

# Automatic Registration of Airborne Images with Complex Local Distortion

Desheng Liu, Peng Gong, Maggi Kelly, and Qinghua Guo

## Abstract

*Accurate registration of airborne images is challenging because complex local geometric distortions are often involved in image acquisition. In this paper, we propose a solution to this registration problem in two parts. First, we present an area-based method to extract sufficient numbers of well-located control points, and second, we use the extracted control points with local transformation models to register multi-temporal airborne images. The proposed image registration methods were applied to two airborne images with complex local distortion. Performance was evaluated and compared using different transformation models (global models and local models), different numbers of control points, and different similarity measures (correlation coefficient and mutual information). The results showed that local models outperformed global models, more control points could significantly improve local transformation models but not on the global transformation models, and two similarity measures performed similarly. These results revealed two important findings: first, the area-based methods generated larger amounts of evenly distributed control points; and second, local transformation models achieved better registration accuracy when larger amount of evenly distributed control points are used. We concluded that the combination of area-based control point extraction with local transformation models is effective for the registration of airborne images with complex local distortion.*

## Introduction

Image registration is the process of geometrically aligning two or more images of the same scene. It is a critical preprocessing procedure in all remote sensing applications that utilize multiple image inputs, including multi-sensor image fusion, temporal change detection, and image mosaicking. With the recent explosive increase in remotely sensed imagery, and the need for change detection and modeling, automatic image registration has become increasingly important for the integration of multi-temporal and multi-sensor data (Moigne,

2002). Earth system science, for example, will rely on automated tools for the integration and registration of multi-temporal and multi-sensor satellite and airborne imagery as the necessary first step in the analysis of seasonal and annual global climate change and land-use, land-cover change (Townshend *et al.*, 1992; Dai and Khorram, 1998). Additionally, images acquired from aircraft (as opposed to satellites) often need additional registration efforts due to sensor parameter changes from mission to mission (Coulter *et al.*, 2003).

Automatic registration of images has generated extensive research interests in the fields of computer vision, medical imaging, and remote sensing. Comprehensive reviews include Brown (1992) and Zitova and Flusser (2003). One review on remote sensing applications can be found in Fonseca and Manjunath (1996). Image registration generally consists of four steps: (a) control point extraction, (b) transformation model determination, (c) image transformation and resampling, and (d) registration accuracy assessment (Jensen, 2004). Among the four steps, the first two are most complex, and their success essentially determines the final registration accuracy. Since small registration error (less than half pixel) is often hard to achieve, post-registration algorithms (Knoll and Delp, 1986; Gong *et al.*, 1992) have also been proposed to reduce the effect of misregistration in temporal difference images.

While many of the standard remote sensing texts discuss manual control point extraction (Schowengerdt, 1997), this method can be subjective and extremely time consuming. It often results in few usable points, with poor spatial distribution across the image, both of which can reduce the overall registration accuracy (Kennedy and Cohen, 2003). Automated extraction of control points is a solution to this problem, and is attempted by either feature-based methods or area-based methods. Feature-based methods apply feature extraction algorithms to generate a certain number of obvious features first, and then match these features to form the control point pairs. The commonly used features are landmarks (e.g., edges, corners, the centers of gravity of regions, and line intersections) extracted from the images. Feature-based methods are robust in regards to intensity change and geometric distortion, however, they rely on the existence of robust features invariant between images and efficient feature extraction algorithms which are not always easy to implement (Kennedy and Cohen, 2003). Moreover, the possibility of too few or unevenly distributed features may limit their use in some applications (Chen *et al.*, 2003). In contrast, area-based methods work with subset images directly without feature extraction and compare the similarity of sub-images within

---

Desheng Liu and Peng Gong are with the Department of Environmental Science, Policy and Management, University of California, Berkeley and the Center for the Assessment and Monitoring of Forest and Environmental Resources (CAMFER), 137 Mulford Hall, Room 3114, Berkeley, CA, 94720-3114.

Maggi Kelly is with the Department of Environmental Science, Policy and Management, University of California, Berkeley and the College of Natural Resources Geospatial Imaging and Informatics Facility, 137 Mulford Hall, Room 3114, Berkeley, CA, 94720-3114 (mkelly@nature.berkeley.edu).

Qinghua Guo is at the School of Engineering, University of California at Merced, P.O. Box 2039, Merced, CA 95344.

