Robust Multispectral Image Registration Using Mutual-Information Models

Jeffrey P. Kern and Marios S. Pattichis, *Senior Member, IEEE*

***Abstract*—Image registration is a vital step in the processing of multispectral imagery. The accuracy to which imagery collected at multiple wavelengths can be aligned directly affects the resolution of the spectral end products. Automated registration of the multispectral imagery can often be unreliable, particularly between visible and infrared imagery, due to the significant differences in scene reflectance at different wavelengths. This is further complicated by the thermal features that exist at longer wavelengths. We develop new mathematical and computational models for robust image registration. In particular, we develop a frequency-domain model for the mutual-information surface around the optimal parameters and use it to develop a robust gradient ascent algorithm. For a robust performance, we require that the algorithm be initialized close to the optimal registration parameters. As a measure of how close we need to be, we propose the use of the correlation length and provide an efficient algorithm for estimating it. We measure the performance of the proposed algorithm over hundreds of random initializations to demonstrate its robustness on real data. We find that the algorithm should be expected to converge, as long as the registration parameters are initialized to be within the correlation-length distance from the optimum.**

***Index Terms*—Image registration, multispectral imagery, mutual-information models.**

I. Introduction

IMAGE registration is a fundamental operation in the processing of multispectral imagery, which is often required prior to interpreting images or any further processing. Image registration is the operation of aligning images that, in general, are of the same object but collected under different viewing conditions and/or using separate imaging devices.

Thorough surveys of image registration techniques have been published by Brown in 1992 [1], Zitova and Flusser [2], and Pluim *et al.* [3]. When registering image pairs, a common approach is to designate one image as the reference and apply a spatial translation, rotation, scaling, or warping function to the second image in order to bring it into alignment with the reference. Some measure of similarity between the images is used to determine when the optimal alignment has occurred. Although several measures have been studied, the most widely used similarity measures used in image registration are those based on cross correlation [1], [2].

Manuscript received June 29, 2004; revised December 15, 2006.

J. P. Kern is with the Sandia National Laboratories, Albuquerque, NM 87185-0972 USA.

M. S. Pattichis is with the Image and Video Processing and Communications Laboratory, Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131-1356 USA (e-mail: pattichis@ece.unm.edu).

Digital Object Identifier 10.1109/TGRS.2007.892599

A fundamental limitation of prior research in mutual-information methods is that it did not effectively address the issue of convergence of the proposed methods. We use the term "robust" to describe new mutual-information methods that can guarantee their convergence to the correct parameters for a wide range of parameter initializations.

We demonstrate many of the issues in Figs. 1–3. In Fig. 1, we show two spectral bands of desert images from Shiprock, NM. To compute the mutual-information plots of Figs. 2 and 3, we use the popular histogram method. From Figs. 2 and 3, we see that a large number of samples (100 000 in this example) are needed to obtain accurate estimates of the mutual-information surface. When a realistic, limited number of samples are used, the mutual-information surface tends to be extremely noisy, as demonstrated in Fig. 3.

Clearly, the discrete derivatives of the mutual-information surface of Fig. 3 will tend to be extremely noisy. The problem persists even if the underlying probability density functions (pdfs) are estimated using the kernel functions (Section II). Convergence from such noisy estimates cannot be guaranteed, except when the algorithm that is used for maximizing the mutual-information surface is initialized in the neighborhood of the actual maximum (see Fig. 3). In this paper, we present new mathematical and computational models that allow us to estimate the size of the neighborhood around the maximum and then use this information to demonstrate convergence. This problem has not been effectively addressed for remote sensing [4]. In [4, p. 2452], failures to capture the global optimum are handled by changing "the initial search points until the correct optimum is obtained." No details are given on how this is done and how many new search points need to be attempted.

On the other hand, interpolation artifacts and how they can affect convergence have already been addressed in the literature [3]–[5]. The basic problem can be demonstrated in Fig. 4. Image entropy should not be a function of how the image is sampled. It should remain constant. Yet, for nonintegral translations, the interpolation method averages image intensity from nearby pixels, resulting in a reduction of the estimated entropy. On the other hand, over integral translations, no interpolation is needed, and thus, no averaging is needed. These observations result in the peaks at integer translations, as demonstrated in Fig. 4. On the other hand, we note the minima at the midpoints between the integers. The interpolation effects on the estimation of the mutual-information function have also been demonstrated in [4, Fig. 2]. The overall effect of these interpolation artifacts is to significantly limit the registration accuracy [3], [4].

0196-2892/$25.00 © 2007 IEEE

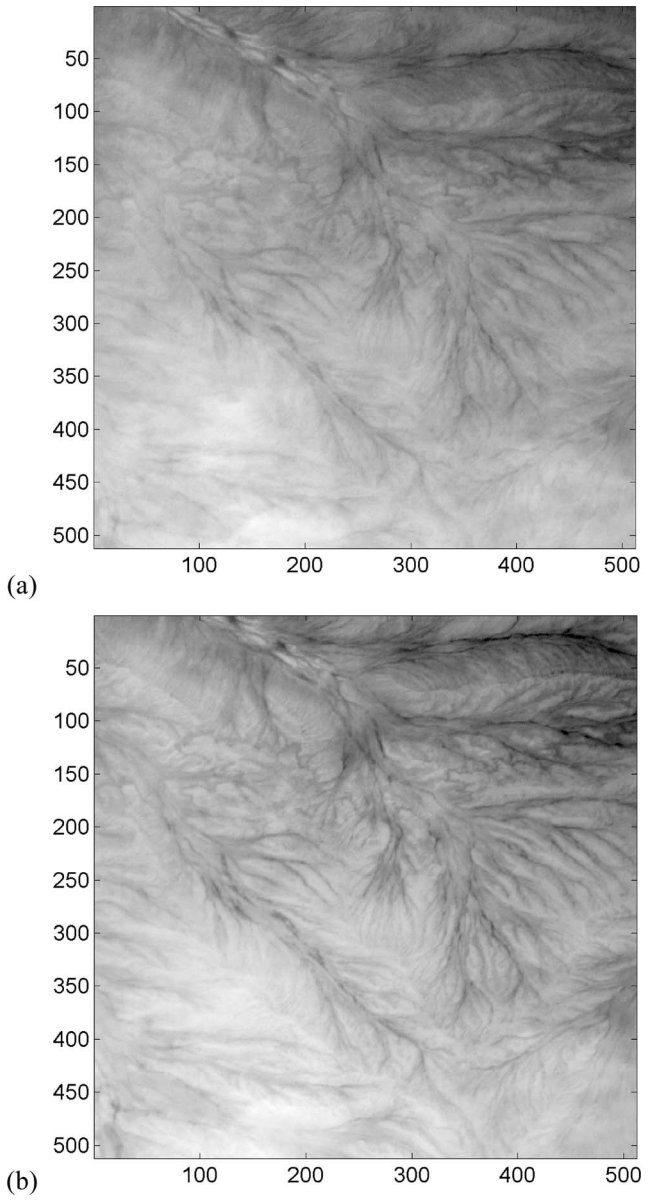

Fig. 1. MTI satellite images of the desert near Shiprock, NM. (a) 0.485-μm band. (b) 0.560-μm band.

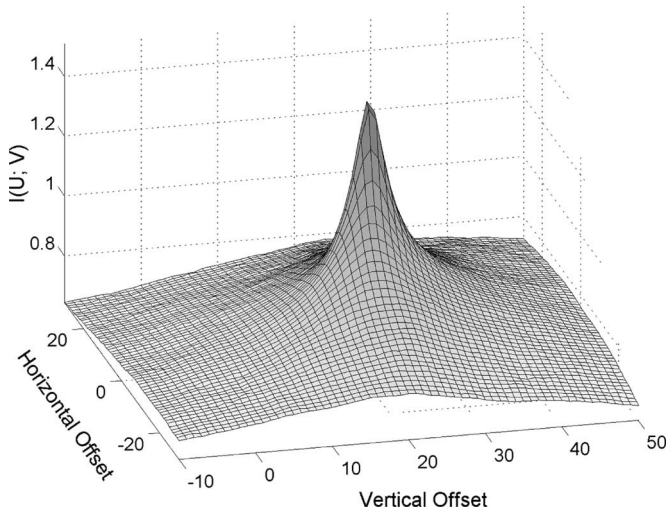

Fig. 2. Mutual-information curve for the two images in Fig. 1 that have been translated with respect to each other. For this smooth surface, we use the histogram method with 50 000 samples for generating each grid point.

