Multiresolution Registration of Remote-Sensing Images Using Stochastic Gradient

Arlene Cole-Rhodes[1], Kisha Johnson[1], Jacqueline Le Moigne[2]

1. Dept of Electrical and Computer Engineering, Morgan State University, Baltimore, MD 21251
2. Applied Information Science Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771

ABSTRACT

In image registration, we determine the most accurate match between two images, which may have been taken at the same or different times by different or identical sensors. In the past, correlation and mutual information have been used as similarity measures for determining the best match for remote sensing images. Mutual information or relative entropy is a concept from information theory that measures the statistical dependence between two random variables, or equivalently it measures the amount of information that one variable contains about another. This concept has been successfully applied to automatically register remote sensing images based on the assumption that the mutual information of the image intensity pairs is maximized when the images are geometrically aligned. The transformation which maximizes a given similarity measure has been previously determined using exhaustive search, but this has been found to be inefficient and computationally expensive. In this paper we utilize a new simple, yet powerful technique based on stochastic gradient, for the maximization of both similarity measures with remote-sensing images, and we compare its performance to that of the exhaustive search. We initially consider images, which are misaligned by a rotation and/or translation only, and we compare the accuracy and efficiency of a registration scheme based on optimization for this data. In addition, the effect of wavelet pre-processing on the efficiency of a multi-resolution registration scheme is determined, using Daubechies wavelets. Finally we evaluate this optimization scheme for the registration of satellite images obtained at different times, and from different sensors. It is noted that once a correct optimization result is obtained at one of the coarser levels in the multi-resolution scheme, then the registration process is much faster in achieving subpixel accuracy, and is more robust when compared to a single level optimization. Mutual information was generally found to optimize in about one third the time required by correlation.

Keywords: image registration, mutual information, remote sensing imagery, optimization, wavelets.

1.INTRODUCTION

Within the context of satellite data geo-registration, this work considers the issue of feature-based, precision-correction and automatic image registration of satellite image data. In this context, image registration is defined as the process which determines the best match between two or more images acquired at the same or at different times by different or identical sensors. One set of data is taken as the *reference data* and all other data, called *input data (or sensed data),* is matched relative to the reference data. The general process of image registration includes three main 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 the data based on the correspondence computed from matched features. A large number of automatic image registration methods have been proposed and surveys can be found in[1-3].

This paper considers the search strategy and similarity metric to be used in step (2) of the registration process. In previous work, exhaustive search was used to compare two different similarity metrics, correlation and mutual information. The registration process was found to be slow, thus in this work we investigate the application of an optimization technique for registration based on either of the two metrics. Correlation measurement has been the most widely used similarity metric[4], although it is computationally expensive and noise sensitive when used on original gray level data. Use of a multi-resolution search strategy provides for large reductions in computing time and increases the robustness of the algorithms. An alternate similarity metric is mutual information, which measures the "resemblance" between two images. But while correlation measures similarity by computing global statistics such as mean and variance, mutual information measures redundancy between two images by looking at their intensity distributions. Prior work[12] has

shown that mutual information works well for remote sensing imagery and that the maximum peak provided by mutual information is sharper and better defined than the correlation maximum peak. This paper utilizes a new method for optimizing either of these similarity measures, which is simply based on evaluations of the objective function, in order to determine an approximation to the gradient.

2. BRIEF REVIEW OF MUTUAL INFORMATION

2.1. Definition of Mutual Information

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[6-10]. If A and B are two images to register, $P_A(a)$ and $P_B(b)$ are defined as the marginal probability distributions, and $P_{AB}(a,b)$ is defined as the joint probability distribution of A and B. Then mutual information is defined as :

$$I(A,B) = \sum_{a,b} p_{AB}(a,b).\log \frac{p_{AB}(a,b)}{p_A(a).p_B(b)}$$

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

$$I(A,B) = \sum_{a,b} h_{AB}(a,b).\log \frac{N.h_{AB}(a,b)}{h_A(a).h_B(b)}$$

where N is the sum of all the values in the histogram. This number is equal to the total number of pixels in A and B, but it is higher if an interpolation, e.g. a bilinear partial surface distribution[6], is used. The histograms are computed using original gray levels or the gray levels of pre-processed images such as edge gradient magnitudes or wavelet coefficients.

From this definition, it can be easily shown that the mutual information of two images is maximal when these two images are identical. Therefore, in the context of image registration, mutual information 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 indicate that in this context, mutual information enables one to extract an optimal match with much better accuracy than cross-correlation.

2.2. Previous Experiments Evaluating Mutual Information versus Correlation

In previous work[11,12], we compared mutual information (MI) to correlation as a similarity measure for the registration of remote-sensing images. Correlation had been more widely used in this context, while mutual information had been used extensively in medical image registration. We showed that mutual information produces consistently sharper peaks at the correct registration values than correlation, which is important for obtaining subpixel registration accuracy. Moreover the sharper peaks are also produced at the lower resolution sub-band images produced by the wavelet decomposition. This indicates that MI can produce more accurate results than correlation in a multi-resolution registration scheme based on wavelets. 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. Previous results showed that MI registration is done correctly even in the presence of noise utilizing both Simoncelli and Daubechies wavelets, though results were only presented[11] for Simoncelli wavelets.

Timing results for MI registration was about 1.5 times more than that for correlation registration based on our prior experiments. We also showed that when noise was present in the input image, both correlation and MI continue to produce perfect registration for Gaussian noise levels up to –12db for our tests, with mutual information being more robust to noise than correlation. Further work[13] involved the extension of our tests to multi-sensor remote-sensing datasets including a comparison of different registration schemes.