---

Photogrammetric Engineering & Remote Sensing  
Vol. 72, No. 9, September 2006, pp. 1049–1059.

0099-1112/06/7209-1049/\$3.00/0  
© 2006 American Society for Photogrammetry  
and Remote Sensing

corresponding windows at different displacements. Area-based methods are usually computationally intensive and may not be applicable to image pairs with large rotational distortion or displaying phenological change, but they have the advantage of being relatively easy to implement and require no pre-processing of images. Moreover, the regular grid distribution of the resulting control points is often desirable for the registration of images with complex local distortion, because sufficient numbers of control points can cover every possible local distortion region that might be found across an image.

Transformation models used in image registration are either global transformations or local transformations depending on whether the transformation function is location dependent (Brown, 1992). The registration of satellite imagery is commonly modeled as a global deformation, but the local geometric distortion is not always negligible depending on the imaging geometry, the terrain variation, atmospheric turbulence, and the sensor non-linearity (Richards and Jia, 1999). In particular, airborne images with high spatial resolution often have complex local deformation caused by the wide view angle, the terrain variation, the low flight height, and the effects of yaw, pitch, and roll (Devereux, 1990). For complex local deformation, piecewise mapping functions (Goshtasby, 1986, 1987, and 1988b) and surface spline mapping functions (Goshtasby, 1988a; Flusser, 1992) are often used as local transformation models. These models usually need a large number of evenly distributed control points or features to represent the local variation. Therefore, the quantity and spatial distribution of extracted control points are crucial to the accuracy of the registration of images with complex local deformation.

Since feature-based control point extraction methods rely on the extraction and matching of prominent features to define the control points, these methods are not always capable of finding enough evenly distributed features for defining the control points in each local deformed region, especially for images that lack structures or patterns. Moreover, the spatial distribution of the resultant control points corresponds to that of the matched features, which may not exist in every possible deformed local region. Additionally, feature extraction and matching are often difficult due to the dominance of ambiguous features. Although global matching (Moigne, 2002) can solve this problem, this method is not applicable to the local transformation models. Consequently, feature-based control point extraction methods are not always suitable for images with complex local deformation, especially when an image contains insufficient features or unevenly distributed features. In this regard, area-based methods are more appropriate in the selection of control points for the use of local transformation models, except in cases where local geometric differences are extremely high; in these cases, area-based matching is not stable. Generally, the control points can be evenly arranged on regular grids to represent the complex local distortion, and the matching can be obtained by optimizing some similarity measures based on the gray value of images.

The discussion above highlights two related facts of transformation models and control point extraction methods: first, local transformation models can work well only if a sufficient number of evenly distributed control points are used, and second, area-based methods provide an effective way to generate large number of well-distributed control points. Clearly, it is beneficial to combine area-based control point extraction methods with local transformation models for the correction of complex image distortions, but few applications reported in the literature explored this possibility. One purpose of this paper is to fill this gap.

In this paper, we focus on the automatic registration of high spatial resolution airborne imagery with complex local

deformation. It is worth noting that a frame center matching technique can be employed during the acquisition of the airborne images to alleviate many of the factors that complicate image distortion (Coulter *et al.*, 2003; Stow *et al.*, 2003). However, it is necessary to develop appropriate algorithms to account for complex local deformation for airborne imagery because in many cases data acquisition tasks are not coordinated. We first conducted a series of experiments on simulated data using different transformation models with the aim of testing the model effects on registration accuracy. We also tested the effects of different numbers of control points on transformation models with the simulated data. We then implemented the proposed image registration methods on real airborne images and conducted two types of comparisons: (a) registration models: global models versus local models, and (b) control points: different numbers and different similarity measures.

The area we are working in is a locally complex forested terrain experiencing a new forest disease called "sudden oak death." Because we wish to map the disease through time, and the resulting changes in the forest, correct spatial registration of multi-temporal imagery is an absolute necessity for automated landscape scale spatio-temporal modeling (Kelly and Meentemeyer, 2002; Kelly *et al.*, 2004). To date, spatial registration has been performed using manual control point extraction and global models, with pixel level accuracies. In this work, we wished to improve upon these results with automated methods. Thus, we see this work as potentially useful for others involved in integrated land change research using multi-temporal remote sensing data inputs, especially those using high spatial-resolution airborne imagery that, because of its mode of capture (individual frames, subject to variable acquisition conditions), can present spatial challenges not found with satellite imagery.