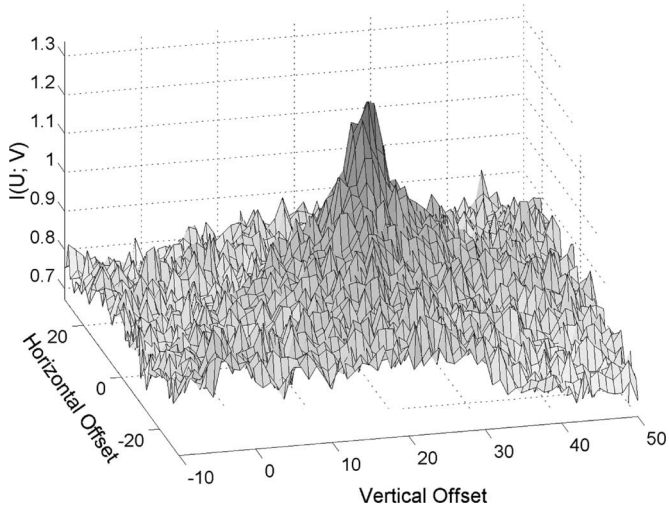

Fig. 3. Repeat of Fig. 2 with 200 image samples per grid point.

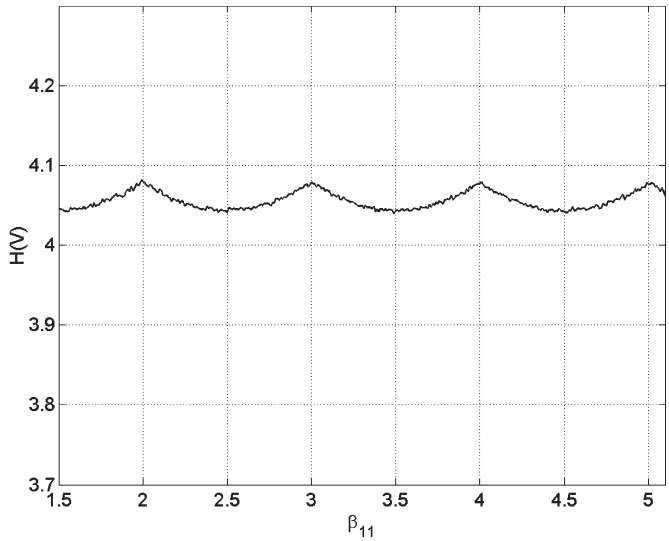

Fig. 4. Image entropy estimation showing false peaks at integer values, for interpolation that is centered at integer pixels. Image entropy should be constant. We get false peaks at integer translations.

To alleviate such interpolation artifacts, there are two basic approaches [3]. In the first approach, we can use a partial volume interpolation method [3] as demonstrated by Chen *et al.* in [4]. In [4], Chen *et al.* introduced a new algorithm called generalized partial volume estimation to alleviate this problem. A second approach that is somewhat more straightforward is to simply interpolate over irrational spacing [3] (also see [5] for a more detailed discussion). We demonstrate this approach in Figs. 4 and 5. We note that the interpolation artifacts have been effectively removed in Fig. 5. This approach results in well-defined mutual-information curves (see Figs. 6 and 7). In what follows, we used bicubic spline interpolation that was found to be two to three times more accurate than the bilinear interpolation that required half the time to execute.

Despite the recent introduction of mutual-information methods, most remote-sensing registration methods are still based on the use of correlation methods. A basic assumption used in

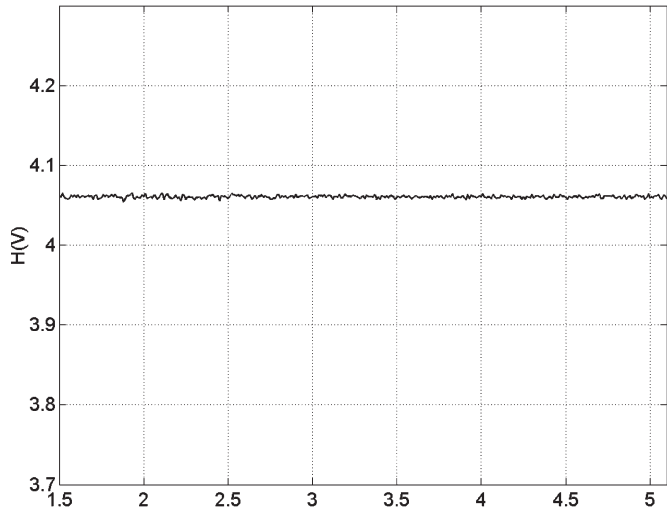

Fig. 5. Repeat of Fig. 1 showing how entropy estimation can be corrected after resampling at pixel spacing of $\sqrt{1.2}$.

cross-correlation methods is that the image (or feature) to be aligned obeys a "signal plus noise" model, where the signal is translated between the images to be aligned. Using normalized cross correlation (NCC), we can also handle the case where the signals are the same up to a multiplicative constant. Significant cases exist, however, where images do not satisfy these assumptions. In remote sensing, these assumptions are commonly violated when the imagery to be registered is collected under different lighting conditions or at different wavelengths. As an example, a phenomenon known as contrast reversal in multispectral remote sensing is due to the differences in the reflectance of various materials as a function of wavelength. For examples that demonstrate the inadequacy of NCC, we refer to [4, Figs. 3–5]. In [4], Chen *et al.* used Landsat images to demonstrate that both the NCC and mean square difference are inadequate similarity measures for image registration, while mutual information proves to be appropriate. We have found that same results hold for Multispectral Thermal Imager (MTI) images (described next).

All of the imagery used in this paper was from the Department of Energy's MTI satellite. The MTI sensor is a push-broom scanner consisting of 15 spectral bands (Fig. 8) from visible to long-wave infrared arranged in a stack of linear arrays, one for each band. Each linear array has 2496 pixels at 5-m nadir ground sample distance (GSD) in the visible to near infrared (VNIR) bands and 624 pixels at 20-m GSD in the remaining infrared bands. As the satellite sweeps over terrain in its sun-synchronous orbit at approximately 600-km altitude, 15 2-D images having a 12-km nominal swath width and an arbitrary height are formed by repetitively scanning each 1-D array. Since the infrared GSD is larger than the VNIR GSD, for registration, the infrared images are upsampled by a factor of four. Overall, for the acquisition geometry that we have described, registration is a simple translation due to timing offsets.

Several methods have been proposed to improve the performance of cross-correlation methods. Most involve preprocessing the image to emphasize certain features such as the magnitude of the image gradient [6], [7], point matching [8], [9], or object boundaries [10], [11], as well as other methods [1], [2], [12].

The use of mutual information as a similarity measure for image registration was proposed independently in 1995 by Viola [13] and Collignon *et al.* [14]. Since its introduction, this method has become widely used in the medical imaging community [2], [15]–[21]. Only recently has it been studied for remote-sensing applications. Johnson *et al.* [22] presented some first results in the comparison of correlation and mutual information for remote sensing. They found that mutual information produced consistently sharper peaks at the correct registration values than cross correlation, which is important for accurate registration. They also found that mutual information was more robust in Gaussian noise when working at low signal to noise ratios. Mutual information is also robust against nonlinear intensity relationships between images, such as contrast reversals.

The remainder of this paper is organized into four sections. In Section II, we present an overview of the image registration using mutual information. In Section III, a model for the mutual-information surface is developed. Section IV focuses on applying the model to a specific optimization technique, a gradient-based approach, as an example of how the model is used to select key parameters of the optimization. The results are given in Section V, and some concluding remarks are given in the final section.

II. Overview of Image Registration Using Mutual Information

We model pixel intensity as a random variable (RV) U, where each pixel intensity represents a sample, and the entire image is viewed as a random sample $\mathbf{U}$. For simplicity, we will designate the reference image by its random variable U and designate a second image as V.

