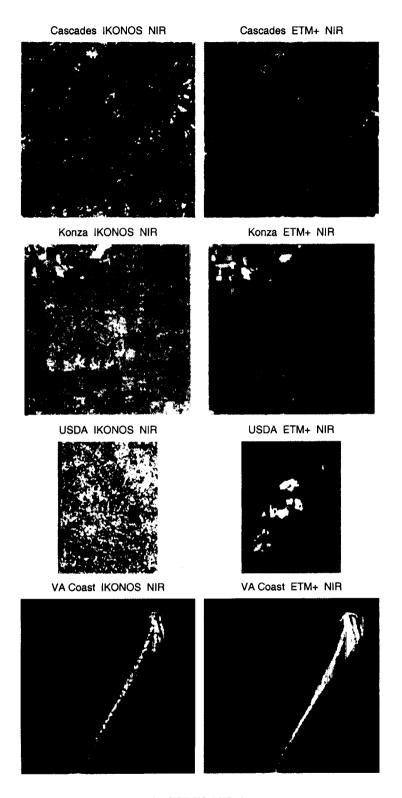
Cascades IKONOS NIR   Cascades ETM+ NIR

Konza IKONOS NIR   Konza ETM+ NIR

USDA IKONOS NIR   USDA ETM+ NIR

VA Coast IKONOS NIR   VA Coast ETM+ NIR

Fig. 9: CORE2 NIR Images

figure

Scene	Band	# inits	Approximate $E(p_\epsilon)$ [Mean Median Stdev]		
			TRU-SplC	TRU-SimB	TRU-SimL
Cascades	RED-RED	11	[0.294, 0.318, 0.070]	[0.361, 0.361, 0.000]	[0.307, 0.309, 0.003]
	NIR-NIR	11	[0.267, 0.267, 0.000]	[0.265, 0.265, 0.000]	[0.278, 0.278, 0.000]
Konza	RED-RED	14	[4.201, 0.409, 7.537]	[0.481, 0.481, 0.000]	[0.422, 0.420, 0.012]
	NIR-NIR	14	[7.944, 8.505, 6.743]	[0.387, 0.387, 0.000]	[0.536, 0.528, 0.020]
USDA	RED-RED	15	[6.946, 7.643, 7.209]	[0.091, 0.091, 0.000]	[0.195, 0.195, 0.000]
	NIR-NIR	15	[18.223, 18.221, 0.018]	[0.156, 0.156, 0.001]	[15.803, 16.306, 3.471]
VA Coast	RED-RED	14	[0.091, 0.091, 0.000]	[21.301, 0.149, 60.243]	[0.220, 0.220, 0.000]
	NIR-NIR	14	[0.224, 0.224, 0.001]	[6.007, 0.154, 21.899]	[0.116, 0.116, 0.000]

TABLE IX: CORE2: Means, Medians and Standard Deviations of Approximate RMS Errors

table

Init Guess	Final Transform $p = [t_x, t_y, \theta, \kappa]$ (Approx GT $p_{GT} = [13.0, 13.2, 0.0, 0.1333]$)		
$p_0 = [t_{x_0}, t_{y_0}]$	TRU-SplC	TRU-SimB	TRU-SimL
[0, 0]	[12.975, 13.220, 0.020, 0.1333]	[−128.912, 96.890, 3.727, 0.2714]	[12.990, 13.356, 0.057, 0.1334]
[1, 1]	[12.975, 13.220, 0.020, 0.1333]	[−12.565, 12.849, −6.579, 0.1443]	[12.990, 13.356, 0.057, 0.1334]
[2, 2]	[12.975, 13.220, 0.020, 0.1333]	[−17.614, 17.966, −6.289, 0.1494]	[12.990, 13.356, 0.057, 0.1334]
[3, 3]	[12.975, 13.220, 0.020, 0.1333]	[12.890, 13.260, 0.021, 0.1333]	[12.990, 13.356, 0.057, 0.1334]
[4, 4]	[12.975, 13.220, 0.020, 0.1333]	[12.890, 13.260, 0.021, 0.1333]	[12.990, 13.356, 0.057, 0.1334]
[5, 5]	[12.975, 13.220, 0.020, 0.1333]	[12.891, 13.260, 0.021, 0.1333]	[12.990, 13.356, 0.057, 0.1334]
[6, 6]	[12.975, 13.220, 0.020, 0.1333]	[12.890, 13.260, 0.021, 0.1333]	[12.990, 13.356, 0.057, 0.1334]
[7, 7]	[12.976, 13.219, 0.020, 0.1333]	[12.890, 13.260, 0.021, 0.1333]	[12.990, 13.357, 0.057, 0.1334]
[8, 8]	[12.976, 13.219, 0.020, 0.1333]	[12.890, 13.260, 0.021, 0.1333]	[12.990, 13.356, 0.057, 0.1334]
[9, 9]	[12.975, 13.220, 0.020, 0.1333]	[12.890, 13.261, 0.021, 0.1333]	[12.990, 13.356, 0.057, 0.1334]
[10, 10]	[12.975, 13.219, 0.020, 0.1333]	[12.891, 13.260, 0.021, 0.1333]	[12.990, 13.356, 0.057, 0.1334]
[11, 11]	[12.975, 13.219, 0.020, 0.1333]	[12.890, 13.261, 0.021, 0.1333]	[12.990, 13.356, 0.057, 0.1334]
[12, 12]	[12.975, 13.219, 0.020, 0.1333]	[12.890, 13.260, 0.021, 0.1333]	[12.990, 13.356, 0.057, 0.1334]
[13, 13]	[12.975, 13.219, 0.020, 0.1333]	[12.890, 13.260, 0.021, 0.1333]	[12.990, 13.356, 0.057, 0.1334]