3. REGISTRATION USING MUTUAL INFORMATION WITH WAVELETS

As described in our previous work[11,12], mutual information is utilized in a multi-resolution wavelet-based framework. Our algorithm is described in Figure 1, where both the reference and input images are decomposed via a wavelet decomposition, and then wavelet subbands are iteratively registered by maximizing their mutual information. This iterative multi-resolution scheme enables one to speed-up computations as well as to refine the deformation transformation with an increasing level of detail, when going up the wavelet pyramid.
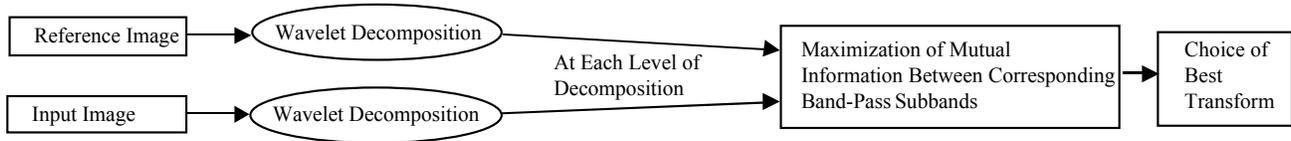

Figure 1
Summary of Our Wavelet & Mutual Information Image Registration Method

4. STOCHASTIC GRADIENT OPTIMIZATION FOR IMAGE REGISTRATION

4.1. Brief Survey

Prior work on optimization techniques for image registration can be found in references [6],[10] and [16]-[18]. The techniques described in [16]-[18] are all based on minimizing a sum of square differences. Maes et. al.[6,7] use the Marquardt- Levenberg technique to optimize mutual information (MI). The required derivatives are explicitly calculated based on a partial volume interpolation of the criterion, and it is implemented in a multiresolution framework. Thevenaz et. al.[16] develop a scheme to optimize an integrated sum of square differences in the intensity values of the images. They use a Marquardt- Levenberg algorithm, and computation of the derivatives and Hessian is based on a spline interpolation of the transformed images, which also works in a multiresolution manner. This work was applied to medical imagery, and is extended in [10] to the maximization of the mutual information similarity criterion. An optimizer is designed specifically for this criterion, which requires the criterion to be differentiable so that the gradient and Hessian matrix can be computed. Differentiability is achieved by the use of Parzen windows. Eastman and LeMoigne[18] describe a registration algorithm based on minimizing the square difference of the intensity values using a gradient descent algorithm. This is done in a multiresolution framework, and results are provided for remote sensing datasets which are multitemporal and/or multisensor. Irani and Peleg[17] 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 [17] does not involve multiple resolutions of the images. The technique to be presented here is a multiresolution scheme based on a wavelet decomposition, and it does not require explicit derivation of the required gradient vector.

4.2. A New 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[14], and it has recently attracted considerable attention for solving challenging optimization problems for which it is difficult or impossible to directly obtain a gradient of the objective function with respect to the parameters being optimized. 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.

In this study, we consider a parameter search space of two-dimensional rigid motions, consisting of an x and y-translation and a rotation in the plane. There are thus three parameters to be optimized, and these can be put in vector form as $\gamma = [t_x, t_y, \Theta]$. At each iteration, the gradient approximation is based on only two function measurements (regardless of the dimension of the parameter space). At iteration k, the update law for the parameters is steepest ascent:

$$\gamma_{k+1} = \gamma_k + a_k\, g_k$$

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

$$(g_k)^i \;=\; \{L(\gamma_k + c_k\, \Delta_k) - L(\gamma_k - c_k\, \Delta_k)\}/\{2\, c_k\, (\Delta_k)^i\}, \quad \text{for } i=1, 2 \ldots m$$

For i=1,2 … m, $(\Delta_k)^i$ takes a value of +1 or –1, as generated by a Bernouilli distribution. a_k and c_k are positive sequences of numbers which may be chosen to decrease to zero, as described in [14]. These must be optimized for the particular problem in order to improve the convergence rate. Spall[14] has outlined suggested forms for these parameters which will ensure convergence of the algorithm. Note that for this problem m=3, since three parameters are to be updated at each iteration, as indicated above. This technique is a very powerful method, 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. The two objective functions to be optimized in this work are mutual information (MI) and correlation.

The reference image is rotated and translated to search for that transformation, which will provide the best match with the input image. In our tests, the transformed image is obtained using nearest neighbor interpolation, since this is the simplest to implement and is not very computationally intensive. But this type of interpolation at the subpixel level, produces an MI surface with the step-like form shown in Figure 2(a). This surface was plotted for x- and y-translation increments of 0.1, with Θ fixed at a value of zero. Note that the MI-value remains constant over a range, and changes at parameter translations of 0.5. Based on suggested parameter values provided in [14] and some heuristics based on our observation of the surface under consideration, the following parameter values were chosen for our implementation of the SPSA algorithm:

$$c_k = 0.5$$
$$a_k = 0.5(1 + 1/(k+1)^\alpha)$$

The parameter α was chosen to be 0.5, and the results provided in Section 5 below were all produced using these parameter values.
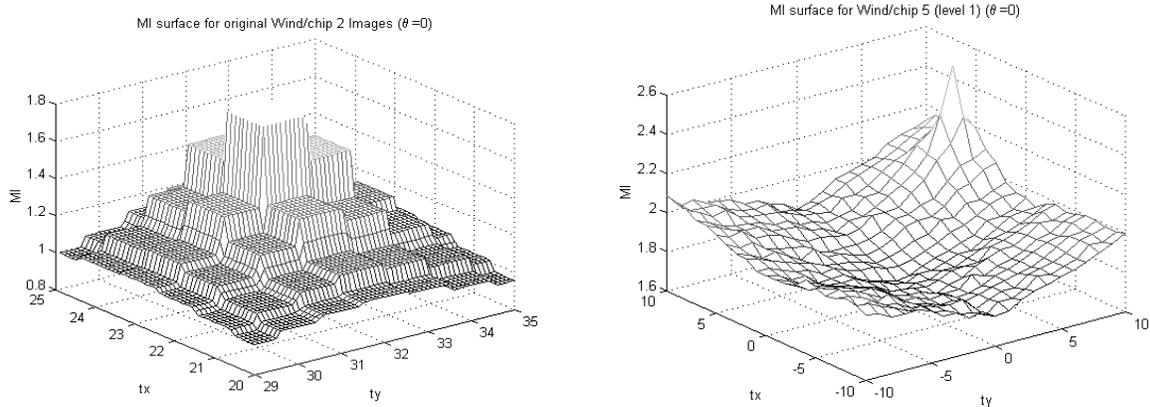
(a) At the subpixel level, for wind2 & chip2   *(b) For wind5 & chip 5 (Level 1) showing the global maximum, some local minima, and the 'lifting effect' as the offset increases.*

Figure 2
Mutual Information Surfaces

It is important to note that in searching for the maximum MI value, the rotation parameter also varies, and so a definite peak in the MI-value probably exists in the three dimensional parameter space, but this is not easy to verify graphically.

5. EXPERIMENTS AND RESULTS

5.1. Description of the Datasets and the Experiments

In this study two datasets are used. The first dataset, the Landsat dataset, consists of 7 pairs of images of size 256X256, each of which was 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)[19]. These pairs of images shown in Figures 3, are referred to as wind and chip respectively, and the Landsat-5 windows are registered to the corresponding Landsat-7 chips. Our second dataset, the AVHRR dataset, corresponds to a series of multi-temporal NOAA Advanced Very High Resolution Radiometer (AVHRR) scenes acquired over South Africa. These differed from the reference by very small translations and no rotations, and are shown in Figure 4. Note the varying locations of clouds in the images.