To register the floating image to the reference, each pixel in the random sample $\mathbf{V}$ is spatially transformed by an operator T, resulting in a new random sample $\mathbf{V}'$. We register the image associated with U with the image associated with V by determining the spatial transformation that maximizes the mutual information over T, i.e.,

$$\hat{T} = \underset{T}{\operatorname{argmax}}\, I(U; V'). \tag{1}$$

A. Efficient Parameter Search

For higher order warping functions, exhaustive search methods are often impractical. Optimization algorithms, such as the gradient ascent algorithm first proposed by Viola [13] can be more efficient. This approach is based on the method of steepest descent used in classical adaptive signal processing [23], [24] where a cost function, in this case $I(U; V')$, is maximized (or minimized) by applying incremental adjustments to the parameters in the direction which increases (or decreases) the cost function, i.e.,

$$T^{(k+1)} = T^{(k)} + \eta \frac{dI}{dT} \tag{2}$$

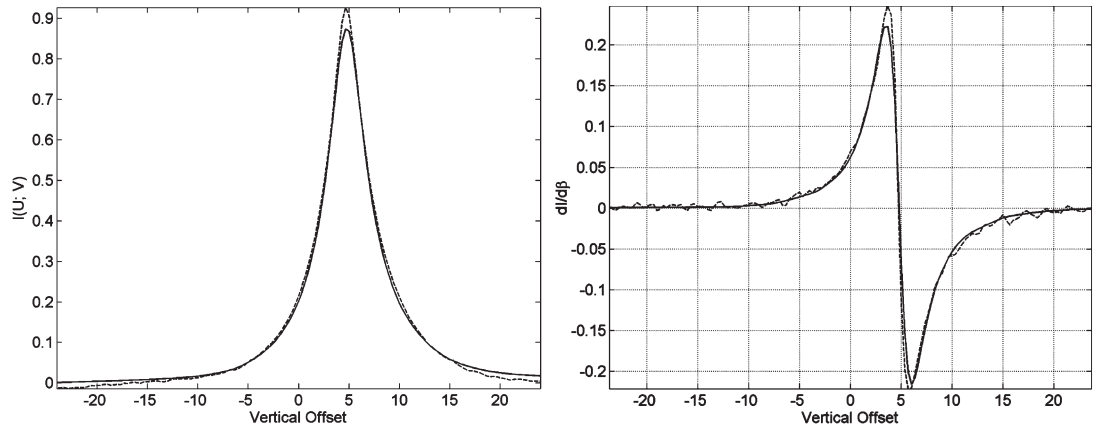

Fig. 6. Mutual-information model fits for the images in Fig. 8. The models are plotted using solid lines, and the actual information curve is shown using dashed lines. The discrete derivative is shown in the right image.

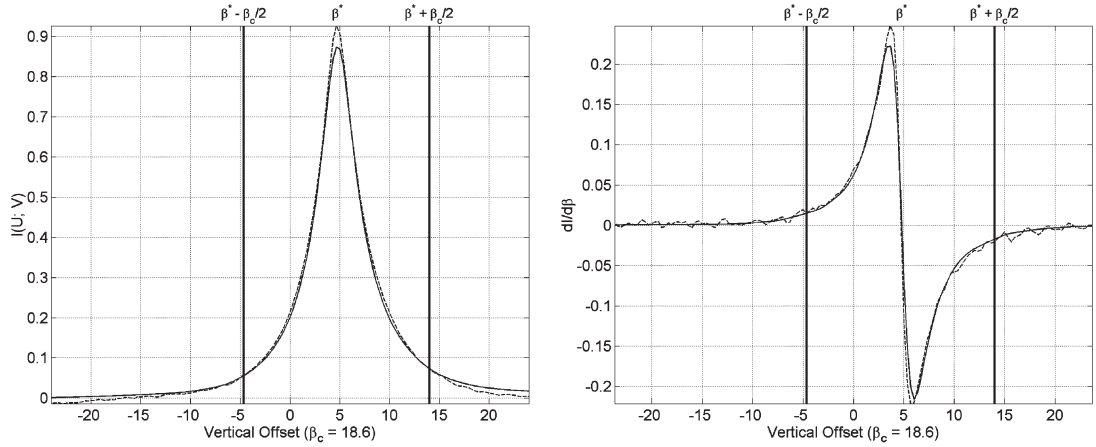

Fig. 7. Mutual-information curves and its derivative shown relative to the correlation length β_c centered about the optimal parameter β^* for the images in Fig. 4. Note that most of the peak is captured within the correlation length.

Band	Wavelength Range (μm)
A	0.45-0.52
B	0.52-0.60
C	0.62-0.68
D	0.76-0.86
E	0.86-0.89
F	0.91-0.97
G	0.99-1.04
H1 & H2	1.36-1.39
I	1.54-1.75
O	2.08-2.37
J	3.49-4.1
K	4.85-5.05
L	8.01-8.39
M	8.42-8.83
N	10.15-10.7

Fig. 8. MTI sensor wavelengths.

where k is the iteration number, η is the step-size parameter, and dI/dT is the image gradient.

There are a number of ways to accelerate the convergence of this algorithm, such as the second-order algorithms discussed in [23]–[25]. These will not be discussed in this paper. It should be noted, however, that the models presented in this paper could be applied to other optimization approaches.

An important feature of the optimization algorithms such as (2), which employ iterative approximation, is that they generally work well when the gradient is estimated stochastically using a relatively small number of samples, improving the performance even further. In what follows, instead of using the histogram method, we will use multidimensional Gaussians to approximate the underlying pdfs. For details, we refer the reader to Cacoullos [26] for the multidimensional case and to Parzen [27] for the 1-D case.

III. Mutual-Information Models for Remote-Sensing Applications

In this section, we begin with a quadratic approximation to the mutual-information surface and show how this approximation is to address a number of convergence issues. We then present a more complete model of the mutual-information surface that is useful for predicting the quadratic approximation as well as several other parameters related to the optimization process. Finally, we relate the model to the frequency-domain

characteristics of the image showing how these affect the shape of the mutual-information surface.

A. Quadratic Approximation to the Mutual-Information Surface

Let the transformation T consist of several parameters β, and let the mutual-information surface be approximated by the following Taylor series expansion about the peak (around β^*):

$$I(\beta) \approx I(\beta^*) + \frac{1}{2}(\beta - \beta^*)^T \mathbf{H} (\beta - \beta^*) \tag{3}$$

$$\nabla I \approx \mathbf{H}(\beta - \beta^*) \tag{4}$$

where $\mathbf{H}$ is the Hessian matrix

$$H_{\mathrm{ij}} \equiv \left. \frac{\partial^2 I}{\partial \beta_{\mathrm{ij}}^2} \right|_{\beta=\beta^*}. \tag{5}$$

Let λ_i be the eigenvalues of $\mathbf{H}$. Although $\mathbf{H}$ is generally not diagonal, we can approximate it as such and treat the β_{ij} as independent without loss of generality.

B. Stability and Convergence

For a single parameter β, the stochastic gradient ascent algorithm in (2) becomes

$$\beta^{(k+1)} = \beta^{(k)} + \eta \frac{\partial I}{\partial \beta^{(k)}}. \tag{6}$$

An expression for β at an arbitrary iteration k can be written as

$$\beta^{(k)} \approx (1 + \eta\lambda)^k \beta^{(0)} - \eta\lambda\beta^* \sum_{i=0}^{k-1} (1 + \eta\lambda)^i. \tag{7}$$

Note that λ is negative at the maximum. For values of $0 < |1 + \eta\lambda| < 1$ and large values of k, the summation is approximated by

$$\sum_{i=0}^{k-1} (1 + \eta\lambda)^i \to \sum_{i=0}^{\infty} (1 + \eta\lambda)^i = \frac{-1}{\eta\lambda}. \tag{8}$$

Therefore, we have an expression for the kth value of the parameter

$$\beta^{(k)} \approx (1 + \eta\lambda)^k \beta^{(0)} + \beta^*. \tag{9}$$

Note that the restriction $0 < |1 + \eta\lambda| < 1$ was placed on this derivation. It should be clear that if this restriction is not met, the summation grows without bound, and the process is unstable. Conversely, the process is stable and approaches the solution in (9) if the condition $0 < \eta < 2/|\lambda|$ is met. For multiple parameters β_i, with corresponding Hessian eigenvalues of λ_i, it is easy to see that the stability criterion requires

$$0 < \eta < \frac{2}{\max_i |\lambda_i|}. \tag{10}$$

As $k \to \infty$ in (7), we can see that the parameter converges correctly to the peak mutual information

$$\beta^{(k)} \approx \beta^*. \tag{11}$$

When using noisy gradient estimates, for constant η, we can only guarantee convergence in the mean value

$$E\left[\beta^{(k)}\right] \approx \beta^* \tag{12}$$

assuming that the estimation noise is asymptotically zero mean.

