Multiresolution Registration of Remote Sensing Imagery by Optimization of Mutual Information Using a Stochastic Gradient

Arlene A. Cole-Rhodes, *Member, IEEE*, Kisha L. Johnson, Jacqueline LeMoigne, *Senior Member, IEEE*, and Ilya Zavorin

***Abstract*—Image registration is the process by which we determine a transformation that provides the most accurate match between two images. The search for the matching transformation can be automated with the use of a suitable metric, but it can be very time-consuming and tedious. In this paper, we introduce a registration algorithm that combines a simple yet powerful search strategy based on a stochastic gradient with two similarity measures, correlation and mutual information, together with a wavelet-based multiresolution pyramid. We limit our study to pairs of images, which are misaligned by rotation and/or translation, and present two main results. First, we demonstrate that in our application mutual information may be better suited for sub-pixel registration as it produces consistently sharper optimum peaks than correlation. Then, we show that the stochastic gradient search combined with either measure produces accurate results when applied to synthetic, as well as multitemporal or multisensor collections of satellite data. Mutual information is generally found to optimize with one-third the number of iterations required by correlation. Results also show that a multiresolution implementation of the algorithm yields significant improvements in terms of both speed and robustness over a single-resolution implementation.**

***Index Terms*—Image registration, mutual information, remote sensing imagery, stochastic optimization, wavelets.**

I. Introduction

DIGITAL image registration is a process by which the most accurate match is determined between two images, which may have been taken at the same or different times, by the same or different sensors, from the same or different viewpoints. The registration process determines the optimal transformation, which will align the two images. This has applications in many fields as diverse as medical image analysis, pattern matching and computer vision for robotics, as well as remotely sensed data processing. In all of these domains, image registration can be used to find changes in images taken at different times, or to build 3-D models from 2-D images taken from different viewpoints, or for object recognition.

In the remote sensing framework in particular, with the increasing number of multiple platform remote sensing missions, different sensors may simultaneously observe the same features. These sensors may produce data at different resolutions or in different spectral ranges, over multiple times, thus providing very large amounts of redundant or complementary data. The combination of all these data will allow for better analysis of various phenomena, as well as allow the validation of global low-resolution analysis by the use of local high-resolution data analysis. For all these applications, accurate geo-referencing is the first step in integrating such data from multiple sources, and it is thus becoming a very important issue in remote sensing. By using a model-based systematic correction, newly acquired remote sensing data is usually geo-referenced to within a few pixels. Starting with this information, we focus on precision correction or automatic image registration, which refines the accuracy to within one pixel or a sub-pixel. For applications such as data fusion, it is very important to reach sub-pixel accuracy, and automatic image registration offers a practical means of achieving this.

In this context, we define image registration as follows: Given a pair of two-dimensional gray-level images, $F_R(x,y)$ and $F_I(x,y)$ that we denote by the reference and input (or sensed) images respectively with coordinates $(x,y) \in \Delta \subset \mathcal{R}^2$, where Δ is a region of interest; To register the images is to find a geometric transformation $T_P(.)$ of a certain class such that for all (x,y), $F_R(T_P(x,y))$ best matches $F_I(x,y)$, where P is a set of transform parameters. In this paper, we limit $T_P(.)$ to a class of transforms that include shift (tx,ty) and rotation (θ) and can be written as

$$T_P(x,y) = \begin{bmatrix} \cos(\theta) & \sin(\theta) & tx \\ -\sin(\theta) & \cos(\theta) & ty \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \qquad (1)$$

Thus we can write $T_P(x,y) = Q_P\ [x\ y\ 1]^T$, where we define Q_P to be the transformation matrix given above, for $P = \{tx\ ty\ \theta\}^T$. Later we can incorporate isometric scaling into our study.

In order to find the optimum transformation, the image registration process may include the following steps: 1) the extraction of features to be used in the matching process, 2) the feature matching strategy and metrics, and 3) the resampling of

Manuscript received August 15, 2002; revised July 21, 2003. This work was supported by the National Science Foundation under Grant ECS-074929 and by NASA under NRA-NAS2–37143 on "Research in Intelligent Systems." The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Thierry Blu.

A. A. Cole-Rhodes and K. L. Johnson are with the Electrical and Computer Engineering Department, Morgan State University, Baltimore, MD 21251 USA (e-mail: acrhodes@eng.morgan.edu; kjohnson@eng.morgan.edu).

J. LeMoigne is with Applied Information Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771 (e-mail: lemoigne@backserv.gsfc.nasa.gov).

I. Zavorin is with UMBC/GEST Center, NASA Goddard Space Flight Center, Greenbelt, MD 20771 USA (e-mail: zavorin@backserv.gsfc.nasa.gov).

Digital Object Identifier 10.1109/TIP.2003.819237

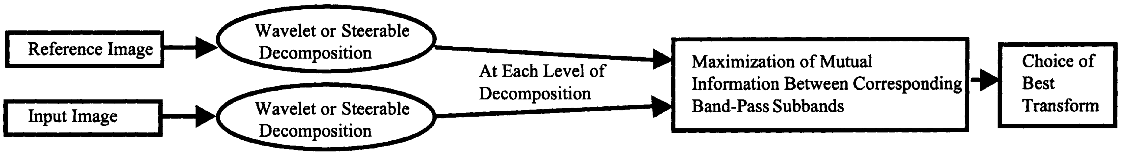

Reference Image
Input Image
Wavelet or Steerable Decomposition
Wavelet or Steerable Decomposition
At Each Level of Decomposition
Maximization of Mutual Information Between Corresponding Band-Pass Subbands
Choice of Best Transform

Fig. 1. Summary of our wavelet-based mutual information registration method.

the data based on the correspondence computed from matched features. Many automatic image registration methods have been proposed and a survey can be found in Brown [1]. Our work considers the search strategy and similarity metric to be used in step 2) of the registration process. Many objective functions exist in the literature, which can be used in automated image registration schemes. These objective functions may be feature-based or intensity-based. Feature-based methods establish geometric correspondences by matching salient features, which have been extracted by pre-processing the images. A drawback of these algorithms lies in the difficulty of recognizing matched features in the images, and they require the use of reliable and robust algorithms for image segmentation and edge detection. By contrast intensity-based methods require no prior pre-processing of the images. Commonly used intensity-based objective functions include intensity correlation, the mean square difference of the image intensity values, and mutual information (MI). Cross-correlation is one of the most common similarity metrics used in registration. It measures similarity by computing global statistics such as mean and variance, and it performs well if the two images are similar in nature, with an underlying linear relationship between the image intensities. On the other hand, mutual information measures redundancy between two images by looking at their intensity distributions, and it represents a measure of the relative entropy between two sets. Mutual information (MI) has been extensively studied for the registration of medical imagery [3]–[5], and it has been found to be especially robust for multimodal image registration.

In this paper, we show how mutual information can be successfully merged with an optimization scheme and applied to the registration of remotely sensed imagery. Our first tests are designed to compare the sharpness of the MI and correlation curves, and they show that MI produces consistently sharper peaks at the correct registration values than correlation. Moreover, when used with a multiresolution search strategy, this comparative result is also verified for the lower resolution sub-band images of the Simoncelli pyramid described in Section II. The use of a multiresolution search provides for large reductions in computing time, and this result is very important for producing consistently accurate results within such a scheme.

In our earlier work [2], [6], [14] a simple search strategy, based on exhaustive search, was used to provide a thorough comparison of the two different metrics. But exhaustive search is computationally expensive, and the computational cost increases exponentially with the number of transformation parameters and the size of the dataset. Therefore, in this work we describe a more sophisticated search technique, which uses a gradient approximation, that is applied within a multiresolution framework based on a wavelet-like pyramid decomposition. Section II describes our registration framework, while Sections III and IV present cross-correlation and MI, together with a comparative study of the performance of these two metrics when applied to image registration. Section V then describes our optimization search technique and associated results are presented in Section VI. Section VII discusses other related work, in particular comparing the algorithm presented here to that of Thevenaz *et al.* [5], and it gives conclusions and directions of future work. The main innovation of this paper is in the use of the simultaneous perturbation stochastic approximation (SPSA) gradient strategy for the optimization of the mutual information similarity criterion. It provides a simple, more practical approach to MI-based registration problems than what is currently found in the literature.

II. Multiresolution Image Registration

Most of our previous work in image registration has focused on the use of wavelets or wavelet-like features in step 1) of the registration process. Fig. 1 summarizes our registration scheme [2], [6], [14] when wavelet or wavelet-like information is utilized. Both the reference and input images are first decomposed following a multiresolution wavelet or frame decomposition. In order to achieve computational efficiency, our search strategy follows the multiresolution decomposition, working iteratively from the deepest level of decomposition (where the image size is the smallest) to the top level of decomposition, i.e., going from coarse to fine spatial resolution. For all levels of decomposition, MI or correlation between sub-band images of the reference image and input image is successively computed and maximized. The accuracy of this search increases when going from coarse resolution to fine resolution. At each level the search focuses in on an interval around the “best” transformation found at the previous level and is refined at the next level up. As a preliminary study, our search space is restricted to 2-D rotations and translations, and this will be extended later to affine transformations. To obtain the transformed images, data interpolation is done using cubic B-splines [18]. Maximization of the metric can be performed by exhaustive search, but it is more efficient and more accurate if an automated optimization technique is used.

Different wavelet or wavelet-like filters could be chosen, but our previous work [7] showed that Steerable Simoncelli filters [8] are more robust to translation, rotation and noise than the standard Daubechies wavelet filters. The method described by Simoncelli [8] enables one to build translation- and rotation-invariant filters by relaxing the critical sampling condition of the wavelet transforms. By invariance, it is meant that the information contained in a given sub-band will be invariant to translation or rotation. The resulting representation is equivalent to an overcomplete wavelet transform; it is not an orthogonal repre-

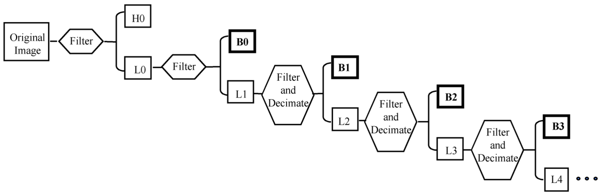


Fig. 2. Four-level decomposition by a steerable pyramid. Sub-bands *Bi* are utilized to extract features.

sentation but is an approximation of a "tight-frame," i.e., invertible. The Simoncelli Steerable Pyramid is summarized in Fig. 2, where only the analysis decomposition is shown. In Fig. 2, H0 is the result of high-pass filtering, $\{L0, \ldots, Lm\}$ are results of low-pass filtering, and $\{B0, \ldots, Bm\}$ represent the results after filtering by a set of oriented band-pass filters which ensure that the representation is rotation-invariant. In order to ensure some translation-invariance, the outputs of the high-pass filter and of the band-pass filters are not sub-sampled. In addition, that portion of the signal $L0$, which is iteratively decomposed by the band-pass and the low-pass filters, does not contain the larger high frequency components and has been preprocessed by a low-pass filter, thus removing most aliased components. This representation is overcomplete by a factor of $4k/3$, where k is the number of oriented band-pass filters [8]. In the study described in LeMoigne *et al.* [6], the steerable filters studied in a correlation framework, showed very accurate and reliable responses for registration purposes. Therefore, in the experiments shown here, we will use Simoncelli steerable filters, and in order to optimize the computational speed, we chose $k = 1$.

When using the multiresolution approach for registration, a wide variety of search methods can be utilized to obtain an approximation to the solution at each pyramid level. Different search strategies may even be used at different levels. The simplest approach is to apply an exhaustive search method at all pyramid levels, where one varies one or more of the transformation parameters over a certain discrete range of values, which is assumed to include the "true" transformation (or "Ground Truth," GT). For each combination of parameters, the similarity metric is computed and the combination that yields the largest metric value is chosen as the final approximation at the current level. How this discrete mesh is determined depends on the pyramid level. At the coarsest resolution, the initial range is usually specified by the user. When moving up the pyramid, the new range is chosen as a given interval centered around the solution computed at the previous step. Details of this approach can be found in [2].