## Methods

### Control Point Extraction: Template Matching

Area-based control point extraction methods seek to optimize some predefined objective functions based on template matching. The basic assumption of these methods is that the objective functions are optimized at the position where two templates are correctly aligned. Similarity measures between two templates are often used to define the objective functions. A template can be a sub-image or the whole image depending on the complexity of image deformation. For a rigid deformation, the template is the whole image. An optimal matching leads to the final transformation model, thus control points are not explicitly generated. For a non-rigid deformation, a template is usually a sub-image so that a number of control points could be extracted. These control points are then used to determine the transformation model. Ideally, templates should be positioned at those unique and highly distinct local neighborhoods so that accurate matching can be achieved. However, control points extracted in those highly distinct areas may be insufficient to allow desirable transformation models of the entire image to be constructed, because highly distinct templates may not exist in every possible local distortion region where control points are needed to adjust the local influence of the transformation models. This is particularly true for images with complex local deformation, as is the case for our study area. Therefore, we propose an alternative approach to balance between the optimal template matching on informative local neighborhoods and local control on every possible deformation region. In this approach, the templates are placed on dense regular grids in order to cover every possible local deformation region. At the same

time, outlier detection is applied to the extracted control points in order to remove mismatched points on those non-unique and non-informative areas as a result of the adoption of regular grids.

The template matching process implemented here is illustrated in Figure 1. It starts with placing  $N \times N$  evenly spaced grids on both the master image (Figure 1a) and the slave image (Figure 1b). At each grid point  $[I, J]$ , three steps follow. First, an  $n \times n$  sub-image centered on grid point  $[I, J]$  in the master image is extracted as the master template (Figure 1c). Second, assuming the largest possible translation is  $m$  pixels, an  $(n + 2m) \times (n + 2m)$  sub-image centered on the grid point  $[I, J]$  in slave image is extracted as the buffer zone (Figure 1d) for the subsequent selection of slave template. Within the buffer zone, an  $n \times n$  sub-image

centered on  $[I + i, J + j]$  is extracted as the slave template (Figure 1e), where  $i, j \in [-m, \dots, -1, 0, 1, \dots, m]$  are translations in horizontal and vertical direction, respectively. In total,  $(2m + 1)^2$  such slave templates are extracted in the buffer zone. Third, a similarity measure is calculated between the master template and one of the  $(2m + 1)^2$  slave templates. On the 3D mesh plot of the similarity measure versus  $i$  and  $j$  (Figure 1f), the maximum similarity measure is obtained at  $[i_{max}, j_{max}]$ , where  $i_{max}, j_{max} \in [-m, \dots, -1, 0, 1, \dots, m]$ . Then, the control point pair is located at  $[I, J]$  in the master image and at  $[I + i_{max}, J + j_{max}]$  in the slave image. Sub-pixel level accuracy can be achieved by fitting a smooth surface to the similarity measures around  $[i_{max}, j_{max}]$  (Figure 1g) and interpolating between integer  $i$  and integer  $j$  (Figure 1h) (Kennedy and Cohen, 2003).