The SPSA optimization algorithm was first run on the 256X256 wind and chip datasets with no wavelet decomposition, and then on multiple resolutions from a Daubechies decomposition. Our experiments investigate the performance of the optimization algorithm for the mutual information and correlation metrics when used in conjunction with the Daubechies wavelet decomposition.

Using Daubechies wavelets, 3 levels of decomposition are processed (i.e. levels 0-2) for the Landsat images and the feature space for matching is composed of the gray levels of the Low-Pass output images at each level of decomposition. The images at each of these three levels of decomposition correspond to a decimation of 2, 4 and 8 of the original image, respectively. For the larger AVHRR dataset, four levels of decomposition were processed, up to a decimation of 16.
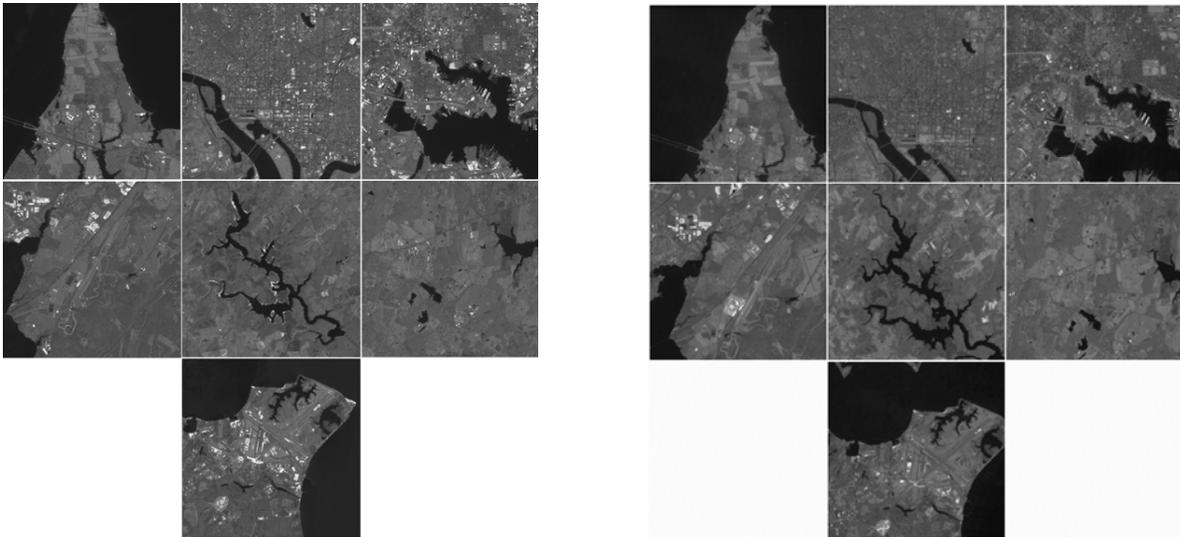
Figure 3 - *The Landsat Dataset*

Seven Chips (256x256) Extracted   *Seven Corresponding Windows (256x256) Extracted*
from Band 4 of a 1999 Landsat-7 Scene   *from Band 4 of a 1997 Landsat-5 Scene*
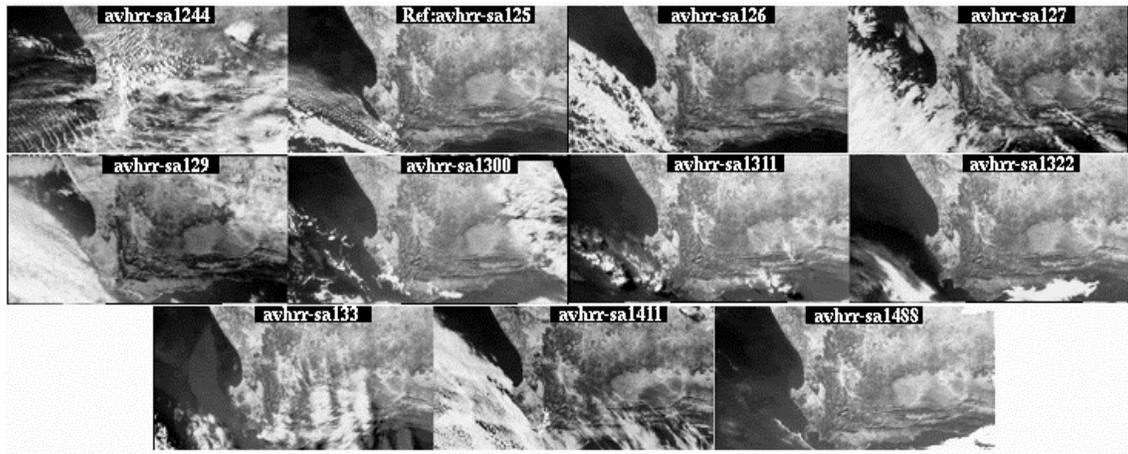
Figure 4 - *The AVHRR Dataset*
Series of AVHRR Scenes (1024x512) Acquired Over South Africa

An observed effect for small images of fixed size was that the MI value tends to artificially rise as the offset in the registration value increases (referred to as the 'lifting effect' of the surface). See Figure 2(b) showing an MI surface for the problematic wind/chip_5 image pair. This effect can be especially pronounced for the smaller-sized lower resolution images produced by the wavelet decomposition. It can be remedied by using a 'weighted MI', as suggested by Carranza & Loew[15], but this modified (weighted) mutual information, which takes into account the area of overlap between the images being registered, only works well with a full exhaustive search. The suggested weighting is fairly arbitrary and non-smooth so this modification does not work well in an optimization scheme.

5.2 Results

The different components of our algorithm are tested and compared using our two datasets. The results presented here were obtained by using multiple levels of decomposition and a search using the SPSA optimization technique at each level of decomposition.

Results for the datasets are presented in Tables 1-5. In the last column of these tables a/b indicates that the maximum function value (MI or correlation) is reached at a iterations, and remains constant over the total number of iterations, b. An entry, such as 400~ means that the optimum was not yet achieved, and the function value was still changing. An entry, such as 200* means that the result of the optimization was incorrect for the starting point chosen, and correct results were not produced at an arbitrary starting point of $[t_x, t_y, \Theta] = [0,0,0]$ for that level. In addition, for the 'typical' data pair, [wind2,chip2], the results from Tables 1-3 are expanded as shown in Figure 5. These show the convergence rate of the relevant parameters for the optimization of MI and also for correlation. Note that in all cases the parameters, which undergo MI optimization achieve their optimum value faster than those which undergo correlation optimization. In addition for the original image at the top level of the decomposition, the effect of a near optimum initial condition provided by the wavelet approach, is also seen to produce convergence in much fewer iterations than from an 'arbitrary' starting point of $[t_x, t_y, \Theta] = [20,35,0]$.