Although this method is quite robust, it is not very practical for two reasons. First, it is computationally expensive even for a small number of search parameters. Second, it yields results of limited accuracy since the accuracy depends on how fine the discrete mesh is.

III. Correlation and Mutual Information as Similarity Metrics

A. Correlation

Correlation is one of the most widely used similarity metrics in image processing [16]. One of its principal applications is in the area of template, or prototype matching, where the problem is to find the closest match between an unknown image and a set of known images. One approach is to compute the correlation between the unknown and each of the known images. The closest match can then be found by selecting the image that yields the correlation with the largest value. Matching of images A and B can be performed by using the correlation coefficient, which is defined as

$$C(A,B)=\frac{\sum\limits_i\sum\limits_j[a_{ij}-mean(a)]*[b_{ij}-mean(b)]}{\left[\sum\limits_i\sum\limits_j(a_{ij}-mean(a))^2*\sum\limits_i\sum\limits_j(b_{ij}-mean(b))^2\right]^{1/2}} \tag{2}$$

where the double sums indicated are taken over the rows and columns of the two images, and a_{ij}, b_{ij} are the pixel values of images A and B at row i and column j, respectively. This statistical measure has the property that it measures correlation on an absolute scale ranging from [-1, 1]. Under the assumption that the transformation is small enough, it can be shown that maximizing this correlation measure is equivalent to minimizing the least-mean-square of the difference in the intensity values of A and B, see [17]. For many registration methods, correlation is the primary tool, where A may be an input image to be registered against a reference image, B. It is equal to one for identical images, and thus provides the degree of similarity between the two images.

The cost of a single computation of the spatial correlation of two images is $O(N^2)$, where N is the number of pixels in each image. When used for image registration, the total cost is then a function of the number of steps where the correlation is computed.

B. Mutual Information (MI)

The concept of mutual information represents a measure of relative entropy between two sets, which can also be described as a measure of information redundancy [3]–[5]. From this definition, it can easily be shown that the MI of two images is maximal when these two images are perfectly aligned. Therefore, in the context of image registration, MI can be utilized as a similarity measure which, through its maximum, will indicate the best match between a reference image and an input image. Experiments show that, in this context, MI enables one to extract an optimal match with a much better precision than cross-correlation.

If A and B are two images to register, $p_A(\mathrm{a})$ and $p_B(\mathrm{b})$ are defined as the marginal probability distributions, and $p_{AB}(a,b)$ is defined as the joint probability distribution of A and B. Then MI is defined as

$$I(A,B)=\sum_a\sum_b p_{A,B}(a,b)*\log\left(\frac{p_{A,B}(a,b)}{(p_A(a)*p_B(b))}\right). \tag{3}$$

This quantity can be computed using the histograms of the two images A and B, $h_A(\mathrm{a})$ and $h_B(\mathrm{b})$ respectively, as well as their joint histogram $h_{AB}(a,b)$. The MI is then defined by

$$I(A,B) = \frac{1}{M}\sum_{a}\sum_{b} h_{A,B}(a,b) * \log\left(\frac{M h_{A,B}(a,b)}{(h_A(a) * h_B(b))}\right) \tag{4}$$

where M is the sum of all the entries in the histogram, see [3]. The histograms are computed using original gray levels or gray levels of pre-processed images, such as edge gradient magnitudes or wavelet coefficients.

In this work a histogram with 64 bins is used, since it produces a significantly smoother MI surface than the 256-bin histogram. The smoother surface works better with the optimization algorithm, and the reduced number of bins dramatically improves the runtime for MI registration. The joint histogram is obtained by the following computation. The transformed reference image is obtained using cubic B-spline interpolation [18]. The gray values of the input image and the transformed reference image are linearly rescaled into the range [0,255]. The gray values (a,b) of those pairs of pixels, which lie in the same position are then used to build the histogram, using the following update law:

$$h_{AB}\left(\left[\frac{a}{4}\right],\left[\frac{b}{4}\right]\right) \rightarrow h_{AB}\left(\left[\frac{a}{4}\right],\left[\frac{b}{4}\right]\right) + 1 \tag{5}$$

where $(a,b) = (F_I(x,y)\ ,\ F_R(T(x,y)))$ for $0 \leq a,\, b \leq 255$. Note that $[z]$ represents the integer part of z, and a 64-bin histogram is produced.

The cost of computing the MI of two images depends both on the number of data points or pixels in each image, N, and also on the number of bins used to form the histogram. If both images have the same number of pixels, N, the computational cost of computing the histogram is $O(N)$. The computational cost relative to the number of histogram bins, K used in the computation, is $O(K^2)$.

IV. Evaluation of Mutual Information Versus Correlation for the Registration of Remote Sensing Imagery

In this section, we present results of a number of different tests, which provide a comparison between MI and correlation as two potential similarity measures for remote sensing image registration. In order to obtain high registration precision, it is important to use a similarity measure that produces a sharp peak at the correct transformation point with significantly smaller values elsewhere, especially in the vicinity of the correct transformation. Other important considerations for the choice of a similarity measure include the resolution and/or accuracy of the final solution, speed of computation, and the presence or absence of local extrema. These will be discussed in later Sections. In this Section, the following set of tests has been designed to compare sharpness of MI and correlation curves. The first set of tests illustrates that MI provides a sharper peak than correlation at the correct registration value of either a rotation, or a translation in one of the x- or y- directions, when searching exhaustively over a range of values. This sharper peak enables one to obtain a higher precision of the registration. These experiments are performed on a 512×512 image (Fig. 3) with no wavelet decomposition, and also on multiple resolutions of a Simoncelli decomposition. The second set of experiments investigates the sensitivity of the MI and correlation metrics to compositions of translations and rotations of the reference image when used in conjunction with the Simoncelli steerable filter decomposition. This sensitivity is then investigated for input images with varying levels of noise.

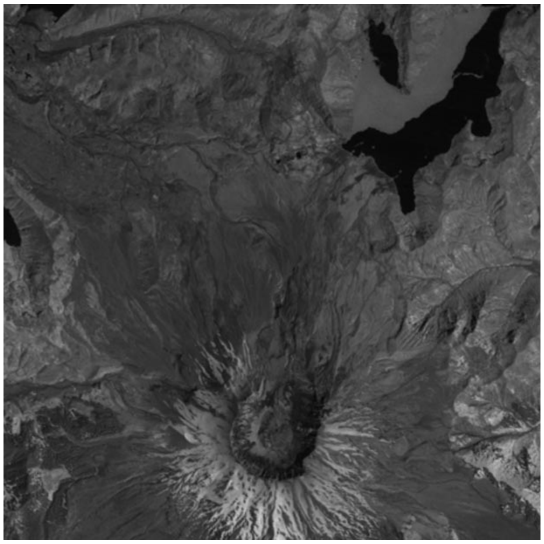


Fig. 3. Landsat – U.S. Pacific Northwest reference image for sharpness of MI and correlation curves.

The experiments described in this section include many of the issues that will be present in "real-life" imagery, although the list is not exhaustive. In particular, this set of experiments deals only with uncorrelated noise and single-modality inputs, but the results are still informative, and show the main characteristics of the two similarity metrics.

A. Sharpness of MI and Correlation Curves

After the curves for both metrics have been normalized to lie in the range [0,1] we restrict the neighborhood *V0*, for which the *area under the curve* is computed, to one centered around the maximal point and bounded by the points where the two curves intersect, when this does occur. Then for the correlation and mutual information curves which are produced, the following assumptions are noted to be true:

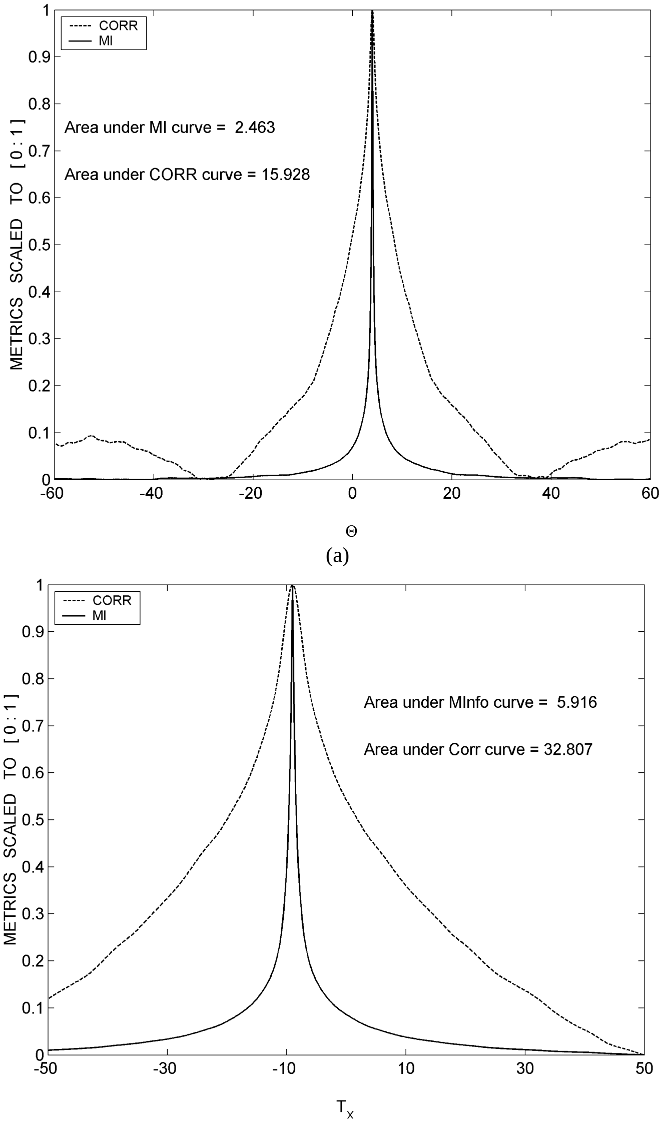
- the two functions are defined and continuous in *V0*;
- the two functions are both positive in *V0*;
- the two functions do not intersect in *V0* except at the maximum.

Under these assumptions, we say that a function f *is sharper than a function* g in a neighborhood *V0* if there exists a neighborhood *V1*, that is a subset of *V0* centered on the maximal point, such that the magnitude of the slope of f is larger than that of g for all points in *V1*. Since the two curves do not intersect in this neighborhood, and they are both normalized to the same maximal value of 1, it is then easy to show that this definition is equivalent to stating that the area under the curve f in *V0* is smaller than the area under the curve of g in *V0*. An alternative definition for the neighborhood, *V0* could be that of the region around the peak bounded by the closest inflexion points of the two curves. Inflexion points indicate the presence of other local maxima to which the optimization may possibly be attracted, and this neighborhood *V0*, would then define the *region of attraction* for each of these measures, indicating the maximum distance from which convergence to the optimum can be guaranteed. This issue is discussed further in Section VI-C-I. For the curves of this section we note that these two definitions yield neighborhoods, *V0*, which differ only slightly, and we use the first, more easily computable definition here.

1) Original Grey Level Imagery: First a 1024×1024 image is extracted from Band 4 of a Landsat-TM ("Thematic Mapper") scene of the Pacific Northwest, and a 512×512 reference image is produced from the center of this scene (Fig. 3). From the same scene, forty-two 512×512 input images are produced, with either a single translation or a single rotation of the reference image. The ground truth translations range from -10 to $+10$ pixels in the x-direction, and the rotations, θ, range from -10° to $+10^\circ$. We then register each of the 42 reference-input pairs by executing a one-dimensional exhaustive search where the reference is transformed either by a rotation ranging between -60° and 60°, or by a shift ranging between -50 and 50 pixels. Both correlation and MI are measured between the input and the transformed reference, and we compare the sharpness of the peak in the neighborhood, *V0* between each of the 42 correlation curves and the corresponding MI curves.