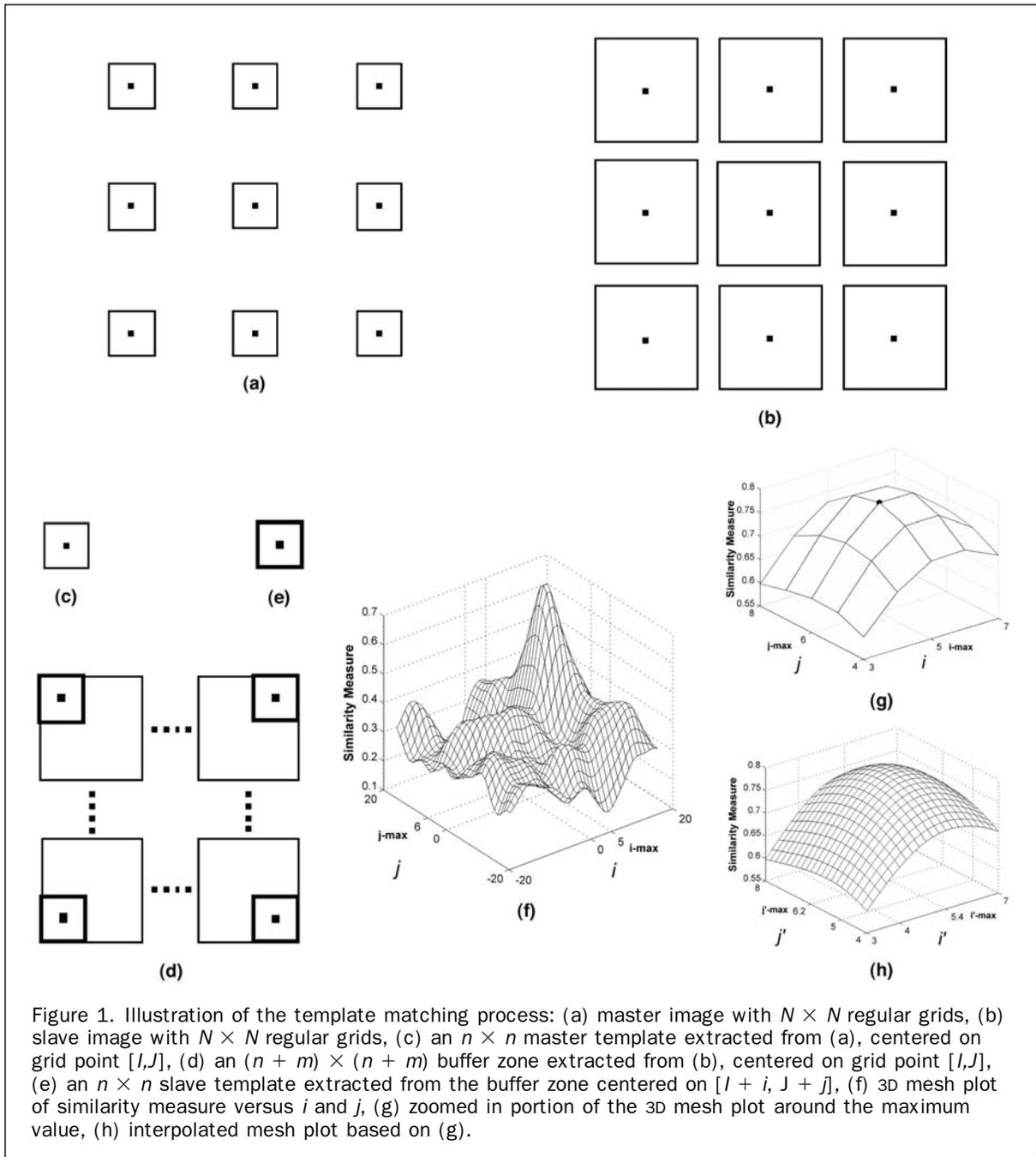


Figure 1. Illustration of the template matching process: (a) master image with  $N \times N$  regular grids, (b) slave image with  $N \times N$  regular grids, (c) an  $n \times n$  master template extracted from (a), centered on grid point  $[I, J]$ , (d) an  $(n + m) \times (n + m)$  buffer zone extracted from (b), centered on grid point  $[I, J]$ , (e) an  $n \times n$  slave template extracted from the buffer zone centered on  $[I + i, J + j]$ , (f) 3D mesh plot of similarity measure versus  $i$  and  $j$ , (g) zoomed in portion of the 3D mesh plot around the maximum value, (h) interpolated mesh plot based on (g).

### Control Point Extraction: Similarity Measures

Common similarity metrics used in image registration include sum of squared difference (SSD), normalized cross-correlation (NCC), correlation coefficient (CC), and mutual information (MI). In this paper, we only compared two similarity measures CC and MI for control point extraction since SSD and NCC are similar to CC in principle. The definitions and the properties of CC and MI are reviewed as follows.

CC is a standard statistic used in many similarity analyses. It treats two templates as two matrices of the same size and computes the two-dimensional correlation coefficient of two matrices. It is a statistical based similarity measure and ranges from  $-1$  to  $1$  thus giving normalized similarity. It is calculated as

$$CC(u,v) = \frac{\sum_x \sum_y \left\{ [M(x,y) - \mu_M][S(x-u,y-v) - \mu_S] \right\}}{\sqrt{\sum_x \sum_y [M(x,y) - \mu_M]^2} \sqrt{\sum_x \sum_y [S(x-u,y-v) - \mu_S]^2}} \quad (1)$$

where  $M$  is the master template (or reference image) and  $S$  is the slave template (or input image), the sum is over  $x, y$  of the master template under the window containing the slave template which is translated by  $u$  and  $v$  in horizontal and vertical direction with respect to master template, and  $\mu_M$  is the mean of the master template and  $\mu_S$  is the mean of the slave template.