C. Model of the Mutual-Information Surface

In this section, we consider the simpler case where the image pair is modeled as a pair of jointly Gaussian random variables. Let $\boldsymbol{Z}$ be a p-dimensional zero-mean multivariate random vector with pdf

$$f_Z(\boldsymbol{z}) = (2\pi)^{-p/2} |\Sigma_Z|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{z}^T \Sigma_Z^{-1} \boldsymbol{z}\right) \tag{13}$$

where Σ_Z is its covariance matrix. Hence, we can represent the entropy as

$$H(Z) = p \log\sqrt{2\pi e} + \frac{1}{2}\log\left(|\Sigma_Z|\right). \tag{14}$$

Let $\boldsymbol{Z}$ be comprised of two random variables U and V that represent the pixel intensities of two images by the same names. From (14), we have

$$H(U, V) = \log(2\pi e) + \log|\Sigma_{UV}|^{1/2} \tag{15}$$

where

$$\begin{aligned} \log|\Sigma_{UV}|^{1/2} &= \log\sqrt{\sigma_U^2 \sigma_U^2 - \sigma_{UV}^2} \\ &= \log\sigma_U + \log\sigma_V + \log\sqrt{1 - \rho_{UV}^2} \end{aligned} \tag{16}$$

and $\rho_{UV} = \sigma_{UV}/\sigma_U\sigma_V$ is the correlation coefficient of U and V. Similarly, we can show that

$$\begin{aligned} H(U) &= \frac{1}{2}\log 2\pi e + \log\sigma_U \\ H(V) &= \frac{1}{2}\log 2\pi e + \log\sigma_V \end{aligned} \tag{17}$$

and finally

$$\begin{aligned} I(U; V) &= H(U) + H(V) - H(U, V) \\ &= -\log\sqrt{1 - \rho_{UV}^2}. \end{aligned} \tag{18}$$

Considering a single parameter as before, the mutual information varies with respect to the parameter β. For the moment, let us assume linearity and shift invariance. Then, (18) can be rewritten as

$$I(\beta) = -\log\sqrt{1 - \rho_{UV}^2(\beta)} \tag{19}$$

where we once again assume that U and V are implicit in the mutual-information function. The correlation coefficient can be defined in terms of the autocorrelation functions $R_{UU}(\beta)$ and $R_{VV}(\beta)$, and cross-correlation function $R_{UV}(\beta)$ giving

$$I(\beta) = -\log\sqrt{1 - \frac{R_{UV}(\beta)^2}{R_{UU}(0)R_{VV}(0)}} \tag{20}$$

in terms of the autocorrelation and cross-correlation functions.

D. Frequency-Domain Model

Let U be a wide sense stationary (WSS) stochastic process. For translation-based registration, the cross-power spectral density (cpsd) function is given by the Fourier transform relationship

$$S_{UV}(f) = \mathsf{F}\left(R_{UV}(\beta)\right). \tag{21}$$

Also, note that the autocorrelation terms evaluated at a shift of zero are equal to the average power

$$R_{UU}(0) = P_{UU} \qquad R_{VV}(0) = P_{VV}. \tag{22}$$

We can now represent the mutual-information approximation in terms of the cpsd according to

$$I(\beta) \approx -\log\sqrt{1 - \frac{1}{P_{UU}P_{VV}}\left(F^{-1}\left(S_{UV}(f)\right)\right)^2}. \tag{23}$$

E. Effective Bandwidth Model

The model described in (23) is generally applicable for any WSS stochastic process. Next, we consider a common image model that can be used to effectively model the cpsd. For a Gaussian Markov process model of the input image, we have

$$S_{UV}(f) = \frac{(P_{UV}/2B_{\text{eff}})}{1 + (\pi f/2B_{\text{eff}})^2} \tag{24}$$

where P_{UV} is the cross-power of images U and V, and the effective bandwidth B_{eff} is equal to

$$B_{\text{eff}} = \frac{1}{2}\frac{\int_{-\infty}^{\infty} S_{UV}(f)df}{\max\left[S_{UV}(f)\right]}. \tag{25}$$

A reasonable model of the spectral characteristics of the cpsd is sufficient for computing the effective bandwidth in (25). Perhaps, more importantly, this model suggests a direct relationship between the image spectral characteristics and the mutual-information surface. We will explore this next.

F. Relationship Between Image Spectra and Mutual-Information Surface

We define the width of the cross-correlation function in the spatial domain to be equal to the correlation length

$$\beta_{\text{c}} = \frac{\int_{-\infty}^{\infty} R_{UV}(\beta)d\beta}{R_{UV}(0)}. \tag{26}$$

The correlation length β_{c} is related to the effective bandwidth B_{eff} in the frequency domain according to

$$\beta_{\text{c}} = \frac{1}{2B_{\text{eff}}}. \tag{27}$$

Therefore, the narrower the bandwidth of $S_{UV}(f)$, the longer the cross-correlation function $R_{UV}(\beta)$ and the wider the mutual-information surface in (20), and vice versa.

We know from the properties of the cpsd that

$$|S_{UV}(f)|^2 \leq S_{UU}(f)S_{VV}(f). \tag{28}$$

We see that the cpsd spectrum, and therefore its effective bandwidth, is bounded by the spectra of the individual images. The effective bandwidth in turn determines the width of the mutual information peak. We can conclude that, in general, images with large bandwidths (high-frequency features) have narrower mutual information peaks, and the images with small bandwidths (low-frequency features) have broader mutual information peaks.

IV. Using Mutual-Information Models for Optimization

We now illustrate how the model in (23) is used to provide better control over the optimization process. Several optimization methods may be used for finding the peak of the mutual-information function. The examples here are particular to the stochastic gradient approach in (2), but clearly, the model has applicability to other methods.

A. Stability in the Linear Region

In (10), we saw that if λ is known, stability was assured within the linear region by choosing an appropriate step-size parameter η. Let us define

$$D = dI/d\beta. \tag{29}$$

Referring to Fig. 7 (also discussed in Section V), we note that the width of the linear region can be defined as the range of β between the minimum and maximum values of D according to

$$L = \beta_{D_{\min}} - \beta_{D_{\max}}. \tag{30}$$

From this, we can approximate the slope as

$$\lambda \approx \frac{D_{\min} - D_{\max}}{L} \approx -2\frac{D_{\max}}{L}. \tag{31}$$

Therefore, using the mutual-information model, we are able to estimate λ and use this to calculate a value for η that will insure stability.

B. Stability in the Nonlinear Region

We found in the experiments performed that the largest magnitude of the derivative D always occurred at the boundary between the linear and nonlinear regions. Using this knowledge, we are able to insure stability in the nonlinear region as well. Considering a single parameter for simplicity, from (6), we see that the largest possible iteration step is bounded by $\eta D_{\max}$. Using (10) and (31), we see that

$$\eta \frac{dI}{d\beta} \leq \pm L \tag{32}$$

insuring that the largest possible step size is never larger than the width of the linear region. This guarantees that the process will eventually step into the linear region where it will be stable.

C. Rate of Convergence

The transient equation in (9) allows us to calculate the rate of convergence in the linear region. Since we have a model for the mutual-information surface, we can also approximate the rate of convergence in the nonlinear region using (6). What we find is that, for a fixed value of η, convergence is very slow, away from the peak where $dI/d\beta$ is very small.