Examples of these curves are shown in Fig. 4. The scaled MI and correlation curves are shown in Fig. 4(a) for an input image which has a transformation of the reference given by $(tx, ty, \theta) = (0, 0, 4)$. Rotation, θ is varied over the range [-60, 60]. Fig. 4(b) shows the same curve for an input image with a transformation $(tx, ty, \theta) = (-9, 0, 0)$ as tx is varied in the range [-50, 50]. The solid curve represents MI and the dashed curve represents correlation. We showed [14] that MI produces a much sharper peak than correlation in both cases. More specifically, we find that for rotations, the average value of the area under the MI curve is 2.46 as compared to an average correlation value of 15.26, while for the translations the average MI value is 5.76, as compared to the correlation average of 32.02. These results quantitatively indicate how much sharper the MI curve is, compared to the correlation curve.

2) Simoncelli Band-Pass Imagery: In the second part of this experiment we use a single reference-input pair with both images produced from the same source as above. The reference is the 512×512 center of the source and the input is the 512×512 center of the source shifted by 32 pixels in the x-direction (horizontally). Thus the correct transformation between the reference and the input is $(tx, ty, \theta) = (32, 0, 0)$. The tested pair is then decomposed using single-orientation Simoncelli filters. Four levels of decomposition are produced, which correspond to scaling of the images by 1, 2, 4 and 8. At wavelet level J, we fix the parameters, $\theta = ty = 0$ and vary tx in the interval $[tx_J - 10, tx_J + 10]$, where tx_J is the correct transformation scaled to the level J resolution. Thus, for instance, at the 4th level, $tx_4 = 32/8 = 4$ and tx is varied between -6 and 14 with a step of 1, which corresponds to varying tx between -48 and 112 with a step of 8 at the original (finest) level. For MI and correlation, we generate a set of 4 curves of the measure value corresponding to the 4 levels of the Simoncelli decomposition. These curves are shown in Fig. 5.



(a)

(b)

Fig. 4. Scaled MI & correlation curves for registration of 512 images, with single-resolution. (a) Image transformation: $\theta = 4$, $\mathrm{Tx} = \mathrm{Ty} = 0$. (b) Image Transformation: $\mathrm{Tx} = -9$, $\theta = \mathrm{Ty} = 0$.

The solid curve represents MI and the dashed curve represents correlation. As expected, at all decomposition levels both correlation and MI produce their largest values at the points that correspond to the correct transformation. However, MI produces consistently sharper peaks than correlation. As in previous experiments, the area under the scaled correlation and MI curves indicated in Fig. 5, is used as a measure of sharpness of the curves, and again at all levels MI produces smaller areas. It is important to note that the correlation curves tend to be concave around the maximum, while the MI curves are often convex. This property, which explains the sharpness of the curves, could pose problems for the application of second order optimization methods.

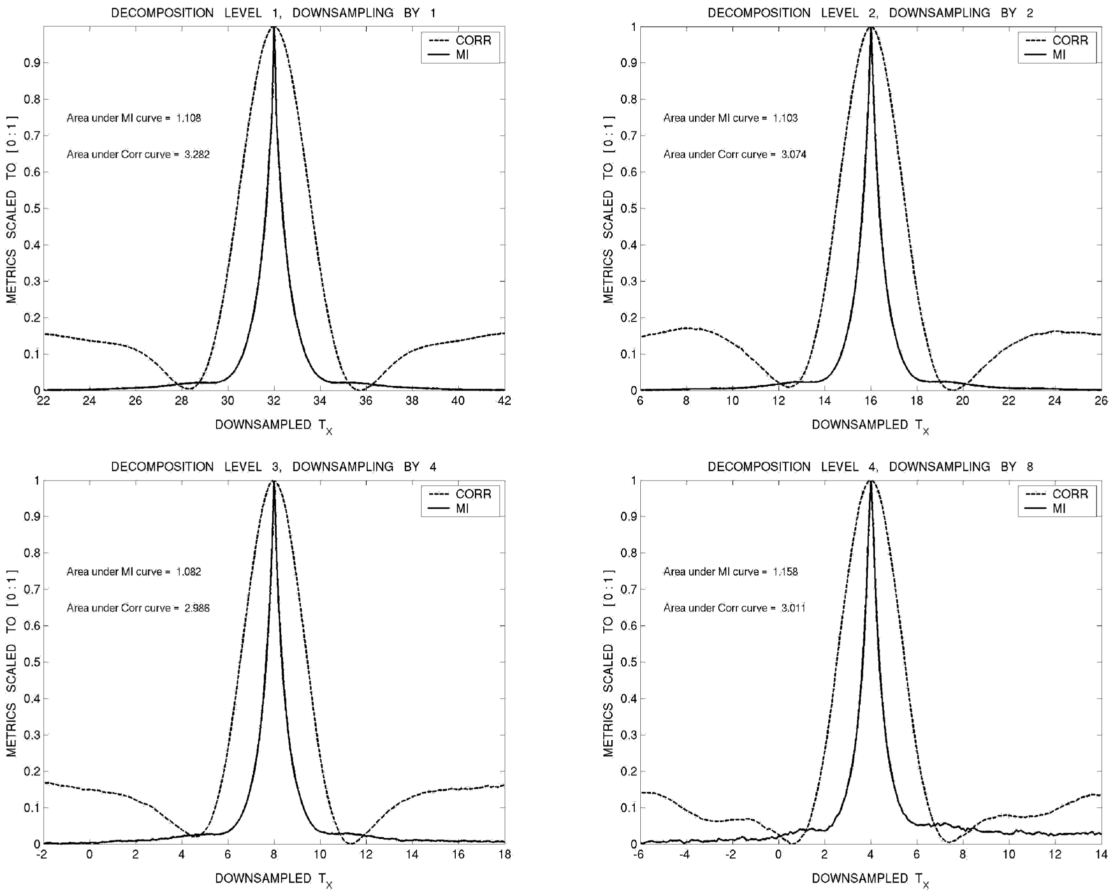


Fig. 5. Correlation and mutual information curves with values scaled to range [0,1] for the different levels of the Simoncelli decomposition.

B. Sensitivity to Noise

The following set of tests is designed to compare the sensitivity of the registration results to the amount of noise present in the data, when utilizing either correlation or MI. A collection of input images is generated using only one transformation, namely $(tx, ty, \theta) = (6, 4, 3)$, and then adding different amounts of Gaussian white noise. The added noise is measured by the signal-to-noise ratio, b expressed in decibels (dB), defined as

$$b = 10 \log_{10} \frac{\mathrm{Var(Image)}}{\mathrm{Var(Noise)}}. \tag{6}$$

In this experiment, the SNR is varied between 20 dB (almost noise-less) and -15 dB (extremely noisy). Two levels of single-orientation Simoncelli wavelet decomposition are computed for all images. Results are presented in Figs. 6 and 7, which show rotation and shift errors, respectively. We observe that both measures produce perfect results even with levels of noise as large as -12 dB. However, correlation-based results deteriorate faster than the MI-based results.

As a summary, we have shown that for these experiments, MI produces consistently sharper peaks at the correct registration values than correlation, which is important for obtaining sub-pixel registration accuracy. Moreover, sharper peaks are also produced at the lowest resolution of the sub-band images produced by a wavelet-like decomposition. This indicates that MI can produce more accurate results than correlation in a multiresolution registration scheme based on wavelet-like filters. Registration is achieved in a more efficient manner in this framework, since one can start with a smaller image for the initial search, and successfully narrow down the search range for the larger images. Our results show that even when noise is present in the input image, both correlation and MI produce perfect registration for Gaussian noise levels up to -12 dB for our tests with Simoncelli filters, and MI is more robust to noise than correlation.

V. Stochastic Gradient Optimization for Image Registration

In the previous sections, the search for the optimum transformation was done by an exhaustive search over an allowable range of parameters. But as previously stated, this computational cost increases exponentially with both the dimension of the parameter space and the dimension of the dataset. Exhaustive search becomes even more expensive when the goal is sub-pixel accuracy, thus an alternate iterative search method is considered in this Section.

A. Brief Survey of Optimization Techniques

The choice of optimization search technique depends on the type of problem under consideration. Traditional nonlinear programming methods, such as the constrained conjugate gradient, or the standard backpropagation in neural network applications, are well suited to deterministic optimization problems with exact knowledge of the gradient of the objective

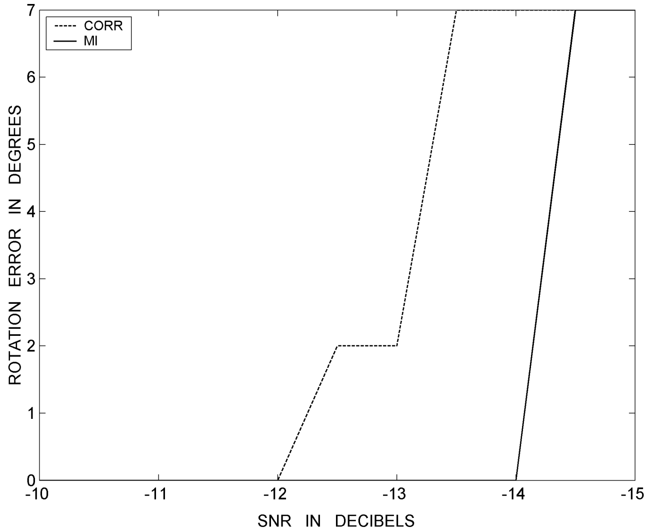


Fig. 6. Rotation error as a function of noise.

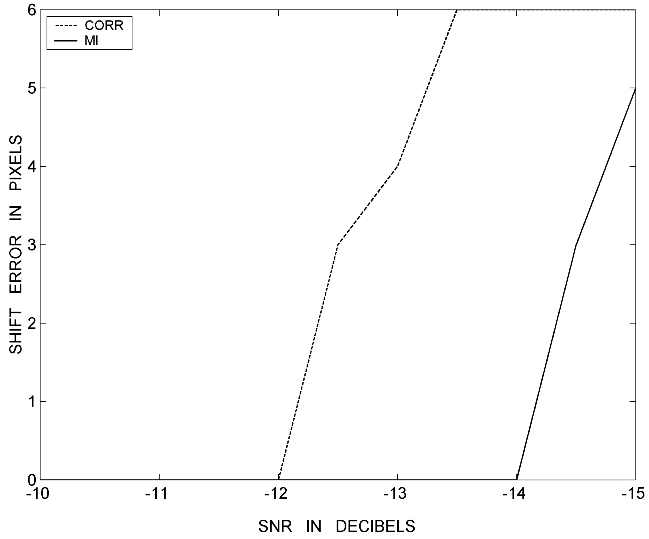


Fig. 7. Translation error as a function of noise.

function. Optimization algorithms have been developed for a stochastic setting where randomness is introduced either in the noisy measurements of the objective function and its gradient, or in the computation of the gradient approximation. These optimization algorithms can be divided into two categories: *Gradient-based algorithms*, such as the Robbins-Monro stochastic approximation algorithm can be considered to be a generalization of the deterministic steepest descent. It requires that direct measurements of the gradient are available, but these measurements are generally a gradient estimate because the underlying data is usually noisy. *Gradient-free algorithms* include some general-purpose optimizers such as the simple random search, or the genetic algorithm, which works with a population of candidate solutions and randomly alters the solution over a sequence of generations. Both these methods can be useful for a broad search over the domain of the parameters being optimized, and can provide initialization for a more powerful local search algorithm. Other nongradient optimization methods include Simulated annealing, the Nelder-Mead Simplex method which attempts to minimize a scalar-valued nonlinear function using only function evaluations, and the Kiefer-Wolfowitz algorithm which is a finite-difference method for optimization of noisy data. Approaches based on the use of gradient estimations tend to be fast, but are sensitive to the presence of local optima. Additional discussion of these methods can be found in [23].