A check of the level 3 decompositions (not shown) of the wind and chip images of size 16, showed that most of the MI results were incorrect, while correlation produced correct results at level 3 for about half of our tests. The incorrect results were sometimes due to the algorithm being trapped at a local maximum, and this frequently occurred because the search proceeded from the starting point in a direction away from the global optimum due to the 'lifting effect' on the surface, caused by the offsets as described above. It may be an indication that the MI criterion is more susceptible to this 'lifting effect' than is correlation for the smaller-sized images. This also brings up the question of the choice of starting point and availability of prior information on the registration values. Tables 4 and 5 provide the results for the AVHRR images at four levels of decomposition (i.e. levels 0-3).

At level 2, with a decimation of 8 of the original image, MI produced incorrect results for the wind/chip 5 and 6 pairs, as shown by the results in Table 1, while the correlation results for avhrr 1311 were incorrect at level 0. It is critical, with the multiresolution approach, that a correct result is obtained at the coarsest level of the decomposition so as not to propagate and multiply errors. An underlying assumption in all our tests is some prior knowledge of the range of the actual registration values, and our goal is to achieve a result with subpixel accuracy by the use of our algorithm. The use of this prior information allows us to discard incorrect optimization results at any level, and it also allows us to chose a reasonable starting point at the coarser levels of the wavelet decomposition.

In the multiresolution approach provided by the wavelet decomposition, when the optimization algorithm works at a coarser level, it provides a near optimal starting point for the next level. This can produce convergence in as few as one iteration, and it also increases the robustness of the algorithm since it becomes less likely that it will be trapped in a local minimum at the next level. Thus a considerable speed up in the overall registration process can be achieved. In general, MI was found to optimize (faster) in fewer iterations than correlation did. At the subpixel level, there was a small discrepancy between the optimal values obtained for MI versus those obtained for correlation, but this cannot be resolved without accurate, independent ground truth estimates of the true transformations.

5.3 Comparison to Exhaustive Search and Discussion

At each level of decomposition, evaluation of the MI or correlation objective function takes up a major portion of the computational time. For the exhaustive search, the required number of function evaluations is n^3, if we use the same number of data points, n, for each of the search ranges of tx , ty and Θ. In the multiresolution implementation, the optimum parameter values from the coarser level are passed to the next level and used as the center of the search range at that level.

For the exhaustive search used in prior work[12], the search range decreases at each decomposition level from the lowest resolution image to the original size. Consider a 4-level search involving decomposition levels 0-2. At the top (or finest) level, each parameter is searched over a 5 data point range, obtained by adding 2 points in both directions; the next coarsest level adds 4 data points, giving a total of 9 points; then 8 points are added, and so on. Thus the total number of function evaluations required at each level of the exhaustive search is shown below.

Daubechies Level	Number of function evaluations (Exhaustive search)
2	varies
1	4913
0	729
original	125

At the coarsest level, the initial search depends on the availability or not, of prior information of the correct registration values. At this level, the default search range for the translation parameters is chosen to be [-1,1] which contains 3 parameter values, but the number of data points for rotation may vary between 3 and 359. Thus the number of function evaluations can vary between 27 to 3231 for the default setting, and may in fact be much more.

In the SPSA optimization, all parameters are updated simultaneously, and convergence occurs at most levels in less than 200 iterations. Each iteration requires two function evaluations to obtain the gradient, so the number of function evaluations is double the number of iterations. From the results of the multiresolution optimization algorithm provided in Tables 1-5, we observe that we achieve a speed up by a factor of at least 3 at top level for mutual information, over the exhaustive search. Correlation optimization, which requires more iterations, is still faster than the corresponding exhaustive search if the search time over all levels is added up. The total computational cost can be evaluated by using the assignment of unit processing time to a level 2 image, then a level 1 image is processed in 4 time units, etc..

Finally, it should be noted that since the exhaustive search is done in discrete unit increments it is not possible to get subpixel registration results using this search method.

6. CONCLUSIONS

The study presented in this paper investigated a new optimization technique, and successfully applied it to register remote sensing images in a multiresolution framework, using wavelets. We observed that with the SPSA optimization algorithm, MI tends to optimize 3 to 4 times faster than correlation does (see Table 3). Current work involves evaluating and comparing the application of Simoncelli wavelets for this process, and investigating the effect of a smoother interpolation scheme for the transformed image on the accuracy of the optimization result. Such a scheme will involve an increase in the computational cost of the algorithm but it may be justified if there is an improvement in the accuracy of the result. Further work will also involve the extension of our tests to a larger number of remote sensing datasets, testing the effect of noise and of large rotations on the algorithm.

ACKNOWLEDGMENTS

REFERENCES

1. L. Brown, "A Survey of Image Registration Techniques," *ACM Computing Surveys*, vol.**24**, no.4, 1992.
2. L.M.G. Fonseca and B.S. Manjunath, "Registration Techniques for Multisensor Sensed Imagery," *Photogrammetric Engineering and Remote Sensing Journal*, **62** (9), pp. 1049-1056, Sept. 1996.
3. J. Le Moigne, "Parallel Registration of Multi-Sensor Remotely Sensed Imagery Using Wavelet Coefficients," *SPIE's OE/Aerospace Sensing, Wavelet Applications*, Orlando, pp. 432-443, April 1994.
4. W.K. Pratt, "Correlation Techniques of Image Registration," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. AES10, No. 3, 353-358, 1974.
5. D.I. Barnea, and H.F. Silverman, "A Class of Algorithms for Fast Digital Registration," *IEEE Transactions on Computers*, Vol. C-21, 179-186, 1972.
6. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, P. Suetens, "Multimodality Image Registration by Maximization of Mutual Information," *IEEE Trans. on Medical Imaging*, Vol.**16**, No.2, April 1997.