CC largely depends on the radiometric intensity of the target imagery, and so it requires images to be of the same type and be radiometrically correlated. It also is sensitive to considerable intensity changes between images. These requirements may limit its application with images from different sensors, different sensor capture characteristics, and images with significant changes.

The second similarity measure discussed here, Mutual Information (MI) was first introduced in medical image registration (Viola *et al.*, 1997; Collignon *et al.*, 1995). Since then, it has become one of the favorite similarity measures for many researchers because of its superior performance. MI has its origin in information theory. It measures the statistical dependence or information redundancy of two random datasets. Unlike CC, MI does not assume a linear or monotonic functional relationship (Roche *et al.*, 2000); and it can be applied to multi-modal images with different radiometric properties and may be robust to actual intensity changes (Chen *et al.*, 2003; Inglada and Giros, 2004).

Given master template  $M$  and slave template  $S$ , the mutual information between  $M$  and  $S$  can be equivalently defined in the following three equations:

$$I(M,S) = H(M) + H(S) - H(M,S) \quad (2)$$

$$I(M,S) = H(M) - H(M|S) \quad (3)$$

$$I(M,S) = H(S) - H(S|M) \quad (4)$$

where  $H(M)$  and  $H(S)$  are the entropies of  $M$  and  $S$ ;  $H(M,S)$  is the joint entropy of  $M$  and  $S$ ;  $H(M|S)$  and  $H(S|M)$  are the conditional entropy of  $M$  given  $S$  and  $S$  given  $M$ . Denote  $P_M(i)$  and  $P_S(j)$  as the marginal probability distributions of  $M$  and  $S$  and  $P_{M,S}(i,j)$  as the joint probability distribution of  $M$  and  $S$ , MI between  $M$  and  $S$  can be calculated from:

$$I(M,S) = \sum_i \sum_j P_{M,S}(i,j) * \log \left( \frac{P_{M,S}(i,j)}{P_M(i) * P_S(j)} \right). \quad (5)$$

To estimate MI between  $M$  and  $S$  based on Equation 5, we only need to estimate the joint histogram between  $M$  and  $S$ ,  $h_{M,S}(i,j)$ , from which all the terms in Equation 5 can be

estimated:  $P_M(i) = \frac{1}{N} \sum_j h_{M,S}(i,j)$ ,  $P_S(j) = \frac{1}{N} \sum_i h_{M,S}(i,j)$ , and  $P_{M,S}(i,j) = \frac{1}{N} h_{M,S}(i,j)$ , where  $N$  is the total number of pixels within each template.

### Control Point Extraction: Outlier Detection

Spatial outliers are those spatial objects whose non-spatial attributes are significantly different from that of their spatial neighbors (Shekhar *et al.*, 2003). Control points extracted by the aforementioned methods are not always meaningful and may contain outliers (Kennedy and Cohen, 2003). Since the image deformation is assumed locally smooth and homogeneous, the outliers are detected as those deformation vectors differing greatly from their spatial neighbors in both direction and amplitude. A simple z-score test is used to detect the outliers in control points generated by the area-based

methods. Z-score is defined as:  $Z_i = \frac{(x_i - \bar{x}_i)}{\theta}$ , where  $x_i$  is

the testing spatial object,  $\bar{x}_i$  is the mean value of  $x_i$ 's neighbors, and  $\theta$  is the standard deviation of  $x_i$ 's neighbors. When the absolute z value is greater than a user predefined threshold,  $x_i$  is considered as an outlier. We used the difference in between  $x$  and  $y$  coordinates for a pair of points, instead of the absolute distance between the pair of control points in the z-score test because the difference in  $x$  and  $y$  coordinates contain information about both distance and direction. Therefore, the pair of control points is rejected if the distance between either the  $x$  or  $y$  coordinate is found to be an outlier.

### Transformation Model Determination

The transformation of the slave image coordinate system to the master image coordinate system follows the control point extraction from the image pairs. This process can be mathematically expressed by two sets of mapping functions with respect to two coordinate components of the master image. The coordinate mapping from any point  $[u,v]$  in the slave image coordinate system to the corresponding point  $[x,y]$  in the master image coordinate system is modeled as:

$$\begin{cases} x = f(u,v|\alpha) \\ y = g(u,v|\beta) \end{cases} \quad (6)$$

where  $\alpha$  and  $\beta$  are parameters of the mapping function  $f$  and  $g$ , respectively and are determined by control points, and are commonly determined by a global transformation.