The stochastic gradient technique, which is used in this work is a gradient-free approach. It does not require an explicit derivation of the required gradient vector, but it uses instead an approximation to the gradient. In the next Sections we show how it can be applied to image registration, and integrated within the multiresolution framework of the Simoncelli steerable pyramid described in Section II.

B. *Spall's Optimization Technique*

The optimization technique, which is implemented in this work is the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm. It was first introduced by Spall in [12], where a detailed description can be found. It has recently attracted attention for solving challenging optimization problems where it is difficult or impossible to obtain an analytic expression for the gradient of the objective function. This is especially true of the MI function, since the probabilities required in the computation of (3) are estimated using the joint image histogram. The dependence of the MI function on this discrete histogram makes the computation of its derivative complex. SPSA is based on an easily implemented and highly efficient gradient approximation that relies only on measurements of the objective function to be optimized. It does not rely on explicit knowledge of the gradient of the objective function, or on measurements of this gradient.

Let us call L, the objective function to be optimized. In our experiments, $L(.)$ represents either MI or the correlation similarity measure. We consider a parameter search space of two-dimensional rigid transformations, consisting of rotation and translation in the x and y-directions. There are thus three parameters to be optimized, represented in a vector form as $p = [tx, ty, \theta]^T$. At each iteration, the gradient approximation is based on only two function measurements (regardless of the dimension of the parameter space). An additional function measurement is made at each newly computed point, in order to decide (subject to a preset threshold) whether to block or to update the parameters. At iteration k, the update law for the parameters is steepest ascent

$$p_{k+1} = p_k + a_k g_k \tag{7}$$

where the gradient vector $g_k = [(g_k)^1 (g_k)^2 \dots (g_k)^m]^T$ for the m-dimensional parameter space is determined by

$$(g_k)^i = \frac{L(p_k + c_k \Delta_k) - L(p_k - c_k \Delta_k)}{2c_k (\Delta_k)^i}, \text{ for } i = 1, 2 \dots m. \tag{8}$$

In this study, three parameters are to be updated at each iteration, i.e., $\mathrm{m} = 3$. Each element $(\Delta_k)^i$ of the vector, Δ_k takes on a value of $+1$ or -1, as generated by a Bernoulli distribution, and a_k and c_k are positive sequences of the form

$$a_k = \frac{a}{(k + A + 1)^{\alpha}} \tag{9}$$

$$c_k = \frac{c}{(k+1)^{\gamma}} \tag{10}$$

such that

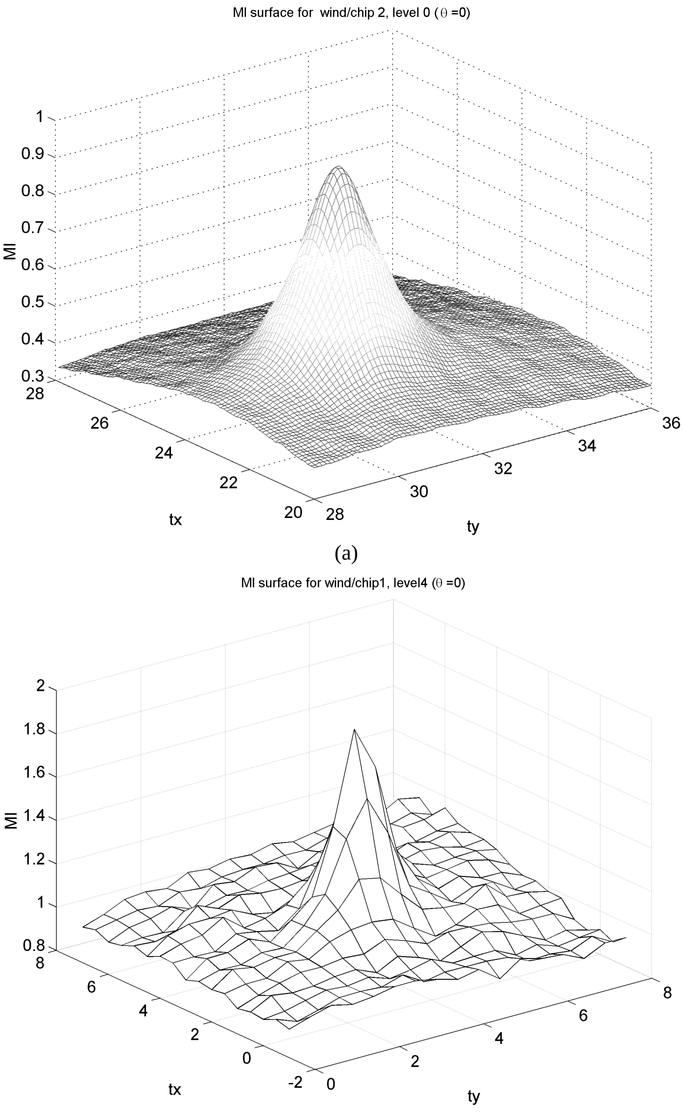
$$\begin{aligned} &a, c > 0,\ A \geq 0,\ 0 < \gamma < \alpha < 1 \\ &\text{and } \alpha - 2\gamma > 0,\ 3\gamma - \frac{\alpha}{2} \geq 0. \end{aligned} \tag{11}$$

The SPSA algorithm is a very powerful technique, which can get through some local maxima of the objective function to find the global maximum because of the stochastic nature of the gradient approximation. All the elements of p_k are randomly perturbed to obtain two measurements of $L(.)$. Each component of the gradient vector is then formed by the ratio defined in (8). The algorithm works by iterating from an initial guess of the optimal parameters, p_0 by using this calculated gradient. Spall [12] presents sufficient conditions for convergence of the SPSA iterative process in the stochastic *almost sure*. Convergence is established by requiring $L(.)$ to be sufficiently smooth (i.e., three times continuously differentiable) near the optimum, and imposing the following conditions on the gain sequences a_k and c_k, such that they go to zero at rates that are neither too fast nor too slow, i.e.,

$$\begin{aligned} &a_k, c_k > 0\ \forall k\ ;\ a_k \to 0,\ c_k \to 0 \text{ as } k \to \infty; \\ &\sum_k a_k = \infty,\ \sum_k \left(\frac{a_k}{c_k}\right)^2 < \infty. \end{aligned} \tag{12}$$

The elements of the perturbation vector Δ_k are required to be independent and symmetrically distributed about 0 with finite inverse moments $E\{1/|(\Delta_k)^i|\}$ for all k, i. The conditions on Δ_k make the gradient approximation, $g_k(.)$ an almost unbiased estimator of the true gradient $g(.)$, i.e., $E\{g_k(p_k)\} = g(p_k) + O(c_k^2)$. For c_k small, these misdirections act like random errors, which average and cancel out over a number of iterations.

When the transformed image is obtained using cubic B-spline interpolation, it produces a smooth MI surface as shown in Fig. 8(a). An important consideration in the application of the optimization scheme, is that the further away the initial guess is from the global maximum, the more local maxima the algorithm may need to overcome to reach the global maximum, and thus the more likely it is to fail. Note that the coarser the images (i.e., the deeper the level of the Simoncelli decomposition) the less smooth is the MI surface, and failure at this coarser level can be catastrophic to the optimization algorithm. For these smaller, lower resolution images, a further reduction in the number of bins in the histogram may be necessary, in order to get a smooth surface. As an illustration, Fig. 8(b) shows the MI surface for level 4 for one pair of images from our test dataset, where ripples on the MI surface can be seen as one moves away from the global maximum. These are indicative of local maxima, which may trap the algorithm causing it to fail. Significant smoothing of the MI surface at the coarsest decomposition level results from using a histogram with 64 bins, as opposed to 256 bins.



(a)

(b)

Fig. 8. Mutual information surfaces. (a) Spline-interpolated sub-pixel MI surface for one data pair at level 1 ($\theta = 0$). (b) MI surface at level 4, showing the global maximum and some local maxima.

VI. Experiments and Results

In this Section, multiresolution registration combining Simoncelli band-pass features, MI and the Spall optimization scheme is thoroughly tested and compared using synthetic test data as well as multitemporal data and remotely sensed imagery from different sensors. Results are also provided to compare MI with correlation. These experiments are conducted on an SGI Octane 195 MHz computer, and timing results are provided for that machine.

A. Description of the Parameters

Using Simoncelli filter size 9 and the Steerable Pyramid decomposition of Fig. 2, four levels of decomposition are computed and the feature space is composed of the gray levels of images $\{B3\}$, $\{B2\}$, $\{B1\}$ and $\{B0\}$. These images correspond to a decimation of 8, 4, 2, and 1 of the original image, respectively. The constants a, c, A, α and γ for the SPSA algorithm are chosen and optimized within the range of values suggested by Spall [12], which would ensure convergence. The chosen parameter values are $A = 100$, $c = 0.5$, $a = 12.0$, $\alpha = 0.602$, $\gamma = 0.101$ using a threshold of 0.1 for blocking; i.e., the parameter values are not updated if the MI value for the new point falls more than 0.1 below the current MI value. These values were found to work well for both MI and correlation for the datasets tested, so they were fixed for all the experiments to follow, providing a single frame of reference for the comparative study. In general, it may be more judicious to set the threshold at some percentage of the starting MI value.

B. Description of the Test Datasets

In this study, four datasets were used. For datasets 1– 3 below, only one band of each sensor was utilized. This is band 4 for Landsat-TM ("Thematic Mapper") data and band 2 for AVHRR-LAC (Local Area Coverage) data. These bands correspond to the Near-Infrared bands and usually show the best contrast of land features. In the future, an investigation could be done of whether a combination of several bands might improve the registration accuracy. The datasets are as follows:

1) From the same Landsat-TM ("Thematic Mapper") scene of the Pacific Northwest used to produce the image of Fig. 3, the 192×192 center of this image is extracted and utilized as the "Reference Image." "Input images" are artificially created by translating and rotating the original image and then extracting the 192×192 centers of the transformed images
    - translation parameters are varied in the horizontal direction by amounts of 0 to 5 pixels;
    - rotation parameters are varied with angles ranging from 0° to 6°.
2) The second set of images comes from a series of multitemporal NOAA Advanced Very High Resolution Radiometer (AVHRR) scenes which differ from the reference by very small translations and no rotations; these are shown in Fig. 9. These images are all of size 512×1024. Note the varying locations of clouds in the images.
3) The third dataset consists of seven pairs of images of size 256×256, each of which extracted from Band 4 of two scenes taken by Landsat-5 (in 1997) and Landsat-7 (in 1999) over the Chesapeake Bay area (Eastern United States). These pairs of images, shown in Fig. 10, are referred to as wind and chip respectively, and the Landsat-5 windows are registered to the corresponding Landsat-7 chips.

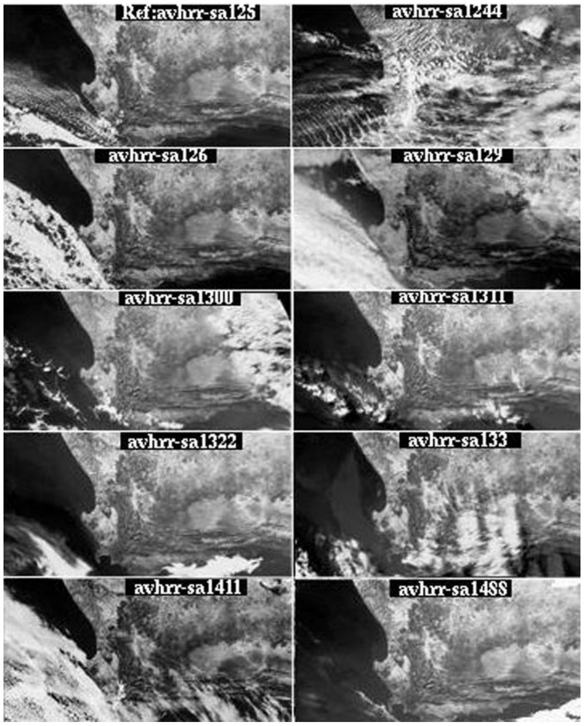


Fig. 9. Second dataset: Series of multitemporal AVHRR images over South Africa.