TABLE X: CORE2: Final Transforms of VA Coast RED-RED Tests $([\theta_0, \kappa_0] = [0, 0.1333])$

table

algorithm combined with low-pass pyramids, either derived from splines (TRU-SplC) or from the Simoncelli's framework (TRU-SimL). Refer to Table XI for a summary of our findings, which are presented in greater detail in Section V-B.

While L_2-based approach to registration is well-suited for the types of data presented in this study, even if combined with the SimB wavelets, it may not work well on multi-sensor data that has large radiometric variations that cannot be easily modelled. Therefore at present time we are evaluating an alternative approach that is based on the concept of Mutual Information [44].

Test Imagery	Largest Radius of Convergence	Best Accuracy When Converged	Best Consistency
Synthetic, Same Radiometry, With Noise	About same	About same	TRU-SimB
Synthetic, Different Radiometry, Noiseless	TRU-SplC/TRU-SimL	TRU-SimB	About same
Real, Multi-Sensor/Multi-Terrain	TRU-SimB	Not Applicable	TRU-SimB

TABLE XI: Summary of TRU Test Results

table

ACKNOWLEDGEMENTS

REFERENCES

[1] L. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.

[2] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[3] J. L. Moigne, W. Xia, J. Tilton, B.-T. Lerner, E. Kaymaz, J. Pierce, S. Raghavan, S. Chettri, T. El-Ghazawi, M. Manohar, N. Netanyahu, W. Campbell, and R. Cromp, "Towards an intercomparison of automated registration algorithms for multiple source remote sensing data," Image Registration Workshop, NASA Goddard Space Flight Center, 1997, NASA Publication #CP-1998-206853.

[4] J. L. Moigne, W. Xia, J. Tilton, T. El-Ghazawi, M. Mareboyana, N. Netanyahu, W. Campbell, and R. Cromp, "First evaluation of automatic image registation methods," in *International Geoscience and Remote Sensing Symposium*, 1998.

[5] P. Thévenaz, U. E. Ruttiman, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 27–41, 1998.

[6] J. Ton and A. Jain, "Registering Landsat images by point matching," *IEEE Transactions in Geoscience and Remote Sensing*, vol. 27, no. 5, 1989.

[7] J. Djamdji, A. Bijaoui, and R. Maniere, "Geometrical registration of images: The multiresolution approach," *Photogrammetric Engineering and Remote Sensing Journal*, vol. 59, no. 5, May 1993.

[8] J. Flusser, "An adaptive method for image registration," *Pattern Recognition*, vol. 25, no. 1, pp. 45–54, 1992.

[9] A. Goshtasby, "Registration of images with geometric distortions," *IEEE Transactions in Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 60–64, 1988.

[10] B. Manjunath, C. Shekhar, and R. Chellappa, "A new approach to image feature detection with applications," *Pattern Recognition*, vol. 29, no. 4, pp. 627–640, 1996.

[11] Q. Zheng and R. Chellappa, "A computational vision approach to image registration," *IEEE Transactions on Image Processing*, vol. 2, pp. 311–326, 1993.

[12] D. Mount, N. Netanyahu, and J. LeMoigne, "Efficient algorithms for robust feature matching," *Special Issue of Pattern Recognition on Image Registration*, vol. 32, no. 1, pp. 17–38, January 1999.

[13] Q. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 12, pp. 1156–1168, 1994.

[14] M. Unser and A. Aldroubi, "A multiresolution image registration procedure using spline pyramids," in *SPIE, Mathematical Imaging: Wavelet Applications in Signal and Image Processing*, San Diego, 1993.

[15] G. Stockman, S. Kopstein, and S. Bennett, "Matching images to models for registration and object detection via clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-4, no. 3, pp. 229–241, 1982.

[16] R. Allen, F. Kamangar, and E. Stokely, "Laplacian and orthogonal wavelet pyramid decompositions in coarse-to-fine registration," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, December 1993.