There are many challenges presented by global transformations that require considerations here. For global transformation models such as affine transformation, projective transformation, and global polynomial transformation, the parameters  $\alpha$  and  $\beta$  are the same for all the points and are determined by all the control points; so, a single function is used to model the transformation for each component of coordinates. When the geometric distortion is complex and location dependent, global models become inadequate to model the image geometry. Linear functions such as affine transformation and projective transformation are too simple to take the local variation into consideration. Nonlinear functions such as global polynomials use the least-squares method to optimize the parameters, thus the local variation will be averaged across the whole image (Zitova and Flusser, 2003; Goshtasby, 1988). Consequently, the registration error of locally deformed images by global functions is usually large and the spatial distribution of the error also varies with the location.

Alternatively, local transformation models have been proposed to solve the problems of global models so that the

transformation models are sensitive to location. As a result, the parameters  $\alpha$  and  $\beta$  in Equation 6 vary across different local regions over the image and are significantly determined by the local control points. In general, the local models fall into two groups: (a) local interpolation methods such as piecewise linear models (Goshtasby, 1986), piecewise cubic models (Goshtasby, 1987), and surface spline models (Goshtasby, 1988a; Flusser, 1992), and (b) local approximation methods such as local weighted least-squares method (Goshtasby, 1988b) and local weighted mean models (Goshtasby, 1988b). The difference between these two types of models is that local interpolation models pass control points exactly, whereas local approximation models may change the values of control points. In this paper, piecewise linear models and local weighted mean models have been applied to our data and are briefly reviewed as follows.

A piecewise linear model (PL) first decomposes the entire image into pieces by triangulation, and then uses different linear mapping functions to model the local geometry for each piece. Triangulation is optimized so that the points inside each piece are closer to the vertices than to any other control points. Also, long thin triangles should be avoided to constrain the local influence of the control points. Consequently, control points generated from area-based methods are very suitable for the use of piecewise linear function to model complex local distortion because the regular grid makes triangulation rather regular, and sufficient numbers of control points can be generated to fit the complexity of the local variation.

A local weighted mean model (LWM) determines the transformation of an arbitrary point by the weighted mean of its nearby polynomials. First, a polynomial function with  $n$  parameters is fitted to each control point by using the control point itself and its nearest  $n-1$  control points. Then, the transformation of an arbitrary point is inferred by the weighted mean of all polynomials passing over that point. The weight is set to be non-zero only in the local region of the point so that the approximation is only influenced by the local control points. Obviously, more control points will make the approximation more localized; thus the local distortion will be better modeled. Note also that problems may arise if the control points are not uniform or sparse. In this case, no or few fitted polynomials fall into a region where no or few control points are generated. Consequently, control points generated from area-based methods would be suitable to this model in terms of the large amount and regular distribution of the distributed control points.

### Simulated Data Analysis

To gain better understanding on the effectiveness of different transformation models, we carried out some experiments by registering a simulated slave image with complex local distortion to a master image using the models discussed previously. The master image (Figure 2a) was generated as a standard checkerboard image, which has typically been used in image geometry analysis because of its regular grid configuration. The checkerboard image consisted of  $40 \times 40$  alternating black and white squares, each of which included  $20 \times 20$  pure black or white pixels. So, the master image consisted of  $800 \times 800$  pixels. The slave image (Figure 2b) was simulated by transforming the master image based on the following distortion function:

$$\begin{cases} u = x + 5 \sin\left(\frac{y\pi}{150}\right) - 5 \cos\left(\frac{x\pi}{150}\right) \\ v = y - 5 \sin\left(\frac{x\pi}{150}\right) + 5 \cos\left(\frac{y\pi}{150}\right) \end{cases} \quad (7)$$