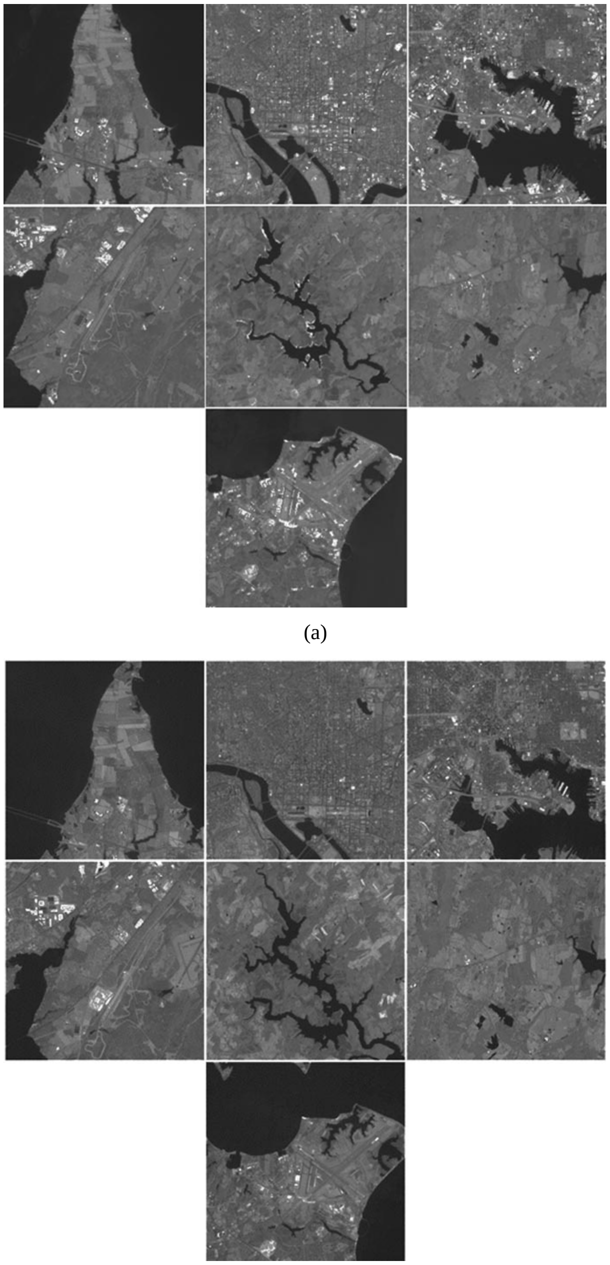
4) The fourth dataset used for this study represents multisensor data acquired by four different sensors over one of the MODIS Validation Core Sites. The site is the Konza Prairie in the state of Kansas, in the Middle West region of the United States. Overall, we consider eight different images corresponding to different bands of different sensors. The four sensors and their respective bands and spatial resolutions involved in this study are
    - IKONOS Bands 3 (Red) and 4 (Near-Infrared), spatial resolution of 4 meters per pixel, resampled to 3.91 m;
    - Landsat-7/ETM+ Bands 3 (Red) and 4 (Near-Infrared), spatial resolution of 30 meters per pixel, resampled to 31.25 m;
    - MODIS Bands 1 (Red) and 2 (Near Infrared), spatial resolution of 500 meters per pixel;
    - SeaWIFS Bands 6 (Red) and 8 (Near Infrared), spatial resolution of 1000 meters per pixel.

    Fig. 11 shows one band of each of these scenes.

C. Algorithm Implementation

First, we conduct a series of experiments using the synthetic images generated from the reference of dataset 1, to test the sensitivity of our algorithm to several parameters. Then based on our results, an automated optimization scheme is designed and applied to the remaining datasets (2–4) in a multiresolution manner. The optimization algorithm is tested on these multisensor and multitemporal datasets using both correlation and MI.

The optimization scheme starts with an "initial guess" of the correct registration value, based on prior information from a coarser registration scheme. The initial guess is then scaled to the corresponding starting value at the lowest decomposition level to be registered, and the optimization scheme is applied for a fixed number of iterations. The final registration translation-values at this level, are then doubled and passed with the



(a)

(b)

Fig. 10. Third dataset: (a) Seven chips (256 × 256) extracted from band 4 of a 1999 Landsat-7 Scene. (b) Seven corresponding windows (256 × 256) extracted from band 4 of a 1997 Landsat-5 scene.

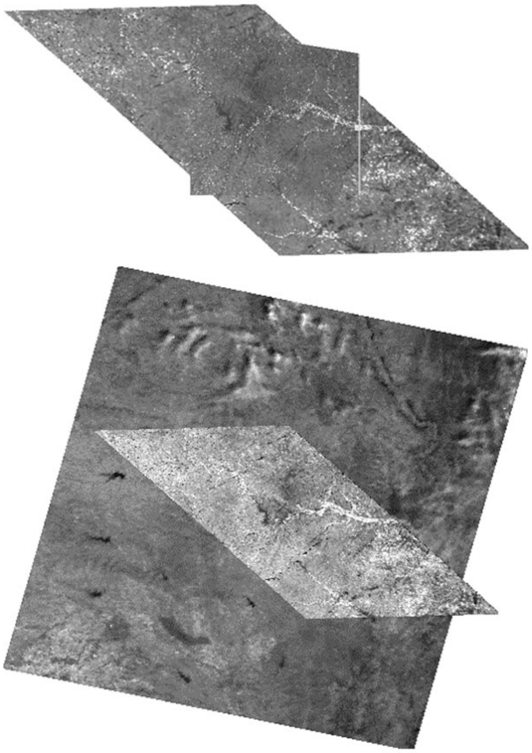


Fig. 11. Fourth dataset: IKONOS, Landsat/ETM, MODIS and SeaWIFS images of the Konza Prairie in Kansas, U.S.

rotation-value, up to the next level as a new starting point. This process is iterated up to level 1, which provides the final registration result. Note with this multiresolution approach, it is critical for a correct result to be obtained at the coarsest level of the decomposition so as not to propagate and multiply errors.

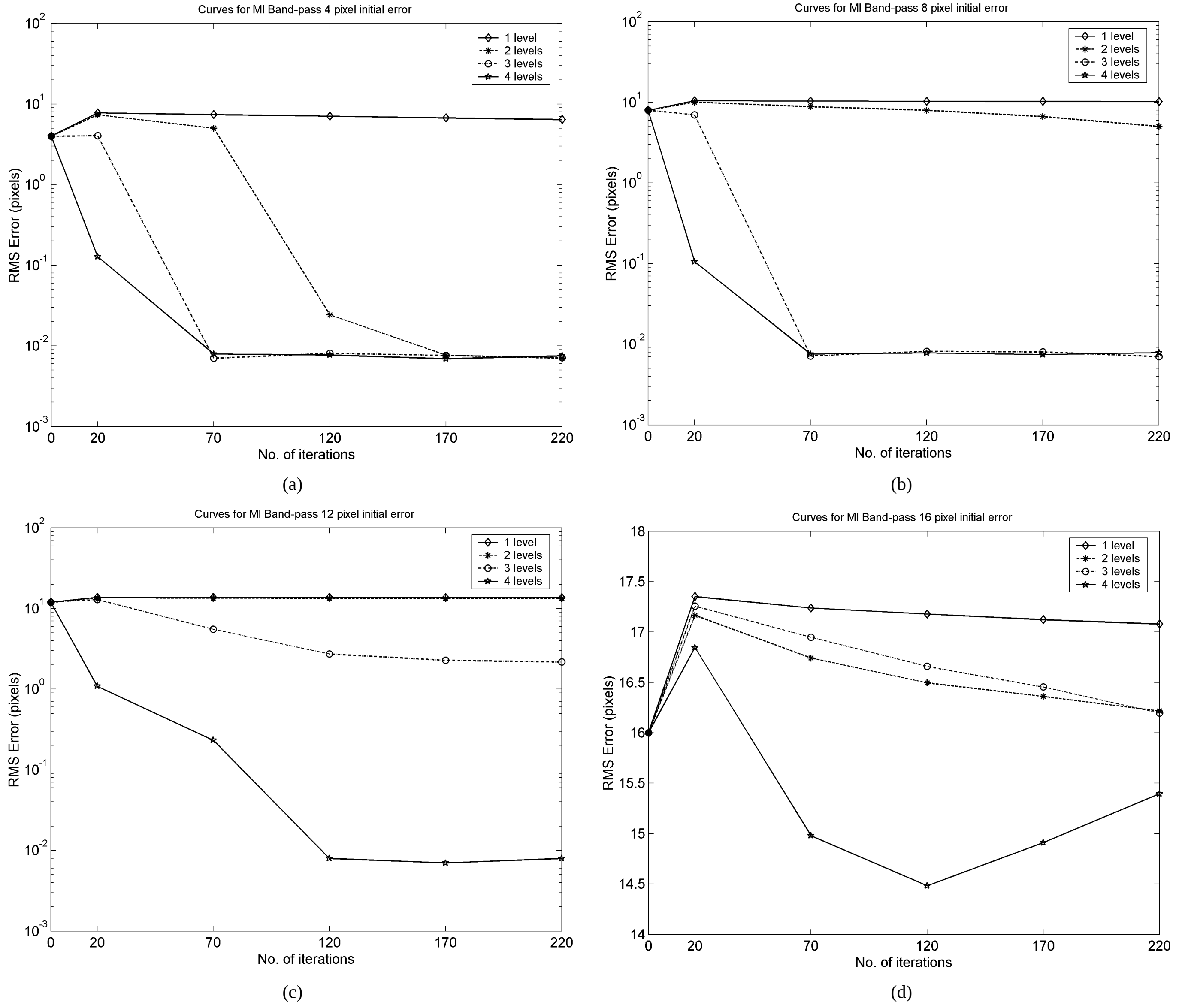
1) Sensitivity to Initial Guess and Number of Decomposition Levels: In this subsection, we test the sensitivity of our algorithm to the following parameters: the choice of the Simoncelli subband (low-pass versus band-pass), the number of levels of decomposition, and the distance between the initial guess and the correct result. Finally we compare MI to correlation in terms of their respective regions of attraction. These tests are performed using dataset 1. The plots of Fig. 12 correspond to MI optimization for the band-pass outputs of the Simoncelli decomposition for the images of dataset 1. They show the average of the final RMS errors measured in pixels, for the images of dataset 1 versus the number of iterations, for starting points (or initial guesses) at various horizontal distances from the correct result (or ground truth). Each starting point has a rotational error of 5°, in addition to the translational error indicated.

The average errors are computed as follows. For each of the 42 reference-input pairs, individual errors are computed by taking the root mean square (RMS) error over all the pixels in each image as follows:

$$\text{RMS Error} = \sqrt{\left(\frac{1}{N}\left(\sum_i \sum_j \|(x_i, y_j) - (x', y')_{ij}\|^2\right)\right)} \tag{13}$$

for $(x', y')_{ij} = T_{err}(x_i, y_j)$ *with* $T_{err} = T_c(T_{GT})^{-1}$; where T_{GT} represents the correct (“Ground Truth”) transformation and T_c is the computed transformation, $||.||$ is the Euclidean distance and N is the total number of pixels in the image. This error is averaged over all the image pairs.