7. F. Maes, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality Image Registration Using Multiresolution Gradient-Based Maximization of Mutual Information," *First NASA/Goddard Image Registration Workshop*, Greenbelt, MD, Nov. 1997, NASA Publication #CP-1998-206853, pp. 191-200.
8. W.M. Wells, III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-Modal Volume Registration by Maximization of Mutual Information," *Medical Image Analysis*, Vol. **1**, pp.35-51, 1996.
9. H-M. Chen, and P. Varshney, "A Pyramid Approach for Multimodality Image Registration Based on Mutual Information," *Proc. of Fusion 2000*, Paris, France, July 10-13, 2000, pp. MoD3-9 - MoD3-15.
10. P. Thevenaz, and M. Unser, "Optimization of Mutual Information for Multiresolution Image Registration," *IEEE Transactions on Image Processing*, Vol. **9**, No. 12, December 2000.
11. K. Johnson, A. Cole-Rhodes, J. LeMoigne, I. Zavorin, "Multi-resolution Image Registration of Remotely Sensed Imagery using Mutual Information," *SPIE Aerosense 2001, Wavelet Applications VIII*, Vol.4391, FL, April 2001.
12. K. Johnson, A. Cole-Rhodes, I. Zavorin, J. LeMoigne, "Mutual Information as a Similarity Measure for Remote Sensing Image Registration," *SPIE Aerosense 2001, Geo-spatial Image & Data Explotation II*, Vol. 4383, Orlando, FL, April 2001.
13. J. LeMoigne, A. Cole-Rhodes, R. Eastman, K. Johnson, J. Morisette, N. Netanyahu, H. Stone, I. Zavorin, "Multi-Sensor Registration of Remotely Sensed Imagery," *8th International Symposium on Remote Sensing*, Vol.4541, Toulouse, France, Sept 2001.
14. Spall, J.C, "Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation," *IEEE Transactions on Automatic Control,* Vol. 37, 332-341 (1992).
15. C.E. Rodriguez-Carranza M.H. Loew, "A Weighted and Deterministic Entropy Measure for Image Registration Using Mutual Information," *SPIE Conf. on Image Processing*, San Diego, CA, Feb. 1998, Vol 3338, pp.155-166.
16. P. Thevenaz, U.E. Ruttiman, and M. Unser, "A Pyramid Approach to Subpixel Registration Based on Intensity," *IEEE Transactions on Image Processing*, Vol.**7**, No.1, January 1998.
17. M. Irani, S. Peleg, "Improving Resolution by Image Registration," *CVGIP: Graphical Models and Image Processing*, Vol. **53**, No. 3, May 1991, 231-239.
18. R. Eastman, J. Le Moigne, "Gradient-Descent Techniques for Multi-Temporal and Multi-Sensor Image Registration of Remotely Sensed Imagery," *FUSION'2001, 4-th Int.Conf. Information Fusion*, Canada, Aug.2001.
19. J. Le Moigne, N. Netanyahu, J. Masek, D. Mount, S. Goward, "Robust Matching of Wavelet Features for Sub-Pixel Registration of Landsat Data," *IGARSS'01,* Sydney, Australia, July 9-13, 2001.
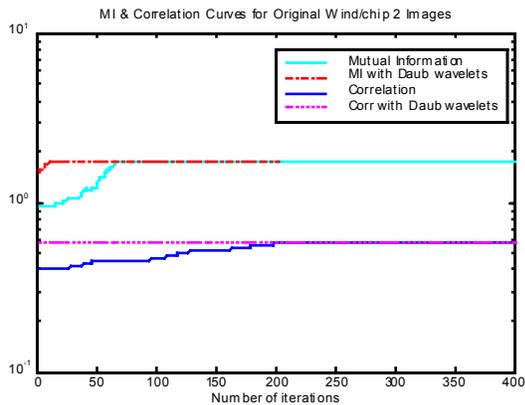
(a) Optimization curves for MI and correlation
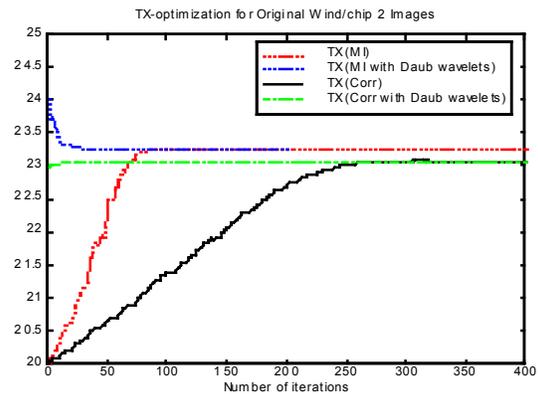
(b) Optimization curves for TX, using MI and correlation
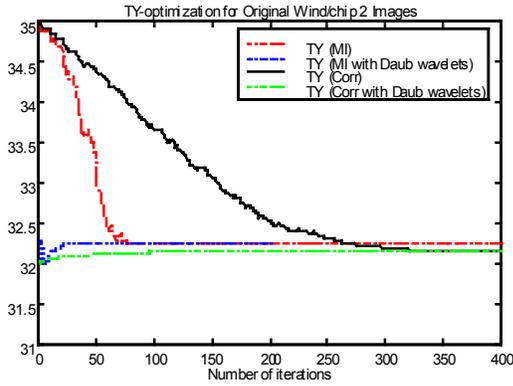
(c) Optimization curves for TY, using MI and correlation
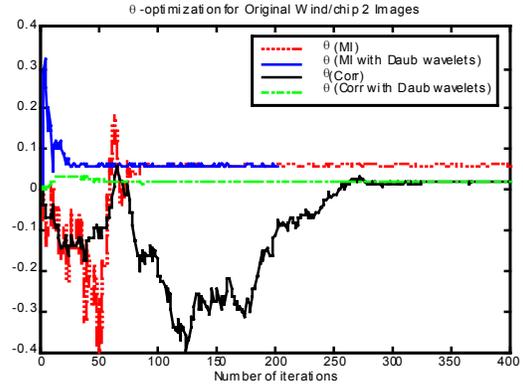
(d) Optimization curves for rotation θ, using MI and correlation

Figure 5
Parameter Optimization Curves for the Registration of Original Wind2 and Chip2 Images

Table 1
Results of Correlation Optimization using Daubechies Wavelets on the Landsat Dataset