The model shows us that a much faster approach is possible. From (32), we see that $\eta dI/d\beta \leq \pm L$, where the sign is determined by the slope of the derivative. Setting this to an equality gives us a fast convergence that is stable when the process enters the linear region. However, this approach will also have an instantaneous error of $\pm L/2$. We can reduce this further as we will see in the next section.

D. Annealing

It is easy to detect when the process has converged to the linear region by noting when it has oscillated about a single value for some number of iterations. Once in the linear region, the process can be forced to converge toward a single value by reducing the step-size parameter through a process known as annealing. A common approach to annealing is to update the step-size parameter of each of the iteration according to

$$\eta_k = \frac{1}{k}. \tag{33}$$

It can be shown that this will converge to the mean of the parameter estimate [28] which is assumed to be the optimal value. The experimental evidence indicates that this is true. If the parameter is initially in the nonlinear region, we want to delay annealing until it is in the linear region. Therefore, we modify (33) to be

$$\eta_k = \begin{cases} \eta_0, & k \leq k_1 \\ \frac{1}{k-k_1}, & k > k_1 \end{cases} \tag{34}$$

where η is initially set to η_0, and k_1 is the last sample prior to where annealing begins. The process is stopped when $E[(\beta - \beta^*)^2]$ reaches a sufficiently small value for the particular application.

It has been shown that if certain methods are followed for interpolating the data to achieve subpixel registration [20], [29], that very accurate results can be achieved. Other annealing methods are similarly successful. For example, η can also be reduced incrementally in a stepwise fashion [29].

E. Attraction Range

The range over which the gradient-type search algorithms will converge can be predicted with the mutual-information model. The capture range (also known as the attraction range) of the optimization process is the range over each parameter for which the gradient is sufficient for optimization to be successful. It was noted in Fig. 7 that the correlation length in (26), calculated from the model, captures the majority of the mutual information peak. The value $\beta_c/2$ provides an empirical estimate of the maximum distance from the optimal value for which sufficient gradient exists. Outside of this range, other search methods are required.

F. Computational Complexity

We provide a brief discussion of the computational complexity of the proposed approach. Overall, execution time is of the order of $N \cdot M^2$, where N denotes the number of iterations and M denotes the number of random samples taken from each image (typically, $M \approx 50$). Here, M^2 refers to the number of Kernel function evaluations that are needed for estimating dI/dT.

V. Results and Discussion

The primary goal of this paper has been to establish robust methods for image registration, where the algorithm will be expected to converge provided that the initial parameters are within some measurable distance from the correct registration parameters. In the experimental results that follow, we note that the dataset has not been geocoded. To validate convergence from a large number of possible distances, we manually register the imagery using 50 carefully selected control points then simulate an offset. This simulation approach is simple and general enough to be duplicated on any given dataset. As we describe below, we validate our claim over 100 random trials per experiment. To further validate the results, we resort to a consistency check (see also [4]). Then, to estimate the accuracy of the registration, we estimate the variance of the estimated registration parameters (a consistency measure).

In Sections V-A and B, we present the results for what we termed as the "linear" and "nonlinear" regions, and the correlation length. As we demonstrate below, we can use the correlation length as a distance measure within which we expect the algorithm to converge. Then, in Section V-C, we present the consistency results, similar to [4], but applied to our new algorithm. We demonstrate the robustness of our approach in

TABLE I
WIDTH OF THE LINEAR REGION (VERTICAL TRANSLATION, IN PIXELS) FOR SEVERAL IMAGE TYPES AND SEVERAL SPECTRAL BANDS. IN EACH CASE, THE INCLUDED BAND WAS REGISTERED WITH BAND A

Band (Wavelength)	A .485	B .560	C .650	D .810	E .875	F .975	G 1.02	I 1.65	O 2.23	J 3.80	K 4.95	L 8.20	M 8.63	N 10.4
Mountains	0	1.3	2.2	2.3	4.1	4.0	4.3	4.8	4.8	5.1	5.3	5.8	6.0	5.8
Suburban	0	2.4	1.6	3.1	6.3	6.1	5.8	5.6	5.8	5.7	6.7	7.0	7.3	7.3
Desert	0	2.2	2.6	3.0	5.2	5.0	5.8	6.8	27	7.0	8.3	22	29	29

TABLE II
CORRELATION LENGTH OF VERTICAL TRANSLATION PARAMETER CALCULATED USING FREQUENCY-DOMAIN TECHNIQUES

Band (Wavelength)	A .485	B .560	C .650	D .810	E .875	F .975	G 1.02	I 1.65	O 2.23	J 3.80	K 4.95	L 8.20	M 8.63	N 10.4
Mountains	20	21	21	21	23	22	23	24	23	24	24	27	26	27
Suburban	18	20	18	19	21	22	20	20	23	46	32	30	28	32
Desert	56	54	47	48	48	64	50	43	181	44	164	166	177	178

Section V-D. In Section V-D, we show that the new algorithm always converges when initialized within this attraction region.

A. Linear and Nonlinear Regions

We indicated earlier that the mutual-information surface tends to have a "linear" region, where the surface is approximated by a quadratic Taylor series expansion. Within this region, the gradient is approximately linear. In the right plot of Fig. 6, the linear region is the segment lying between the positive and negative peaks of the derivative. This plot was repeated for three scenes, each with 14 bands. The results are tabulated in Table I. Outside of the linear region lies the "nonlinear" region. This is where the quadratic approximation does not hold.

Note that for some of the image pairs, the linear region is quite small—on the order of only one to three pixels. Values on this order were commonly observed in the remote-sensing imagery that was used for these experiments.

It is also important to describe how the values listed in Table I relate to the accuracy of the registration process. Basically, the smaller the width of the linear region, the sharper is the peak in the mutual-information surface. With sharper peaks, we also get more accurate registration performance. Thus, in general (but not always), it is easier to accurately register the bands in the mountains and suburban scenes than it is to register all the bands in the desert scene. Overall, convergence accuracy is scene dependent. In general, we might expect that bands that are spectrally far apart, such as A and N, will be the most difficult to register. This is further exaggerated by the infrared pixels having four times lower resolution than the VNIR pixels. We see this as an increase in the width of the linear region between widely separated bands. This is not always true; however, many scenes can have large dissimilarities between bands that are not as widely separated. For the desert scene, we note that the pair of A–O is significantly more difficult to register than the A–J and A–K bands. Beyond the idea of registering a pair of images at a time, we have also considered the problem of concurrently registering all images together by adaptively selecting different pairs of images [29], [30]. It is also important to note that it is the correlation length that determines when registration is possible (see Table II).

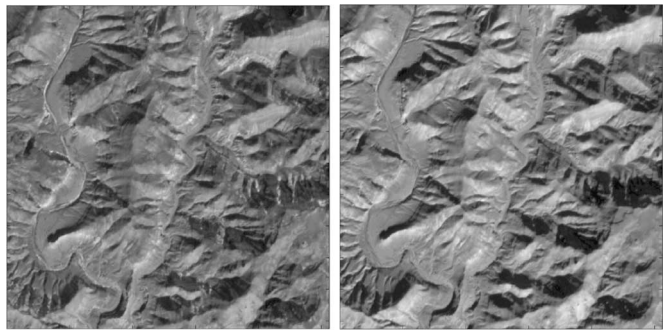

Fig. 9. Band A (0.485 μm) and Band D (0.810 μm) collected by the MTI satellite over the mountains in the region of Utah.

Fig. 10. Band A (0.485 μm) and Band D (0.810 μm) collected by the MTI satellite over Albuquerque, NM.

B. Correlation-Length Estimation Using the Effective Bandwidth

In this section, we present the results from estimating the correlation length. We then use the estimated correlation length to indicate how it can be used to characterize the attraction region.