For all the cases shown in Fig. 12, the algorithm consistently converges using four levels of decomposition, when the starting distance is 12 pixels or less in a single direction from the “ground truth” value. The algorithm fails at 16 pixels, with the error increasing with the number of iterations. This may be due to the algorithm getting trapped at a local maximum at a coarser level, with this incorrect registration being propagated through subsequent levels. For one level of decomposition,



(a)

(b)

(c)

(d)

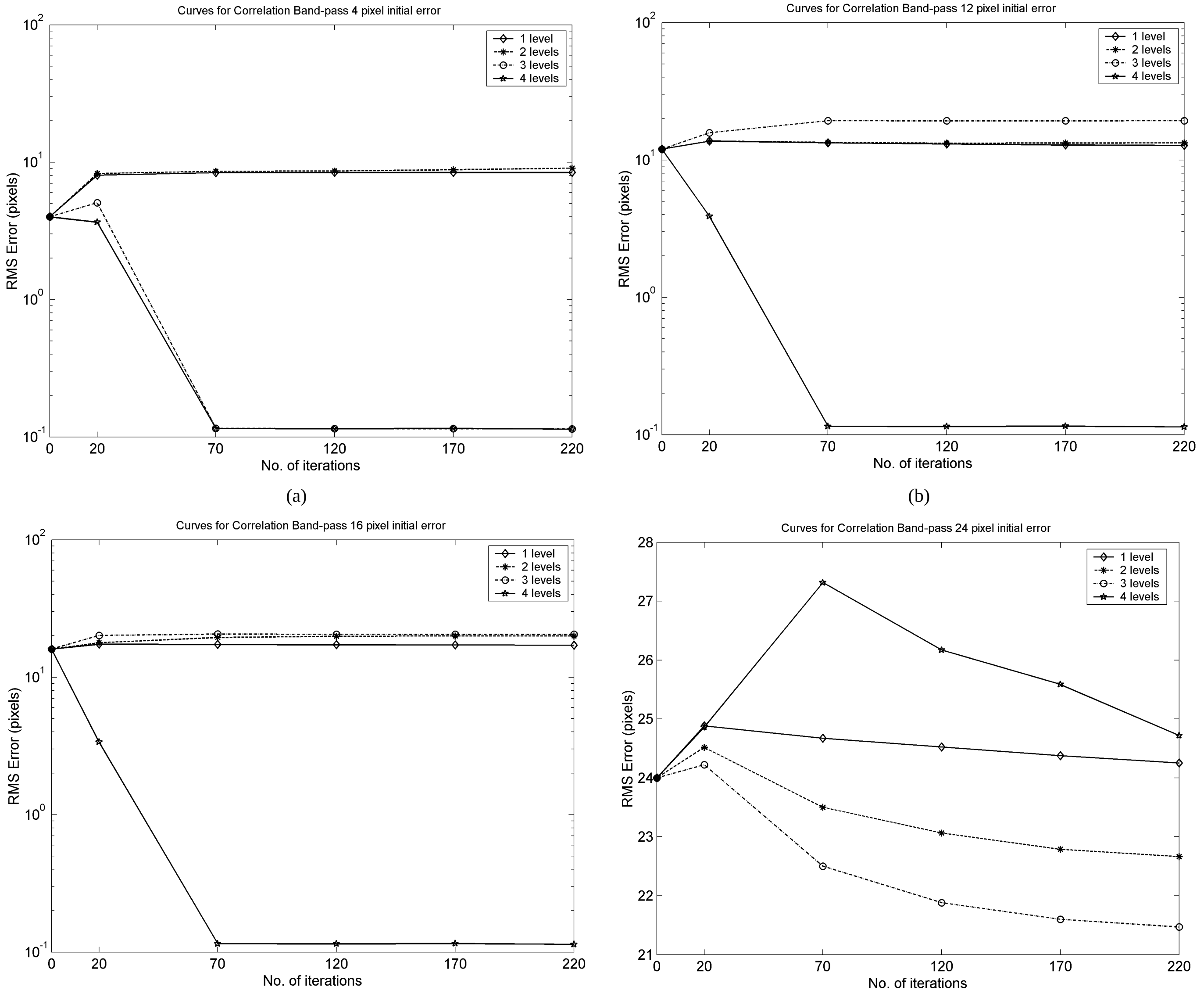
Fig. 12. RMS pixel error curves for MI with different initial distances over varying numbers of decomposition levels (band-pass). (a) $\text{Initial guess} = 4\ \text{pixels}$ from correct result. (b) $\text{Initial guess} = 8\ \text{pixels}$ from correct result. (c) $\text{Initial guess} = 12\ \text{pixels}$ from correct result. (d) $\text{Initial guess} = 16\ \text{pixels}$ from correct result. (Algorithm failure for all decomposition levels.).

registration is done using the band-pass output *B0* only, and for all cases the algorithm does not converge. Convergence can be achieved for this size image by using the original image, with no Simoncelli decomposition.

Similar plots were generated using the low-pass outputs of the Simoncelli decomposition. For the varying numbers of decomposition levels, the final value of the average error after 220 iterations was about the same as for the band-pass outputs (between 10^{-2} and 10^{-3}) with the low-pass being less sensitive than the band-pass to the distance of the initial guess from the correct result. However, when more complex test data is used, such as noisy and/or multisensor imagery, band-pass appears to achieve better precision than low-pass, while being just as robust. This is consistent with results reported in [20]. Based on these observations, the remaining tests are done using four levels of the Simoncelli band-pass output from a starting point, which is less than 12 pixels from the expected solution. We expect that such a starting point can be determined from a coarser registration scheme such as an exhaustive search [2].

The results for the identical experiment optimizing correlation for the band-pass outputs, are shown in Fig. 13. We note that in this case, algorithm failure occurs at a distance of 24 pixels from the "ground truth" values [see Fig. 13(d)].

Comparing the results of the experiments shown in Figs. 12 and 13, we note that correlation converges if the starting distance is less than 24 pixels from the optimum point, and we say that its optimum has a region of attraction of about 24 pixels. With a similar definition, the MI optimum has an attraction region of about 16 pixels. Inspecting the plots of Fig. 5, at level 4 we observe that the neighborhood *V0*, defined by inflexion points, is 3 pixels for correlation, which is consistent with 24 pixels in full resolution units, and it is 2 pixels for MI, which is consistent with 16 pixels in full resolution units. We also note that MI achieves better accuracy than correlation, since after 220



(a)

(b)

(c)

(d)

Fig. 13. RMS pixel error curves for correlation with different initial distances over varying numbers of decomposition levels (band-pass). (a) $\mathrm{Initial\ guess} = 4\ \mathrm{pixels}$ from correct result. (b) $\mathrm{Initial\ guess} = 12\ \mathrm{pixels}$ from correct result. (c) $\mathrm{Initial\ guess} = 16\ \mathrm{pixels}$ from correct result. (d) $\mathrm{Initial\ guess} = 24\ \mathrm{pixels}$ from correct result. (Algorithm failure for all decomposition levels.).

iterations the sub-pixel precision of the final result is 10^{-1} for correlation, compared to about 10^{-2} for MI optimization.

D. Results on Multitemporal and Multisensor Imagery

Tables I and II show details of the optimization algorithm applied to dataset 3, referred to as the wind and chip image pairs, for a total of 10 iterations only. Convergence occurred to a "reasonable" set of final parameters for all the pairs in this dataset, and intermediate results are provided at all four levels of the Simoncelli decomposition. The initial guess for starting the optimization, is about 8 pixels away from the final registration value in the x-direction, and less than 4 pixels in the y-direction. Results of using MI are provided in Table I, while those for correlation are given in Table II. Note the similar timings for the two metrics when using the same number of iterations.

Since no good ground truth is available for this dataset, we evaluate these results visually by obtaining the mosaics using the SPSA registration values of [23,32,0] for [wind2, chip2], and [23,35,0] for [wind5, chip5], as shown in Fig. 14.

Table III provides the results for the AVHRR images (i.e., dataset 2) with four levels of decomposition. For this dataset the average RMS error between the manual registration values and those from the MI optimization is 0.6385 pixels, while the average error from the correlation optimization is 0.5156 pixels.

Results for the multisensor images of dataset 4 are provided in Table IV. For the multisensor images, the average error between the manual registration values and those from the MI optimization is 0.3446, while the average error with the correlation optimization is 1.2522, and sub-pixel accuracy is not achieved on average. The correlation error is skewed by the much larger error produced by the NIR pair of modis and etm. Excluding this data pair, the average error is 0.3538 for MI versus 0.4756 for correlation. It is important to also note that manual registration values were not provided at the sub-pixel level.

TABLE I
WIND AND CHIP, MUTUAL INFORMATION USING SIMONCELLI DECOMPOSITIONS: STARTING POINT (16,32,0), $\mathrm{Max.\ No.\ Iterations} = 10$

Images	Level	Starting MI	Ending Pt (Tx/Ty/θ)	Ending MI	Iterations to Peak	Total Run-Time
			Full Resolution Units			(secs)
wc/ 0	4	1.898177	23.568/30.312/-1.025	2.433614	7	
	3	1.509950	23.032/31.096/-0.252	1.734695	9	
	2	1.185174	23.350/31.152/0.119	1.294970	9	
	1	0.969451	23.327/31.158/0.071	0.973435	3	26.0763
wc/ 1	4	1.267940	22.408/32.920/-0.781	1.806131	6	
	3	1.322530	22.760/33.056/0.873	1.365593	2	
	2	0.743350	23.352/33.100/0.082	1.445919	9	
	1	1.071002	23.353/32.937/-0.156	1.086046	3	25.9383
wc/ 2	4	1.759831	22.920/32.256/-0.970	2.398865	4	
	3	1.392357	24.224/33.436/-0.225	1.448109	3	
	2	0.741066	23.810/32.042/0.382	1.134174	9	
	1	0.716551	23.085/32.254/-0.017	1.010505	9	26.0267
wc/ 3	4	1.329628	22.160/31.840/-1.070	1.799868	6	
	3	1.114694	22.088/30.896/-0.133	1.298333	9	
	2	0.905466	22.576/30.898/0.098	1.055709	9	
	1	0.845633	22.680/31.187/0.027	0.906563	9	26.1428
wc/ 4	4	1.668014	22.816/32.344/-0.518	2.077568	4	
	3	1.245365	22.360/32.664/-0.112	1.337770	9	
	2	0.940130	22.446/32.946/0.033	0.977274	3	
	1	0.714340	22.552/32.975/0.016	0.726304	9	26.0603
wc/ 5	4	0.931030	22.432/34.992/0.124	1.459114	5	
	3	1.280736	23.392/35.400/0.409	1.296779	3	
	2	0.924769	32.222/34.766/-0.209	1.100453	7	
	1	0.762982	23.076/34.996/0.075	0.866124	9	25.7704
wc/ 6	4	1.671403	22.088/31.704/-0.489	2.216321	6	
	3	1.443503	23.100/31.352/-0.392	1.746211	9	
	2	1.144628	23.808/30.474/0.352	1.376146	8	
	1	0.938186	23.669/30.825/-0.090	1.104534	9	26.0889

TABLE II
WIND AND CHIP, CORRELATION USING SIMONCELLI DECOMPOSITIONS: STARTING POINT (16,32,0), $\mathrm{Max.\ No.\ Iterations} = 10$