Image	Lev	Starting Pt Tx/Ty/θ	Starting Correl	Ending Pt Tx/Ty/θ	Correl	Iteration
wind/chip_ 0	2	2/4/0	0.8873	2.99/3.98/0.368	0.9674	9/200
	2	4/4/0	0.8653	2.99/3.99/0.398	0.9674	9/200
	1	6/8/0	0.9632	5.94/7.93/0.248	0.9632	0/200
	0	12/16/0	0.9548	11.73/15.64/0.017	0.9548	0/200
	orig	24/31/0	0.9544	23.44/31.097/0.037	0.9584	200~
wind/chip_ 1	2	2/4/0	0.2713	2.98/4.02/-1.1	0.3943	14/200
	1	6/8/0	0.4086	5.90/8.05/0.02	0.4086	0/200
	0	12/16/0	0.4088	11.82/16.63/0.08	0.4265	48/200
	orig	24/33/0	0.4578	23.45/33.128/-0.001	0.4662	66/200
wind/chip_ 2	2	2/4/0	0.4738	2.94/3.96/-0.036	0.5338	31/200
	1	6/8/0	0.5777	5.886/7.96/-0.0117	0.5777	0/200
	0	12/16/0	0.5750	11.587/16.0/0.066	0.5750	0/200
	orig	23/32/0	0.5892	23.08/32.14/0.018	0.5892	0/200
wind/chip_ 3	2	2/4/0	0.4492	3.38/3.199/2.54	0.554	500~
	1	6/6/0	0.4737	6.01/7.59/0.466	0.551	500~
	0	12/15/0	0.5467	11.50/15.35/0.117	0.557	200~
	orig	23/30/0	0.5470	22.92/30.795/0.146	0.5721	185/200
wind/chip_ 4	2	2/4/0	0.1498	2.73/3.99/0.44	0.1754	391/400
	1	6/8/0	0.1669	5.16/8.00/0.465	0.1772	200
	0	10/16/0	0.1897	10.98/16.07/0.332	0.225	200
	orig	22/32/0	0.2281	22.13/32.53/0.07	0.2329	400~
wind/chip_ 5	2	2/4/0	0.2916	2.96/4.12/0.116	0.3430	14/200
	1	6/8/0	0.3174	5.897/8.91/0.02	0.3592	24/200
	0	12/18/0	0.3618	11.41/17.55/-0.03	0.3664	84/200
	orig	22/35/0	0.3798	22.95/35.098/-0.04	0.4132	31/200
wind/chip_6	2	2/4/0	0.8440	2.933/4.00/-0.124	0.8827	33/200
	1	6/8/0	0.8754	5.91/7.93/-0.18	0.8754	0/200
	0	12/16/0	0.8659	11.85/15.41/-0.008	0.8747	139/200
	orig	24/31/0	0.8796	23.66/30.798/0.003	0.8796	0/200

Table 2
Results of Mutual Information (MI) Optimization using Daubechies wavelets on the Landsat Dataset

Image	Lev	Starting Pt Tx/Ty/θ	Starting MI	Ending Pt Tx/Ty/θ	MI	Iteration
wind/chip_ 0	2	0/0/0	3.501	3.39/4.08/-0.097	4.050	315/350
	2	2/4/0	3.8764	3.113/4.07/-0.145	4.050	4/200
	1	0/0/0	2.259	5.98/7.99/-0.14	3.035	346/400

	1	6/8/0	3.035	5.99/8.04/-0.64	3.035	0/200
	0	12/16/0	2.074	11.8/15.82/-0.08	2.074	0/200
	orig	24/32/0	1.602	23.31/31.15/0.08	1.749	21/200
wind/c hip_1	2	2/4/0	3.404	3.05/4.04/-0.22	3.6158	4/200
	1	6/8/0	2.4065	6.02/8.09/0.134	2.4065	0/200
	0	12/16/0	1.3129	11.91/16.73/0.02	1.473	3/200
	orig	24/34/0	0.8580	23.48/33.2/0.008	1.1716	9/335
wind/chip_2	2	2/4/0	4.2160	2.999/4.011/0.189	4.3662	8/200
	2	4/4/0	4.238	2.99/4.04/0.22	4.3662	26/100
	1	0/0/0	2.1759	6.04/8.09/1.12	3.006	1000
	1	6/8/0	3.1078	5.98/8.05/0.28	3.1078	0/100
	0	12/16/0	2.088	11.86/16.05/0.29	2.088	0/100
	orig	24/32/0	1.492	23.25/32.23/0.06	1.762	21/200
wind/chip_3	2	0/0/0	3.081	2.99/4.04/-0.11	3.6954	91/200
	2	2/4/0	3.4076	2.99/4.02/-0.003	3.6954	5/200
	1	0/0/0	1.379	5.92/7.97/-0.07	2.288	456/500
	1	6/8/0	2.2878	5.92/7.99/0.04	2.2878	0/200
	0	12/16/0	1.4959	11.27/15.70/-0.08	1.735	8/200
	orig	22/30/0	1.090	22.63/31.17/0.026	1.539	5/200
wind/chip_4	2	2/4/0	4.402	3.863/4.114/0.089	4.475	111/200
	1	8/8/0	2.737	5.99/8.04/0.349	3.062	39/200
	0	12/16/0	1.852	11.21/16.54/-0.03	2.02	200~
	orig	22/33/0	1.468	22.61/32.95/0.22	1.5514	8/200
wind/c hip_5	2	2/4/0	3.529	4.38/10.02/0.88	3.955	200**
	1	8/0/0	1.869	6.19/8.75/1.65	2.310	1000~
	0	12/18/0	1.368	11.60/17.64/0.04	1.368	0/200
	orig	23/35/0	1.3486	23.09/35.10/-0.01	1.3486	0/200
wind/chip_6	2	2/4/0	4.442	7.623/7.97/0.39	4.896	200**
	1	0/0/0	2.3559	6.02/7.87/1.26	3.173	300~
	0	12/16/1	2.2469	11.96/15.3/0.08	2.378	21/200
	orig	24/30/0	1.7502	23.76/30.75/0.03	1.944	8/200

Table 3

Results of Optimization, Landsat Dataset, without wavelets from an arbitrary starting point of tx/ty/theta=20/35/0

Mutual Information					
Wind/Chip images	Starting Pt Tx/Ty/θ	Starting MI	Ending Pt Tx/Ty/θ	Ending MI	Iterations
wind/chip_0	20/35/0	1.2809	23.309/31.156/0.0833	1.7492	150/400
wind/chip_1	20/35/0	0.3871	23.48/33.20/0.011	1.1716	43~/700~
wind/chip_2	20/35/0	0.9561	23.253/32.235/0.0597	1.76213	65/400
wind/chip_3	20/35/0	0.80698	22.63/31.169/0.025	1.5391	92/400
wind/c hip_4	20/35/0	0.65795	22.61/32.94/0.028	1.55141	28/400
wind/chip_5	20/35/0	0.65733	23.094/35.097/-0.008	1.34865	8/400
wind/chip_6	20/35/0	1.15283	23.7613/30.762/0.029	1.9440	106/400
Correlation					
wind/chip_0	20/35/0	0.8820	23.25/31.53/0.12	0.9531	700~
wind/c hip_1	20/35/0	0.2191	23.449/33.124/-0.002	0.46616	118/400
wind/chip_2	20/35/0	0.41278	23.08/32.15/0.018	0.55892	232/400
wind/chip_3	20/35/0	0.41582	22.81/30.91/0.08	0.5721	452/700
wind/chip_4	20/35/0	0.1503	22.13/32.54/0.07	0.23365	700~
wind/c hip_5	20/35/0	0.2977	22.97/35.092/-0.03	0.41318	149/400
wind/chip_6	20/35/0	0.7848	23.26/31.07/-0.138	0.8726	700~