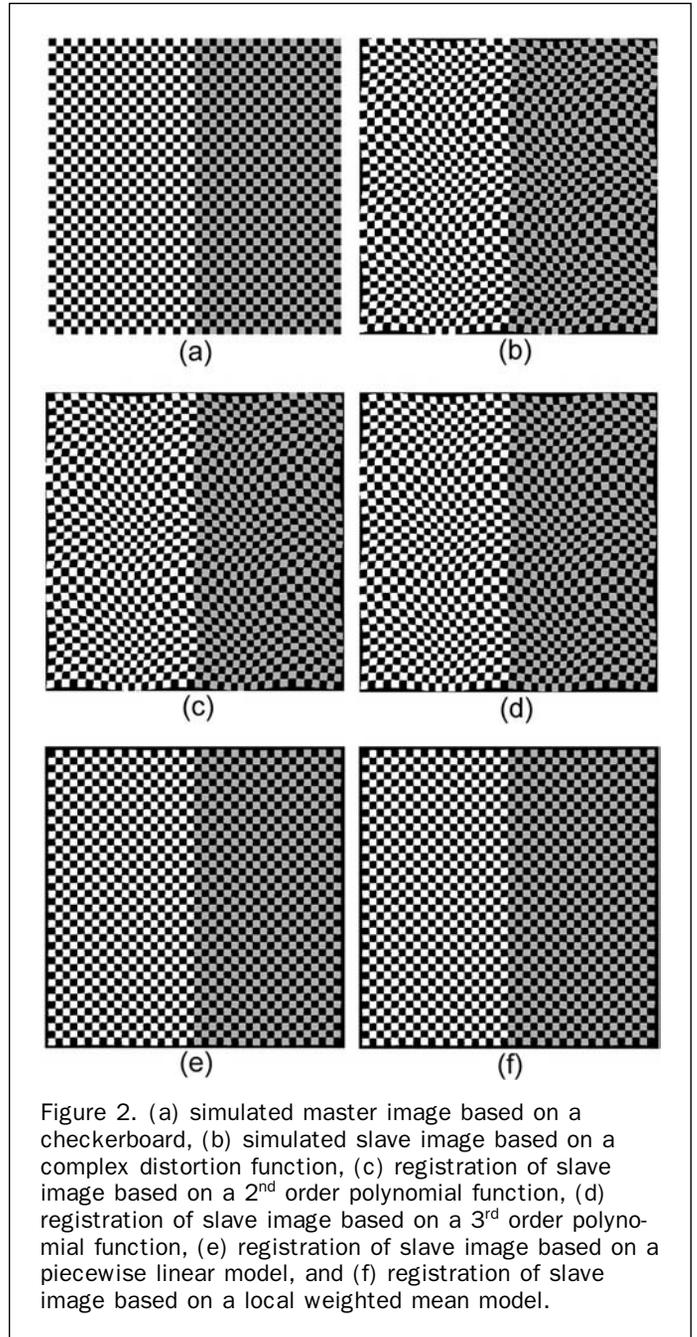


Figure 2. (a) simulated master image based on a checkerboard, (b) simulated slave image based on a complex distortion function, (c) registration of slave image based on a 2<sup>nd</sup> order polynomial function, (d) registration of slave image based on a 3<sup>rd</sup> order polynomial function, (e) registration of slave image based on a piecewise linear model, and (f) registration of slave image based on a local weighted mean model.

The determination of the transformation model is equivalent to inverting the above distortion function so as to obtain an inverse mapping from  $[u,v]$  to  $[x,y]$ . In a real situation, the distortion function would be unknown, whereas in this simulated case, it is known *a priori*. Two important properties of the distortion function in Equation 7 can be found. First, it is nonlinear in the two coordinate components; and second, it has no simple inverse function so the inverse mapping cannot be analytically expressed in a global function. Using the distortion function, we not only simulated a distorted image with complex local properties but also were able to obtain any numbers of perfect control points based on it. In this way, the simulated data enabled us to focus on model effects without considering the uncertainty involved in the generation of control points.

## Real Data Analysis

The study site for this research is a forested peninsula on the east side of Marin County, California, called China Camp State Park. The area has moderate to steep topography, with elevations ranging from the sea level at San Pablo Bay (the northerly lobe of the San Francisco Bay) to over 300 m. High spatial resolution imagery has been collected for the area for four years, using an airborne ADAR 5500 (Airborne Data Acquisition and Registration) sensor. We used imagery from spring 2000 and 2001. The ADAR 5500 imaging system is comprised of a SN4, 20 mm lens with four mounted cameras (spectral bands: band 1 (blue): 450 to 550 nm, band 2 (green): 520 to 610 nm, band 3 (red): 610 to 700 nm, band 4 (near-infrared): 780 to 920 nm) flown at an average aircraft altitude of 2,205 m. The cameras have a large field of view (typically about 35 degrees from nadir) to allow collection of high spatial resolution data from relatively low altitudes with large aerial coverage. The average ground spatial resolution of the imagery is one meter. Imagery was acquired in  $1000 \times 1500$  m frames with 35 percent side- and end-lap, and the frames were mosaicked and georeferenced, all by a private consultant (Positive Systems, Inc. of Montana). More information about the imagery can be found in Kelly *et al.* (2004).