Images	Level	Starting Corre	Ending Pt (Tx/Ty/θ)	Ending Corre	Iterations to Peak	Total Run-Time
			Full Resolution Units			(secs)
wc/ 0	4	0.415278	22.584/30.960/-0.417	0.554160	9	
	3	0.405324	23.072/30.988/-0.314	0.410104	8	
	2	0.304596	23.330/31.006/-0.125	0.310164	9	
	1	0.205008	23.355/31.084/0.024	0.210094	9	25.4361
wc/ 1	4	0.326058	20.744/30.056/-0.236	0.398916	3	
	3	0.311130	22.384/31.996/0.027	0.354633	7	
	2	0.285719	23.152/32.810/0.084	0.332899	9	
	1	0.244948	23.303/33.013/0.033	0.260912	9	25.3361
wc/ 2	4	0.452410	22.120/31.312/-0.292	0.556309	9	
	3	0.387133	22.892/32.052/-0.156	0.395836	8	
	2	0.396167	23.128/32.152/-0.011	0.397172	2	
	1	0.227161	23.170/32.190/0.036	0.228114	9	25.3749
wc/ 3	4	0.243187	20.816/32.240/-0.161	0.302934	8	
	3	0.321629	22.136/31.428/-0.177	0.339906	9	
	2	0.302096	22.444/31.328/-0.065	0.306688	9	
	1	0.221331	22.559/31.253/-0.003	0.224032	9	25.4992
wc/ 4	4	0.290710	20.976/32.240/-0.158	0.353973	9	
	3	0.338284	21.736/32.432/-0.096	0.346130	9	
	2	0.252980	22.036/32.600/-0.021	0.258952	9	
	1	0.138192	22.109/32.661/0.003	0.140181	9	25.4121
wc/ 5	4	0.221524	20.264/33.040/-0.055	0.265966	7	
	3	0.188845	22.068/34.292/0.097	0.223079	8	
	2	0.181546	22.620/34.696/0.131	0.201336	9	
	1	0.175173	22.792/34.800/0.046	0.187496	9	25.2349
wc/ 6	4	0.350946	22.208/31.376/-0.296	0.448415	9	
	3	0.404347	23.172/30.904/-0.283	0.41938	9	
	2	0.373416	23.638/30.766/-0.085	0.387857	9	
	1	0.291206	23.730/30.772/0.013	0.294570	9	25.4178

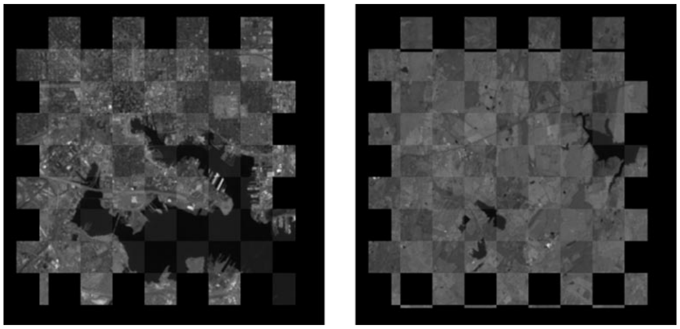
E. Parameter Convergence

By further expanding the results of Tables I and II for the [wind2, chip2] data pair of dataset 3, we can observe the convergence rates using MI optimization versus correlation optimization. In Fig. 15, the plots show the convergence rate of the relevant parameters with the optimization of MI compared with that of correlation. We compare convergence for the original [wind2, chip2] images with an "arbitrary" starting point of $[tx, ty, \theta] = [20, 35, 0]$, and also for Simoncelli decomposition level 1 using the starting point obtained from the previous three-level optimization. For the original image with no pyramid decomposition, one observes that using MI optimization, each of parameters converge in about one third the number of iterations required by correlation optimization. Note that the wavelet starting points at level 1 are very close to the optimum in all cases.

The timing for the 4-level registration of [wind2, chip2] from the starting point $[tx, ty, \theta] = [20, 35, 0]$ over 400 iterations is 999.7 s for MI, while that for correlation is 985.9 s. Nevertheless, it is important to note that for the original [wind2, chip2] image pair over 400 iterations, the maximum MI value is achieved in 72 iterations, while the maximum correlation is achieved at 395 iterations.

VII. DISCUSSION AND CONCLUSIONS

Prior work on optimization techniques for image registration can be found in references [3]–[5] and [9]–[11]. The techniques described in [9]–[11] are all based on minimizing a sum of square

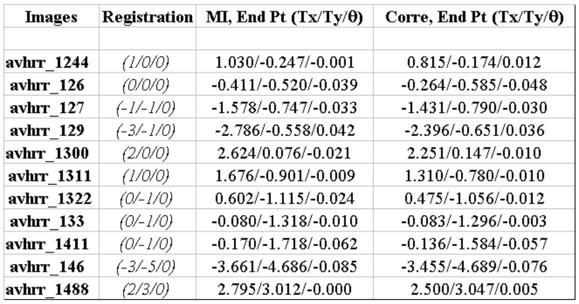


(a) (b)

Fig. 14. Checkerboard Mosaiced images using SPSA optimization values. (a) SPSA registration for [wind2,chip2]. (b) SPSA registration for [wind5,chip5].

TABLE III

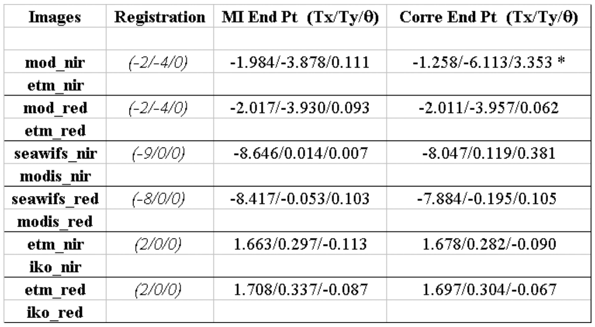
AVHRR, MUTUAL INFORMATION AND CORRELATION, USING SIMONCELLI DECOMPOSITIONS: STARTING POINT (0,0,0), $\mathrm{Max.\ No.\ Iterations} = 250$



Images	Registration	MI, End Pt (Tx/Ty/θ)	Corre, End Pt (Tx/Ty/θ)
avhrr_1244	*(1/0/0)*	1.030/-0.247/-0.001	0.815/-0.174/0.012
avhrr_126	*(0/0/0)*	-0.411/-0.520/-0.039	-0.264/-0.585/-0.048
avhrr_127	*(-1/-1/0)*	-1.578/-0.747/-0.033	-1.431/-0.790/-0.030
avhrr_129	*(-3/-1/0)*	-2.786/-0.558/0.042	-2.396/-0.651/0.036
avhrr_1300	*(2/0/0)*	2.624/0.076/-0.021	2.251/0.147/-0.010
avhrr_1311	*(1/0/0)*	1.676/-0.901/-0.009	1.310/-0.780/-0.010
avhrr_1322	*(0/-1/0)*	0.602/-1.115/-0.024	0.475/-1.056/-0.012
avhrr_133	*(0/-1/0)*	-0.080/-1.318/-0.010	-0.083/-1.296/-0.003
avhrr_1411	*(0/-1/0)*	-0.170/-1.718/-0.062	-0.136/-1.584/-0.057
avhrr_146	*(-3/-5/0)*	-3.661/-4.686/-0.085	-3.455/-4.689/-0.076
avhrr_1488	*(2/3/0)*	2.795/3.012/-0.000	2.500/3.047/0.005

TABLE IV

MULTI-SENSOR, MUTUAL INFORMATION AND CORRELATION, USING SIMONCELLI DECOMPOSITIONS: STARTING POINT (0,0,0), $\mathrm{Max.\ No.\ Iterations} = 250$

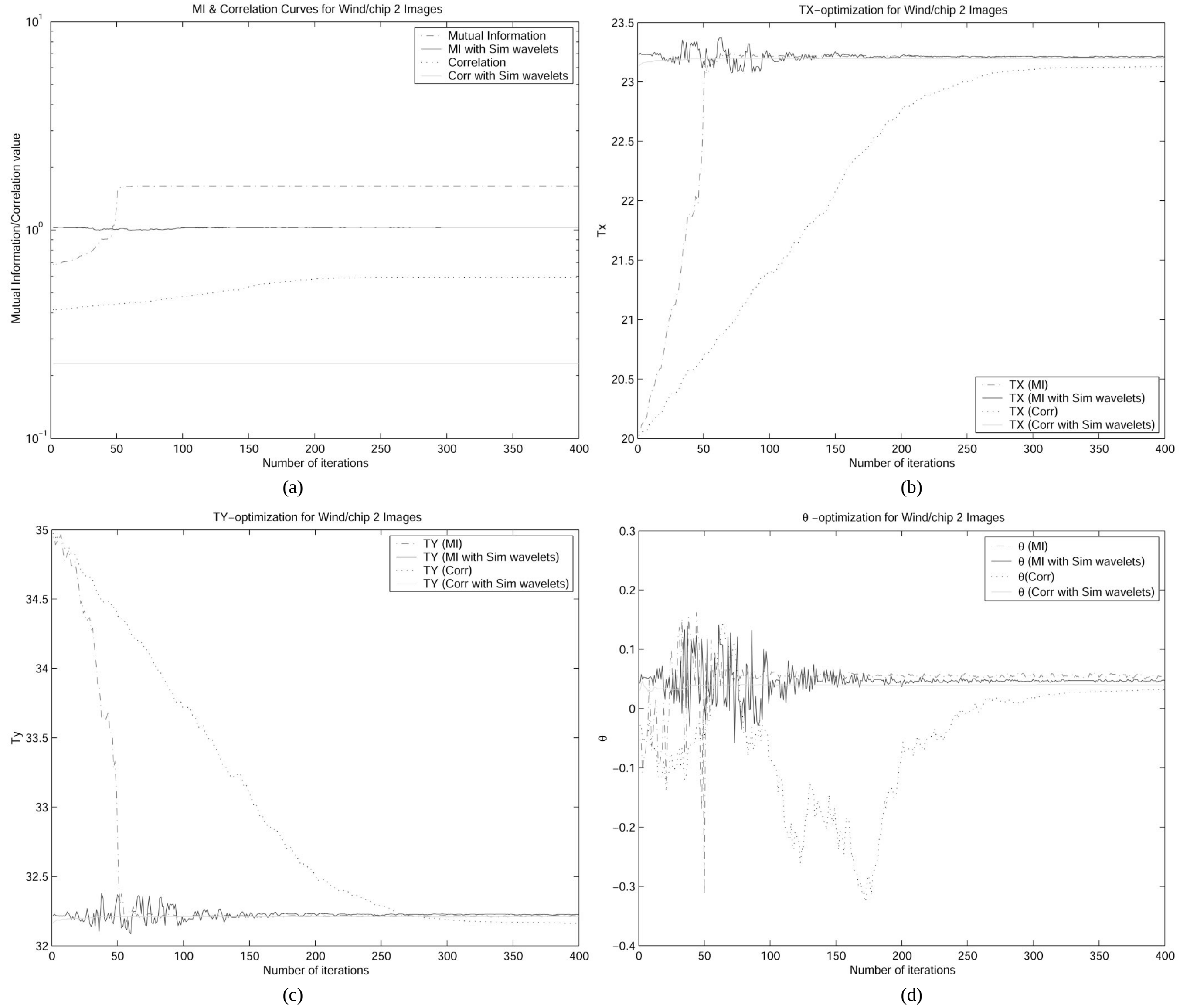


Images	Registration	MI End Pt (Tx/Ty/θ)	Corre End Pt (Tx/Ty/θ)
mod_nir	*(-2/-4/0)*	-1.984/-3.878/0.111	-1.258/-6.113/3.353 *
etm_nir			
mod_red	*(-2/-4/0)*	-2.017/-3.930/0.093	-2.011/-3.957/0.062
etm_red			
seawifs_nir	*(-9/0/0)*	-8.646/0.014/0.007	-8.047/0.119/0.381
modis_nir			
seawifs_red	*(-8/0/0)*	-8.417/-0.053/0.103	-7.884/-0.195/0.105
modis_red			
etm_nir	*(2/0/0)*	1.663/0.297/-0.113	1.678/0.282/-0.090
iko_nir			
etm_red	*(2/0/0)*	1.708/0.337/-0.087	1.697/0.304/-0.067
iko_red			

differences. Maes *et al.* [3] use Powell's method to optimize MI. Following this gradient-based methods were investigated in [22], which uses an explicit calculation of the required derivative based on a partial volume interpolation of the criterion, and the search is implemented in a multiresolution framework. Irani and Peleg [10] choose to minimize the square error of a "disparity vector" between the two images. It proceeds by a Newton-Raphson technique, and also requires computation of the necessary gradients. The scheme described in [10] does not involve multiple resolutions of the images. Finally, Eastman *et al.* [11] integrate the gradient-descent techniques described in [9] and [10] in a multiresolution framework, while focusing on the radiometric component of the registration transform which is associated with the different viewing conditions of multitemporal or multisensor data. Thevenaz *et al.* in [9], develop a scheme to optimize an integrated sum of square differences in the intensity values of the images, which works in a multiresolution manner. They use a Marquardt-Levenberg algorithm, and computations of the derivatives and of the Hessian matrix are based on a spline pyramid. Their work is applied to medical imagery, and is extended in [5] to the maximization of the MI similarity criterion.