Table 4

Results of Mutual Information (MI) using Daubechies wavelets on the AVHRR Dataset

Image (Manual Registration)	Lev	Starting Pt Tx/Ty/θ	Starting MI	Ending Pt Tx/Ty/θ	MI	Iteration
avhrr 126	3	0/0/0	4.279177	0.01/0.03/-0.41	4.279177	0/200
(0/0/0/0)	2	0/0/0	2.820572	-0.01/0.02/-0.39	2.820572	0/200
	1	0/0/0	1.711433	-0.06/-0.15/-0.01	1.71143	0/200
	0	0/0/0	1.028833	-0.31/-0.25/0.01	1.028833	0/200

Image	Lev	Starting Pt Tx/Ty/θ	Starting Correl	Ending Pt Tx/Ty/θ	Correl	Iteration
	orig	0/0/0	1.637575	-0.42/-0.22/-0.01	1.637575	0/29
avhrr 127	3	0/0/0	4.176531	0.05/0.09/0.15	4.176531	0/200
(-1/-1/0)	2	0/0/0	2.574838	0.01/0.01/0.43	2.574838	0/62
	1	0/0/0	1.288376	-0.26/-0.35/-0.00	1.288376	0/200
	0	-1/-1/0	0.601903	-0.71/-0.64/-0.00	0.601903	0/200
	orig	-1/-1/0	0.972912	-1.50/-0.82/-0.03	1.014969	450 ~
avhrr 129	3	0/0/0	4.093746	1.77/1.00/-0.16	4.154608	177/200*
(-3/-1/0)	2	0/0/0	2.448219	-1.001/2.01/0.17	2.459606	389/500*
	1	0/0/0	1.161929	-0.88/0.21/0.14	1.174389	600 ~
	0	-2/0/0	0.521364	-1.21/-0.07/0.12	0.529380	600 ~
	orig	0/0/0	0.799681	-2.57/-0.66/0.07	0.841950	750 ~
avhrr 133	3	0/0/0	4.169636	0.01/0.03/-0.45	4.169636	0/300
(0/-1/0)	2	0/0/0	2.659957	0.02/-0.02/-0.46	2.663372	0/300
	1	0/0/0	1.524865	0.01/-0.32/-0.08	1.524865	0/300
	0	0/-1/0	0.953245	-0.11/-0.65/-0.02	0.953245	0/300
	orig	0/-1/0	1.386609	-0.26/-1.23/-0.01	1.386609	0/300
avhrr 1244	3	0/0/0	4.316372	2.15/1.43/0.10	4.363943	194/380*
(1/0/0)	2	0/0/0	2.657322	-0.02/0.32/0.16	2.657322	0/81
	1	0/0/0	1.276138	0.49/-0.01/0.03	1.285468	600 ~
	0	0/0/0	0.465121	0.49/-0.17/0.004	0.465121	0/248
	orig	0/0/0	0.649399	0.76/-0.32/-0.07	0.654045	600 ~
avhrr 1300	3	0/0/0	4.263280	1.30/1.17/-0.32	4.291931	69/200
(2/0/0)	2	1/1/0	2.596047	0.85/1.10/0.11	2.596047	0/200
	1	2/2/0	1.272731	1.05/0.16/-0.13	1.372877	224/300
	0	2/0/0	0.715468	1.52/-0.01/-0.06	0.718038	300 ~
	orig	3/0/0	1.290954	2.71/0.7/-0.03	1.290954	0/300
avhrr 1311	3	0/0/0	4.339528	0.01/-0.02/-0.38	4.339528	0/300
(1/0/0)	2	0/0/0	2.845347	0.03/-0.02/-0.36	2.845347	0/300
	1	0/0/0	1.701039	0.38/-0.16/-0.09	1.701039	0/300
	0	1/0/0	1.136558	0.86/-0.37/-0.03	1.137582	300 ~
	orig	2/-1/0	1.529245	1.74/-0.87/-0.01	1.529245	0/300
avhrr 1322	3	0/0/0	4.411436	0.00/-0.02/0.39	4.411436	0/300
(0/-1/0)	2	0/0/0	2.956676	0.01/-0.00/-0.44	2.956676	0/263
	1	0/0/0	1.803839	0.13/-0.22/-0.09	1.803839	0/300
	0	0/0/0	1.149540	0.30/-0.59/-0.01	1.88324	271/300
	orig	1/-1/0	1.721903	0.58/-1.04/-0.04	1.725832	300 ~
avhrr 1411	3	0/0/0	4.300184	0.08/0.22/-0.15	4.300184	0/300
(0/-1/0)	2	0/0/0	2.667522	0.04/0.14/-0.33	2.669109	293/300
	1	0/0/0	1.288783	0.03/-0.27/-0.10	1.288783	0/24
	0	0/-1/0	0.657162	-0.03/-0.99/-0.01	0.657162	0/300
	orig	0/-2/0	1.160988	-0.05/-1.82/-0.10	1.173493	300 ~
avhrr 1488	3	0/0/0	4.333458	0.60/-0.04/-0.38	4.338499	122/200
(2/3/0)	2	1/0/0	2.726772	0.34/0.17/-0.29	2.739452	154/300
	1	1/0/0	1.491790	0.77/0.71/-0.02	1.621418	26/300
	0	2/1/0	0.955991	1.27/1.48/-0.01	0.995065	300 ~
	orig	3/3/0	1.613441	2.82/3.08/0.01	1.613441	0/300