To demonstrate the variation in the estimation of the correlation length, we use the mountainous images of Fig. 9 and the suburban image of Fig. 10, and a desert image region of Fig. 11. The mountainous and suburban images have a relatively high frequency content, which can be seen as rapid contrast variations. For each of these (Figs. 9 and 10), the difference in frequency content between the visible and infrared bands is not dramatic. We therefore only need to concentrate

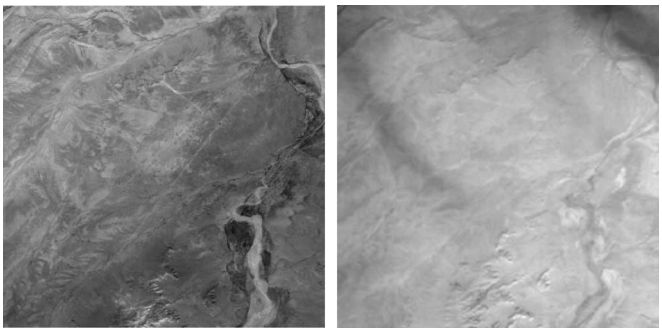

Fig. 11. Desert region collected by the MTI satellite near Shiprock, NM. The left image is Band A (0.485 μm), and the right image is Band N (10.4 μm).

on the visible bands (left images). The desert image of Fig. 11 has lower frequency content than the other two. Furthermore, the infrared image (right) appears to be significantly smoother, suggesting a lower frequency content than the visible image (left). Using (25) and (27), we calculated the correlation length for each image band.

The estimated correlation-length results using the effective bandwidth model are tabulated in Table II. As predicted, the visible band (Band A) of the desert image has a correlation length that is larger than the other two images. Likewise, the correlation length of the infrared band (Band N) has a larger correlation length of the visible band.

As we noted earlier, we can use the correlation length to provide an estimate of the attraction region, as shown in Fig. 7 (for the mountainous images). In Fig. 7, we can see that the correlation length captures most of the information peak. Beyond the correlation length, the gradient magnitude tends to be small and noisy.

C. Consistency Results

Using the methods presented, several spectral bands from the three different scenes were registered—the suburban scene in Fig. 10, the mountain scene in Fig. 9, and the desert scene in Fig. 11. For each scene, Band D (0.810 μm) and Band G (1.015 μm) were each registered horizontally and vertically to Band A (0.485 μm) using the method presented in (2) and the Gaussian kernel functions. For the images in Figs. 9 and 10, we use just 50 randomly selected pixels from each image (see Section II-A), while 400 pixels were needed for the desert images in Fig. 11. The initial step size was initially set to $L/2$, which is half of the maximum allowed by the model. After the algorithm settled to within a range of values less than L for 25 time steps, it was assumed to be in the linear region. At this time, the step-size parameter was annealed.

Two methods were used to validate the results. In each case, the images were segmented into five horizontal slices of equal size so that five results could be compared for each band pair. First, a manual registration was performed using five to ten visually selected control points. For the second validation, we noted that if Band D were registered to Band A, Band G to Band D, and Band A to Band G, the net result should be zero (see also [4]). Fig. 12 illustrates the results for the vertical component of each registration. In each case, the result of this cyclic validation was very close to zero. To show that this

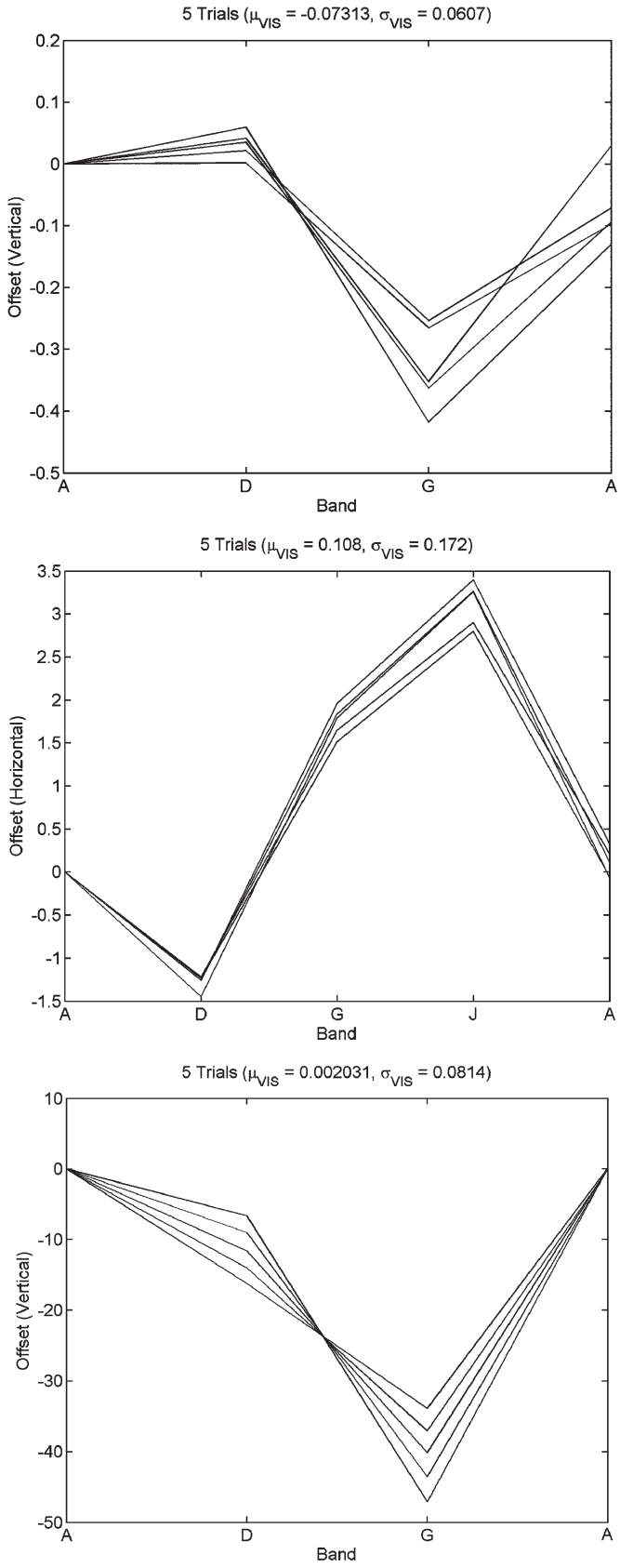

Fig. 12. Consistency test results (vertical shift) for the five slices of the suburban (top), mountain (middle), and desert (bottom) scenes. Note that the desert scene has significant differences between the intermediate results for each slice. This was because that scene was collected in a manner that resulted in significant stretching between the bands. Regardless, the cyclic result was close to zero as expected.

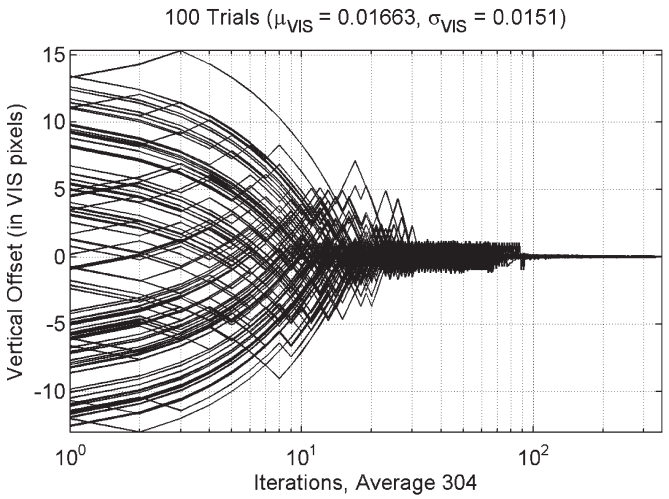

Fig. 13. Convergence of Band A (0.485 μm) and Band D (0.810 μm) of the mountain image, beginning with a misregistration of the 12 pixels in a random direction (the total horizontal and vertical offset is 12 pixels, but only the vertical is shown here).