[17] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," DARPA Image Understanding Workshop, 1981.

[18] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231–239, 1991.

[19] R. Eastman and J. LeMoigne, "Gradient-descent techniques for multi-temporal and multi-sensor image registration of remotely sensed imagery," in *FUSION 2001*, 2001.

[20] J. LeMoigne, "Parallel registration of multi-sensor remotely sensed imagery using wavelet coefficients," in *SPIE OE/Aerospace Sensing, Wavelet Applications Conference*, Orlando, Florida, 1994.

[21] J. LeMoigne, W. Campbell, and R. Cromp, "An automated parallel image registration technique based on the correlation of wavelet features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 8, pp. 1849–1864, August 2002.

[22] Y. Wu and H. Maitre, "A multiresolution approach for registration of a spot image and a sar image," in *10th Annual International Geoscience and Remote Sensing Symposium*, May 1990, pp. 635–638.

[23] H. Stone, J. LeMoigne, and M. McGuire, "The translation sensitivity of wavelet-based registration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 21, no. 10, pp. 1074–1081, 1999.

[24] G. Wolberg, *Digital Image Warping*. IEEE Computer Society Press, 1990.

[25] J. LeMoigne, A. Cole-Rhodes, R. Eastman, K. Johnson, J. Morisette, N. Netanyahu, H. Stone, and I. Zavorin, "Multi-sensor image registration for on-the-ground or on-board science data procesing," Science Data Processing Workshop, SDP'2002, Greenbelt, MD, February 2002.

[26] R. G. Lyon, J. E. Dorband, and J. M. Hollis, "Hubble space telescope faint object camera calculated point spread functions," *Applied Optics*, vol. 36, no. 8, 1997.

[27] H. S. Stone, M. T. Orchard, E.-C. Cheng, and S. A. Martucci, "A fast direct Fourier-based algorithm for subpixel registration of images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 10, pp. 2235–2243, 2001.

[28] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase-correlation to sub-pixel registration," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 188–200, 2002.

[29] I. Daubechies, *Ten lectures on wavelets*. Philadelphia: SIAM, 1992.

[30] E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger, "Shiftable multiscale transforms," *IEEE Transactions on information theory*, vol. 38, no. 2, pp. 587–607, 1992.

[31] Z. Liu, Y. K. Ho, K. Tsukada, K. Hanasaki, Y. Dai, and L. Li, "Using multiple orientation filters of steerable pyramid for image registration," information fusion, to appear.

[32] I. Zavorin, H. Stone, and J. LeMoigne, "Iterative pyramid-based approach to subpixel registration of multisensor satellite imagery," in *SPIE International Symposium on Optical Science and Technology 2002, Earth Observing Systems VII*, Seattle, WA, July 2002.

[33] M. Unser, A. Aldroubi, and M. Eden, "The L2-polynomial spline pyramid," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 364–379, 1993.

[34] I. J. Schoenberg, "Cardinal interpolation and spline functions," *Journal of Approximation Theory*, vol. 2, pp. 167–206, 1969.

[35] R. Coifman and D. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics, Lecture Notes in Statistics 103*, A. Antoniadis and G. Oppenheim, Eds. Springer-Verlag, 9999, pp. 125–150.

[36] J. Magarey and N. Kingsbury, "Motion estimation using a complex-valued wavelet transform," *IEEE Transactions on signal processing*, vol. 46, no. 4, pp. 1069–1084, 1997.

[37] W. Pratt, "Correlation techniques of image registration," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES10, no. 3, pp. 353–358, 1974.

[38] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of SIAM*, vol. 11, pp. 431–441, 1963.

[39] P. Thévenaz, "Intramodal registration," http://bigwww.epfl.ch/thevenaz/registration/index.html.

[40] J. Morisette, J. Privette, and C. Justice, "A framework for the validation of MODIS land products," *Remote Sensing of Environment*, vol. 83, no. 1-2, pp. 77–96, 2002.

[41] I. Zavorin, H. Stone, and J. LeMoigne, "Evaluating performance of automatic techniques for subpixel registration of remotely sensed imagery," in *SPIE 15th Annual Symposium on Electronic Imaging Science and Technology 2002, Image Processing Algorithms and Systems II*, Santa Clara, CA, January 2003.

[42] J. LeMoigne, A. Cole-Rhodes, R. Eastman, K. Johnson, J. Morisette, N. Netanyahu, H. Stone, and I. Zavorin, "Multi-sensor image registration of remotely sensed imagery," in *8th International Symposium on Remote Sensing*, Toulouse, France, 2001, vol. 4541.

[43] P. Thévenaz, November 2002, private communication.

[44] P. Thévenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Transactions on Image Processing*, vol. 9, no. 12, pp. 2083–2099, December 2000.