Table 5

Results of Correlation using Daubechies wavelets on the AVHRR Dataset

Image	Lev	Starting Pt Tx/Ty/θ	Starting Correl	Ending Pt Tx/Ty/θ	Correl	Iteration
avhrr 126	3	0/0/0	0.624229	0.01/-0.18/0.09	0.624229	0/400
(0/0/0)	2	0/0/0	0.609182	0.02/-0.29/0.15	0.609182	0/400
	1	0/0/0	0.594873	-0.002/-0.47/0.01	0.594873	0/55
	0	0/0/0	0.578333	-0.06/-0.36/-0.02	0.578333	0/53
	orig	0/0/0	0.596399	-0.32/-0.87/-0.02	0.597129	237/400
avhrr 127	3	0/0/0	0.369078	0.01/-0.10/0.23	0.369078	0/400
(-1/-1/0)	2	0/0/0	0.371508	-0.0004/-0.17/0.28	0.371508	0/400
	1	0/0/0	0.366964	-0.07/-0.34/0.07	0.366964	0/58
	0	0/0/0	0.354653	-0.17/-0.27/0.05	0.354653	0/60
	orig	0/0/0	0.418214	-1.22/-1.56/0.07	0.411439	700~
avhrr 129	3	0/0/0	0.482614	0.002/-0.04/0.41	0.482614	0/400
(-3/-1/0)	2	0/0/0	0.470858	-0.01/-0.03/0.47	0.471624	400 ~
	1	0/0/0	0.456853	-0.37/-0.30/0.15	0.456853	0-59
	0	0/0/0	0.443043	-0.82/-0.67/0.11	0.446756	400 ~

	orig	-2/-1/0	0.361033	-1.94/-1.34/0.26	0.355661	600~
avhrr 133	3	0/0/0	0.638302	0.03/-0.07/.031	0.638302	0/400
(0/-1/0)	2	0/0/0	0.623101	0.04/-0.13/0.34	0.623101	0/397
	1	0/0/0	0.605265	0.08/-0.37/0.06	0.605265	0/65
	0	0/0/0	0.586242	0.08/-0.94/0.05	0.596173	160/400
	orig	0/-2/0	0.573482	0.13/-1.75/0.07	0.574968	400 ~
avhrr 1244	3	0/0/0	0.227456	0.47/0.18/-0.09	0.227546	0/81
(1/0/0)	2	0/0/0	0.224989	0.24/0.20/-0.01	0.224989	0/600
	1	0/0/0	0.222097	0.44/0.21/0.05	0.222097	0/133
	0	0/0/0	0.217905	0.41/0.17/0.04	0.217905	0/145
	orig	0/0/0	0.289095	0.73/0.35/0.16	0.290349	500 ~
avhrr 1300	3	0/0/0	0.542354	-0.0008/0.89/0.19	0.548916	26/400*
(2/0/0)	2	0/0/0	0.524704	0.01/0.83/-0.49	0.520711	400 ~*
	1	0/0/0	0.508132	0.22/0.33/-0.07	0.508132	0/60
	0	0/0/0	0.493240	0.89/0.82/-0.12	0.495783	500
	orig	0/0/0	0.545601	1.31/0.61/-0.3	0.553674	700~
avhrr 1311	3	0/0/0	0.728088	0.03/-0.24/0.03	0.728088	0/400
(1/0/0)	2	0/0/0	0.710181	0.10/-0.38/0.02	0.710181	0/400
	1	0/0/0	0.690711	0.09/-0.22/-0.11	0.690711	0/12
	0	0/0/0	0.672312	0.75/-1.29/-0.05	0.672734	177/400*
	orig	0/0/0	0.644073	1.05/-0.90/-0.02	0.655385	367/500
avhrr 1322	3	0/0/0	0.621460	0.02/-0.001/0.46	0.621460	0/400
(0/-1/0)	2	0/0/0	0.616881	0.04/-0.07/-0.15	0.616881	0/400
	1	0/0/0	0.605167	0.14/-0.26/-0.000	0.605167	0/400
	0	0/0/0	0.589438	0.21/-0.26/0.05	0.589438	0/105
	orig	0/0/0	0.595403	0.54/-0.80/0.08	0.605295	500 ~
avhrr 1411	3	0/0/0	0.155454	0.06/-0.12/0.15	0.155454	0/400
(0/-1/0)	2	0/0/0	0.175088	0.08/-0.22/0.15	0.175088	0/400
	1	0/0/0	0.176979	0.13/-0.47/0.01	0.176979	144/265
	0	0/0/0	0.169303	0.10/-0.96/0.02	0.195345	132/500
	orig	0/-2/0	0.300815	0.07/-2.01/-0.03	0.300815	0/400
avhrr 1488	3	0/0/0	0.659720	0.03/-0.04/-0.38	0.659720	0/400
(2/3/0)	2	0/0/0	0.637544	0.02/-0.01/-0.43	0.637544	0/100
	1	0/0/0	0.607426	0.56/0.22/0.06	0.607407	298/400
	0	1/0/0	0.589391	1.06/0.40/0.003	0.589391	0/400
	orig	2/1/0	0.571891	2.54/1.74/0.10	0.580913	400 ~

Table 6
Summary of Results using Daubechies wavelets on the AVHRR Dataset

Mutual Information (up to 500 iterations)			
AVHRR image	**Manual Registration (Tx/Ty/θ)**	**Optimizing (Tx/Ty/θ)**	**MI-value**
126	0/0/0	-0.42/-0.22/-0.01	1.637575
127	-1/-1/0	-1.50/-0.82/-0.03	1.014969
129	-3/-1/0	-2.57/-0.66/0.07	0.841950
133	0/-1/0	-0.26/-1.23/-0.01	1.386609
1244	1/0/0	0.76/-0.32/-0.07	0.654045
1300	2/0/0	2.71/0.7/-0.03	1.290954
1311	1/0/0	0.86/-0.37/-0.03	1.137582
1322	0/-1/0	0.58/-1.04/-0.04	1.725832
1411	0/-1/0	-0.05/-1.82/-0.10	1.173493
1488	2/3/0	2.82/3.08/0.01	1.613441
Correlation (up to 700 iterations)			
126	0/0/0	-0.32/-0.87/-0.02	0.597129
127	-1/-1/0	-1.37/-1.65/0.05	0.408933~
129	-3/-1/0	-1.94/-1.339/0.26	0.355661~
133	0/-1/0	0.13/-1.75/0.07	0.574968~
1244	1/0/0	0.73/0.35/0.16	0.290349
1300	2/0/0	1.31/0.61/-0.3	0.553674
1311	1/0/0	1.05/-0.90/-0.02	0.655385
1322	0/-1/0	0.54/-0.80/0.08	0.605295~

| 1411 | 0/-1/0 | 0.07/-2.01/-0.03 | 0.300815 |
| 1488 | 2/3/0 | 2.74/1.97/0.12 | 0.582691~ |