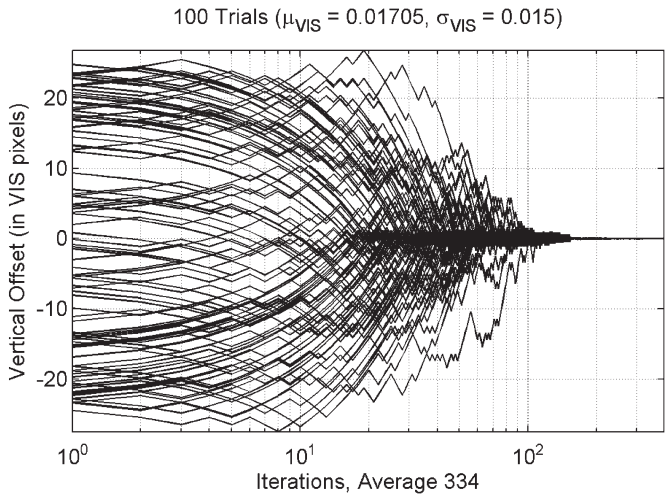

Fig. 14. Convergence of Band A (0.485 μm) and Band D (0.810 μm) of the mountain image, beginning with a misregistration of the 24 pixels in a random direction (only the vertical is shown here).

approach could be extended to more bands, an additional band (Band J, 3.795 μm) was inserted into the cyclic validation for the mountain scene.

We expected the manual registration to be accurate with 1 or 2 pixels (5 to 10 m) for the suburban and mountain scenes and 4 to 8 pixels for the desert scene (due to its lack of easily identifiable control point features).

For each of the suburban and mountain images, the initial offset for each band was within the correlation length predicted by the model, and as such, the algorithm converged without additional effort. At the time of the dessert collect, the sensor was in an operational mode that resulted in a much larger than usual offset between the spectral bands. The initial offsets were larger than the correlation length, and as expected, the algorithm did not converge without additional information. After shifting the imagery by a gross approximation of the offset, each converged normally. We further elaborate on the convergence properties of our approach in Section V-D.

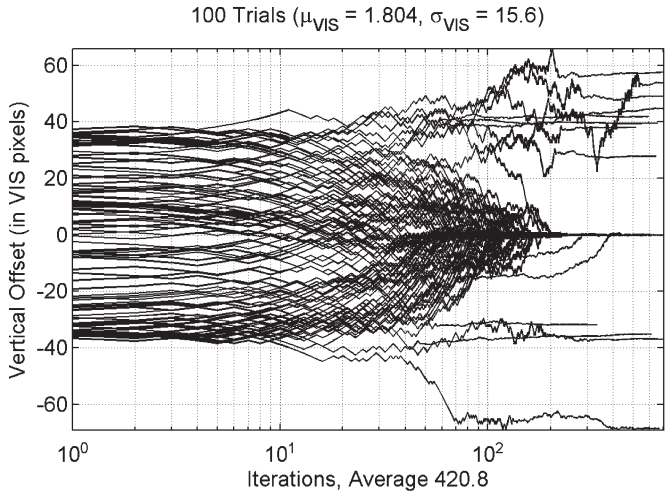

Fig. 15. Convergence of Band A (0.485 μm) and Band D (0.810 μm) of the mountain image, beginning with a misregistration of the 36 pixels in a random direction (only the vertical is shown here).

D. Convergence Results

In Table II, we indicated that the correlation length gives us an approximation of the range over which we expect the registration process to converge. Outside of this range, the gradient of the optimization surface becomes small, and we expect to see an increasing number of failures. To further illustrate this point, we plot 100 random trials of the convergence process for various starting offsets.

We begin with a known registration of two images by carefully registering the two images manually, using 50 control points. We repeat control point selection twice until the average difference between the selected control points is only 0.017 pixels and use the average of the two sets as ground truth. We then shift one of the images by a set number of pixels from that position in a random direction and run the registration algorithm.

Figs. 13–15 illustrate the results for the 100 trials, beginning at offsets of 12, 24, and 36 pixels, respectively. For these plots, we used Bands A and D of the mountain image. In Table II, we indicated that the correlation length for these two bands was 21 pixels. Fig. 13 shows that for an initial offset of 12 pixels, the algorithm converges every time. Only the vertical component is shown here. The horizontal results are similar. Note that the starting offset is constant, but the direction is random. Therefore, sometimes the vertical component is much less than 12 pixels even though the total starting offset is always 12. We see that all of the trials converge to a mean value of 0.01663 pixels with respect to the preregistration and a standard deviation of 0.0151 pixels.

Fig. 14 illustrates similar results for a 24-pixel starting offset—slightly more than the correlation length. Note that, even though the algorithm converges every time, we can see that the process was close to diverging in several of the trials. Also, note that, although the starting offset has been increased

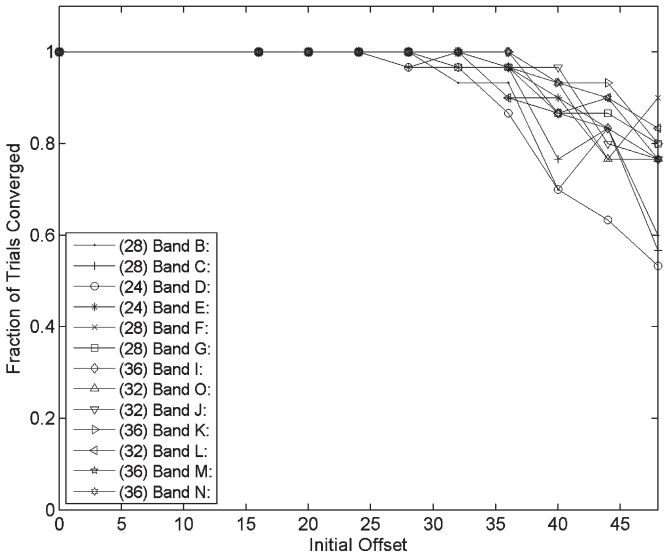

Fig. 16. Success rate for the convergence of Band A with 13 other bands of the mountain image, beginning with the misregistrations of 0 to 48 pixels.

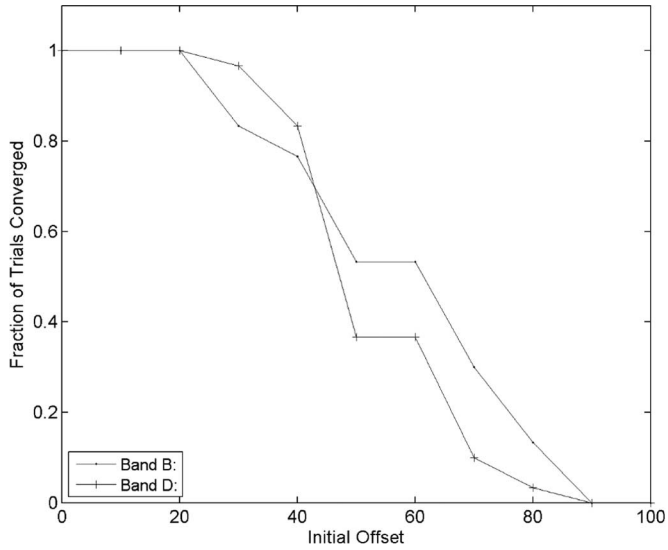

Fig. 17. Success rate for the convergence of Band A with Bands B and D of the mountain image, beginning with the misregistrations of 0 to 100 pixels.

significantly, the mean of the solution (0.01705) is virtually identical to the previous experiment.

In Fig. 15, we see the results for an initial offset of 36 pixels—much larger than the correlation length. It is clear from this plot that the process fails in several of the trials. As expected, when the starting point is well inside the correlation length, a gradient-based algorithm converges easily. However, as the starting offset moves beyond this range, the gradient is insufficient for convergence.

This process was repeated for 13 image pairs. For this same scene, we registered Band A with Bands B, C, D, E, F, G, I, O, J, K, L, M, and N using initial offsets ranging from 0 to 48 pixels with 30 trials at each offset. The results are plotted in Fig. 16. The legend in Fig. 16 indicates the maximum offset at which the results are reliable. Comparing to Table II, it is clear from these plots that the registration algorithm converges reliably for starting offsets up to slightly beyond the correlation length and then becomes increasingly less reliable.

Highly detailed imagery, such as the suburban and mountain scenes, tends to produce strong mutual information peaks. In contrast, scenes with very little detail produce weaker, and generally flatter, peaks that are more difficult to use for registration.

As a counter example to the well-behaved results in Fig. 16, we examine the results for the desert image. This image has very little detail and is a very difficult registration candidate. Nevertheless, it supports the conclusion that in the region of the mutual information peak, the images can be registered reliably, and that the process becomes less reliable as the initial offset moves farther from the peak.