The registration algorithm proposed by Thevenaz and Unser in [5], solves a problem similar to the one described here. Their algorithm is based on a combination of MI together with a multiresolution gradient search. By using the spline data model both for image interpolation and for the probability density estimation with Parzen windows, smoothing is achieved and the gradient components of MI are computed exactly in a deterministic fashion. An optimizer similar to the Levenberg-Marquardt is then designed specifically for this criterion.

The algorithm presented here is generally simpler and thus less computationally intensive, while the optimizer in [5] is more involved and may therefore be more robust. Our gradient components are computed approximately and stochastically, and we also use trivial windowing in the form of a reduced number of histogram bins, to achieve smoothing. In addition, our search strategy is essentially gradient ascent, which is robust when far from the solution but it converges more slowly



(a)

(b)

(c)

(d)

Fig. 15. Parameter optimization curves for mutual information versus correlation for [wind2,chip2] data pair. (a) Optimization curves for MI and correlation. (b) Optimization curves for Tx. (c) Optimization curves for Ty. (d) Optimization curves for rotation, θ.

than the Levenberg-Marquardt type optimizer of [5], when close to the solution. Also because of the stochastic nature of the gradient approximation, our algorithm exhibits a somewhat oscillatory convergence behavior compared to the smooth convergence in [5]. It is unclear which algorithm performs better under various conditions, and more testing is necessary to evaluate this, but this is beyond the scope of this paper. However, we note that due to the simplicity of its components, our algorithm may yield itself more easily to a distributed or parallel implementation, which may be essential for real-time processing of satellite scenes.

The study presented in this paper has applied the SPSA optimization technique for the registration of remote sensing images in a multiresolution framework, using Simoncelli wavelet-like filters. In the multiresolution approach provided by this steerable decomposition, when convergence occurs at a coarser level, it provides a near optimal starting point for the next level. This can produce immediate convergence at that level, providing a considerable speed up in the overall registration process. The multiresolution approach also increases the robustness of the algorithm since it is less likely to get trapped in a local maximum at the higher resolutions. From Figs. 12 and 13, we note that the algorithm consistently converges when using 4 decomposition levels for registration, provided that the initial starting point is not too far from the global optimum.

On average for these experiments, registration of a 256×256 image over the same number of iterations, took about equal time for MI with 64 bins as for correlation on an SGI Octane 195 Mhz computer. The advantage of using MI optimization over correlation can be found in its faster convergence rate in terms of number of iterations. MI was generally observed to converge in about one third the number of iterations required by correlation. In this work, the algorithm was run for a fixed number of iterations, in the future we will investigate the definition of an automatic stopping criterion for the optimization.

Using the area under the curve as a measure of sharpness of the MI and correlation peaks, it was shown in Section IV-A that the MI curve for the original gray levels is about 6 times as sharp as the correlation curve, but it is 2.5 times as sharp when using the Simoncelli sub-bands in a neighborhood about the optimum. Thus it is possible that faster convergence and better precision can be achieved at the finest level of the decomposition by using MI together with the original gray level images, in place of the level 1 outputs of the Simoncelli wavelets. We also observe that while the MI curve is convex around the optimum for the registration of the synthetically generated images in Figs. 4 and 5, this curve becomes concave for the real-life images of Fig. 10, as shown in Fig. 8(a). This may indicate less precise registration for those images, but it also allows for the possibility of applying second order optimization methods.

Current work involves the inclusion of isometric scaling as an additional parameter to be optimized by the algorithm. The experiment using the multisensor images of dataset 4 indicates that the scheme presented here may, in fact, work well for multisensor registration also. We will continue to test this algorithm on other types of datasets in future work, and its performance will be compared to other registration schemes [13].

Acknowledgment

The authors would also like to thank J. Masek and N. Netanyahu for the Landsat dataset of Fig. 10, as well as J. Morisette and the EOS Validation Core Sites Project for the Konza Prairie multisensor dataset shown in Fig. 11. The authors would like to thank the anonymous reviewers for their many helpful suggestions.

References

[1] L. G. Brown, “A survey of image registration techniques,” *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, 1992.

[2] J. Le Moigne, W. J. Campbell, and R. F. Cromp, “An automated parallel image registration technique of multiple source remote sensing data,” *IEEE Trans. Geosci. Remote Sensing*, vol. 40, pp. 1849–1864, Aug. 2002.

[3] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” *IEEE Trans. Med. Imag.*, vol. 16, Apr. 1997.

[4] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, “Multi-modal volume registration by maximization of mutual information,” *Med. Imag. Anal.*, vol. 1, pp. 35–51, 1996.

[5] P. Thevenaz and M. Unser, “Optimization of mutual information for multiresolution image registration,” *IEEE Trans. Image Processing*, vol. 9, pp. 2083–2099, Dec. 2000.

[6] J. Le Moigne and I. Zavorin, “Use of wavelets for image registration,” in *Proc. SPIE Aerosense 2000, Wavelet Applications VII*, Orlando, FL, Apr. 24–28, 2000.

[7] H. Stone, J. Le Moigne, and M. McGuire, “The translation sensitivity of wavelet-based registration,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 1074–1080, Oct. 1999.

[8] E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger, “Shiftable multiscale transforms,” *IEEE Trans. Inform. Theory*, vol. 38, Mar. 1992.

[9] P. Thevenaz, U. E. Ruttiman, and M. Unser, “A pyramid approach to subpixel registration based on intensity,” *IEEE Trans. Image Processing*, vol. 7, pp. 27–41, Jan. 1998.

[10] M. Irani and S. Peleg, “Improving resolution by image registration,” *CVGIP: Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, May 1991.

[11] R. Eastman and J. Le Moigne, “Gradient-descent techniques for multitemporal and multi-sensor image registration of remotely sensed imagery,” in *Proc. FUSION'2001, 4th Int. Conf. Information Fusion*, Aug. 2001.

[12] J. C. Spall, “Multivariate stochastic approximation using a simultaneous perturbation gradient approximation,” *IEEE Trans. Automat. Contr.*, vol. 37, no. 3, pp. 332–341, 1992.

[13] J. LeMoigne, A. Cole-Rhodes, R. Eastman, K. Johnson, J. Morisette, N. Netanyahu, H. Stone, and I. Zavorin, “Multi-sensor registration of remotely sensed imagery,” in *Proc. 8th Int. Symp. Remote Sensing*, vol. 4541, Toulouse, France, Sept. 2001.

[14] K. Johnson, A. Cole-Rhodes, I. Zavorin, and J. Le Moigne, “Mutual information as a similarity measure for remote sensing image registration,” in *Proc. SPIE Aerosense 2001, Geo-Spatial Image and Data Exploitation II*, vol. 4383, Orlando, FL, Apr. 2001, pp. 51–61.

[15] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*. New York: Springer-Verlag, 1997.

[16] W. K. Pratt, “Correlation techniques of image registration,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 3, pp. 353–358, 1974.

[17] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, 2nd ed. New York: Academic, 1982, vol. 1.

[18] M. Unser, A. Aldroubi, and M. Eden, “B-spline signal processing : Part 1 – Theory,” *IEEE Trans. Signal Processing*, vol. 41, pp. 821–833, Feb. 1993.

[19] A. Cole-Rhodes, K. Johnson, and J. LeMoigne, “Multi-resolution registration of remotely sensed images using stochastic gradient,” in *Proc. SPIE Aerosense 2002, Wavelet Applications IX*, vol. 4738, Orlando, FL, Apr. 2002, pp. 44–55.

[20] I. Zavorin, H. Stone, and J. LeMoigne, “Iterative pyramid-based approach to subpixel registration of multisensor satellite imagery,” in *Proc. SPIE Int. Symp. Optical Science and Technology 2002, Earth Observing Systems VII*, Seattle, WA, July 2002.

[21] J. B. Maintz and M. A. Viergever, “A survey of medical image registration,” *Med. Image Anal.*, vol. 2, pp. 1–36, Apr. 1998.

[22] F. Maes, D. Vandermeulen, and P. Suetens, “Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information,” *Med. Image Anal.*, vol. 3, no. 4, pp. 373–386, December 1999.

[23] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation and Control*. Hoboken, NJ: Wiley, 2003.

Arlene A. Cole-Rhodes (S'86–M'88) received the B.Sc. degree in applied mathematics from Warwick University, Warwick, U.K., and the M.Phil. degree from Cambridge University, Cambridge, U.K. She received the Ph.D. in electrical engineering from the University of California at Berkeley in 1989.

She joined the Machine Perception Research Department at AT&T Bell Laboratories in Holmdel, NJ, as a Member of Technical Staff in 1990. Since 1993, she has been with the Department of Electrical and Computer Engineering at Morgan State University, Baltimore, MD, where she is currently Associate Professor. Her early research was in the area of control of robot manipulators. Current research interests are in the development of algorithms for the registration and fusion of multisensor images for remote-sensing and biomedical applications. She is also involved in developing algorithms for the detection and estimation of signals in wireless multiple access communication.

Kisha L. Johnson received the B.S. and M.S. degrees in electrical engineering from Morgan State University, Baltimore, MD, in 1999 and 2001, respectively. Her thesis work is a result of collaborative research with NASA Goddard Space Flight Center's Applied Information Sciences Branch. She is currently pursuing the Ph.D. degree in electrical engineering at Morgan State University.

Her current research interests include signal processing and image registration using mutual information.

Jacqueline LeMoigne (M’92–SM’96) received the Ph.D. degree from INRIA and the University Pierre and Marie Curie, Paris, France, and completed postdoctoral studies at the Computer Vision Laboratory, University of Maryland, College Park.

She is a Senior Computer Scientist in the Applied Information Sciences Branch of the NASA/Goddard Space Flight Center. She joined Goddard in 1990 first as a National Research Council Senior Research Associate and then as a Senior Scientist at the Center of Excellence in Space Data and Information Sciences (CESDIS). Since then, she has focused her research interests on computer vision utilizing high-performance parallel computers and applied to earth and space science problems. Her most recent research focuses on parallel registration of multisensor/multiscale satellite image data, for which she has been studying wavelets and their implementation on high performance parallel computers. Her current work includes the development of an image registration toolbox, the registration of Landsat imagery, the implementation of image registration methods on field programmable gate arrays, and automatic multisensor registration for application to on-board processing and formation flying systems. More recent work is also being performed in the areas of image fusion and dimension reduction. She was an Associate Editor for *Pattern Recognition* from 2001 to 2003.

Dr. LeMoigne has been an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING since 2001.

Ilya Zavorin received the B.S. degree in mathematics and computer science from Santa Clara University, Santa Clara, CA, in 1993, the M.S. degree in applied mathematics in 1997, and the Ph.D. degree in applied mathematics in 2001, both from the University of Maryland, College Park.

He is currently a Research Associate with the Goddard Earth Science and Technology Center, University of Maryland, Baltimore, working at NASA Goddard Space and Flight Center on the development of efficient algorithms for automatic registration and fusion of remotely sensed imagery. His research interests also include accuracy assessment, optimization, and numerical linear algebra.

Dr. Zavorin is a member of SPIE and SIAM.