Fig. 17 illustrates this point for Bands B and D of the dessert image registered with Band A. We see that, because its mutual information peak is so weak, we can take advantage of the central area of the peak. However, we cannot take advantage of as much of the correlation length as we normally could with a strong detailed image. This is why the curves roll off sooner than would be expected when using the correlation lengths in Table II as a reference. Nonetheless, mutual information still does an excellent job of registering poorly detailed imagery as long as the starting offset does not approach the correlation length.

We conclude that modeling the mutual-information surface provides valuable insight into setting up registration algorithms. We have found that the models provide excellent guidance for detailed imagery, and although the peaks are somewhat degraded, we see that mutual information still works well for registering poorly detailed imagery.

VI. Conclusion

In this paper, we related the shape of the mutual-information surface to the frequency-domain characteristics of the imagery. We showed that for the steepest descent, we can predict and control stability and convergence of the optimization process as well as predict the capture range over which the process is expected to converge. By modeling the mutual-information surface, we were able to predict the behavior of the registration algorithm. For multispectral sensing from a single platform, the primary registration offsets are linear, making mutual information a powerful tool for this application.

Beyond what we discussed here, we also had the opportunity to register warped imagery to a first order. The desert image used in this paper was warped due to a pointing problem that occurred on that day. The satellite was rotating along two axes during the collect, resulting in combined stretch and skew of more than 12 pixels from top to bottom. We found that by dividing the image into eight segments (four high by two wide), we were able to build a first-order warping model. Likewise, Johnson *et al.* [22] showed that, for small rotations, mutual-information registration was quite effective. This is important because, in remote sensing, we generally have preliminary information on the offset between images so that we are not starting with large warps or rotations. Thus, based on these findings, we believe that the proposed mutual information approach should be applicable to registering images acquired by the different platforms having different heights, orbits, and angles.

ACKNOWLEDGMENT

The authors would like to thank Sandia National Laboratories and the U.S. Department of Energy for providing the imagery used in this paper. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the U.S. Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

REFERENCES

[1] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 1, pp. 325–376, 1992.
[2] B. Zitova and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.
[3] J. P. W. Pluim, J. B. A Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.
[4] H. Chen, P. K. Varshney, and M. K. Arora, "Performance of mutual information similarity measure for registration of multispectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2445–2454, Nov. 2003.
[5] J. Inglada, V. Muron, D. Pichard, and T. Feuvrier, "Analysis of artifacts in subpixel remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 254–264, Jan. 2007.
[6] A. Roche, X. Pennec, G. Malandain, and N. Ayache, "Rigid registration of 3-D ultrasound with MR images: A new approach combining intensity and gradient information," *IEEE Trans. Med. Imag.*, vol. 20, no. 10, pp. 1038–1049, Oct. 2001.
[7] Z. Jiang and W. Lu, "A new digital image registration algorithm based on the double spatial intensity gradients using pyramids," *Pattern Recognit. Lett.*, vol. 8, no. 5, pp. 335–340, Dec. 1988.
[8] J. Ton and A. K. Jain, "Registering landsat images by point matching," *IEEE Trans. Geosci. Remote Sens.*, vol. 27, no. 5, pp. 642–651, Sep. 1989.
[9] Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin, "An automatic image registration for applications in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2127–2137, Sep. 2005.
[10] J. F. Andrus, C. W. Campbell, and R. R. Jayroe, "Digital image registration method using boundary maps," *IEEE Trans. Comput.*, vol. C-24, no. 9, pp. 935–940, Sep. 1975.
[11] H. Li, B. S. Manjunath, and S. K. Mitra, "A contour based approach to multisensor image registration," *IEEE Trans. Image Process.*, vol. 4, no. 3, pp. 320–334, Mar. 1995.
[12] E. Sansosti, P. Berardino, M. Manunta, F. Serafino, and G. Fornaro, "Geometrical SAR image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2861–2870, Oct. 2006.
[13] P. A. Viola, "Alignment by maximization of mutual information," Ph.D. dissertation, MIT, Cambridge, MA, 1995.
[14] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multi-modality image registration based on information theory," in *Information Processing in Medical Imaging*, Y. Bizais, C. Barillot, and R. Di Paola, Eds. Norwell, MA: Kluwer, 1995, pp. 263–274.
[15] D. P. McGarry, T. R. Jackson, M. B. Plantec, N. F. Kassell, and J. H. Downs, III, "Registration of functional magnetic resonance imagery using mutual information," *Proc. SPIE*, vol. 3034, pp. 621–630, 1997.
[16] F. Maes, A. Collingnon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodal image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.
[17] C. E. Rodriguez-Carranza and M. H. Loew, "A weighted and deterministic entropy measure for image registration using mutual information," *Proc. SPIE*, vol. 3338, pp. 155–166, 1998.
[18] L. Xiang, L. Shuqian, and L. Kuncheng, "Registration of CT and MR images using maximization of mutual information," in *Proc. 1st Joint BMES/EMBS Conf.*, Atlanta, GA, Oct. 13–16, 1999, p. 1164.
[19] A. Roche, G. Malandain, N. Ayache, and S. Prima, "Towards a better comprehension of similarity measures used in medical image registration," in *Proc. MICCAI*, Cambridge, U.K., 1999, pp. 555–566.
[20] P. Thévenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Trans. Image Process.*, vol. 9, no. 12, pp. 2083–2099, Dec. 2000.
[21] A. Carrillo, J. L. Duerk, J. S. Lewin, and D. L. Wilson, "Semiautomatic 3-D image registration as applied to interventional MRI liver cancer treatment," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 175–185, Mar. 2000.
[22] K. Johnson, A. Cole-Rhodes, I. Zavorin, and J. LeMoigne, "Mutual information as a similarity measure for remote sensing image registration," *Proc. SPIE*, vol. 4383, pp. 51–61, 2001.
[23] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
[24] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.
[25] C. M. Bishop, *Neural Networks for Pattern Recognition*. London, U.K.: Oxford Univ. Press, 1995.
[26] T. Cacoullos, "Estimation of a multivariate density," *Ann. Inst. Stat. Math.*, vol. 18, no. 2, pp. 179–189, 1966.
[27] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.
[28] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. New York: Academic, 1999.
[29] J. P. Kern, "Spectral image registration using multivariate mutual information," Ph.D. dissertation, Univ. New Mexico, ProQuest Inf. & Learning, Albuquerque, NM, 2003.
[30] J. P. Kern, M. S. Pattichis, and S. D. Stearns, "Registration of image cubes using multivariate mutual information," in *Proc. 37th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2003, vol. 2, pp. 1645–1649.

Jeffrey P. Kern received the B.S. degree in biomedical engineering from Louisiana Tech University, Ruston, in 1981, the M.S. degree in electrical engineering from Louisiana State University, Baton Rouge, in 1982, and the Ph.D. degree from the University of New Mexico, Albuquerque, in 2003.

Since 1983, he has been working with Sandia National Laboratories, Albuquerque, NM, where he has held positions in project management, algorithm development, analog and digital hardware design, and software design. He is currently a Distinguished Member of the Technical Staff in the Remote Sensing and Exploitation Department. His current research interest is image processing for remote-sensing systems.

Marios S. Pattichis (SM'06) received the B.Sc. degree (high honors and special honors) in computer sciences in 1991, the B.A. degree (high honors) in mathematics in 1991, the Master of Science in Engineering in 1993, and the Ph.D. degree in computer engineering in 1998, all from the University of Texas, Austin.

Since 2000, he has been an Assistant Professor with the Department of Electrical and Computer Engineering, University of New Mexico (UNM), Albuquerque. Since 2005, he has been an Adjunct Assistant Professor in radiology with the Department of Radiology, UNM. He is an Associate Editor for *Pattern Recognition*. His research is in the areas of medical image and video processing, digital image and video models, radar image processing, single instruction multiple data, and reconfigurable computer architecture applications.

Dr. Pattichis is the General Chair of the 2008 IEEE Southwest Symposium on Image Analysis and Interpretation to be held in Santa Fe, NM. At UNM, he received the 2003 Electrical and Computer Engineering Distinguished Teaching Award and the 2006 School of Engineering Harrison Faculty Recognition Award.