A subset image of  $1500 \times 1500$  pixels was clipped from the mosaicked image for each year. This area has moderate to deep terrain variation, and includes forested and non-forested areas. The frame centers and the boundary of mosaics are illustrated in Figure 3. As the figure shows, both subset images are mosaics of six single frames and the frame centers are displaced over 300 pixels between the two years. Therefore, the two images show complex local geometric deformation due to the following reasons: (a) difference in the viewing angle and frame center, (b) large variation of the terrain, (c) lack of high spatial resolution and elevation accuracy DEM to orthorectify the frame images, and (d) error propagation in the georeferencing and mosaicking processes. The near-infrared band of each image was used for control point extraction because the study area is highly vegetated,

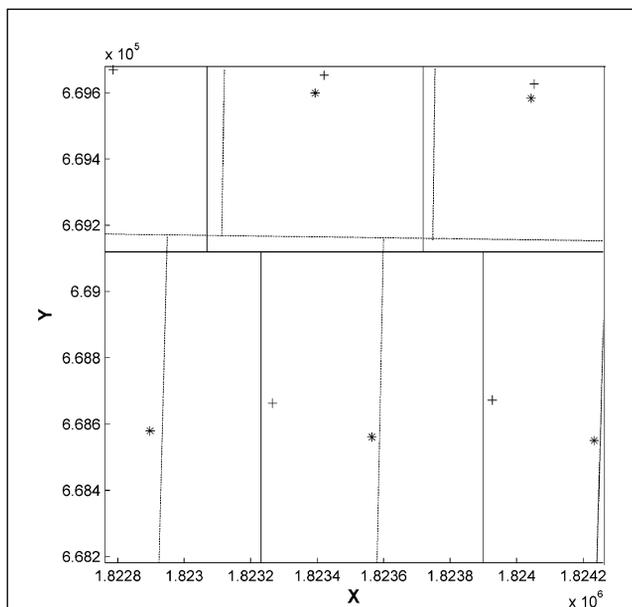


Figure 3. The frame centers (crosses and stars) and the boundary (dashed and solid lines) of mosaics: dashed lines and crosses are for year 2000 and solid lines and stars are for year 2001.

and the near-infrared band provides the best contrast between vegetation and other land covers (Cole-Rhodes *et al.*, 2003).

## Results

### Simulated Data Analysis

The slave image was transformed to match the master image by both global models (2<sup>nd</sup> order and 3<sup>rd</sup> order polynomial functions) and local models (PL and LWM) with perfect control points. The control points were arranged in regular grids with a grid size of 30 pixels, where the coordinates of each grid were calculated based on the known distortion function. The transformed slave images based on the global models and local models are shown in Figure 2c, 2d, 2e, and 2f. Registrations of the slave images by the global models did not show obvious changes, as compared to the original slave image. Most of the distortion with respect to the standard master image still remained. In contrast, the registration of slave images by the local models showed obvious change as compared to the original slave image, and most of the distortions with respect to the master image were successfully corrected.

To further investigate the effect of the number of control points on the transformation models, we varied the grid number from  $5 \times 5$  to  $50 \times 50$  and calculated the registration error for each resulting grid. The registration error was estimated as the mean of 10 sets of 300 independent control points generated at random locations based on the known distortion function. The curves of root mean squared errors (RMSE) with respect to the number of control points for all four transformation models are shown in Figure 4. The RMSE of the global models decreased about 0.5 pixel when the number of control points changed from  $5 \times 5$  to  $10 \times 10$ , but then remained steady with the number of control points increasing from  $10 \times 10$  to  $50 \times 50$ . This indicates that the number of control points has no significant effect on the registration error of global models as long as some minimum numbers of control points are used. A possible reason may be that the intrinsic inability of the global model itself has already resulted in large RMSE; thus the registration result will not be improved significantly even though more control points are used. In contrast, the impact of the number of

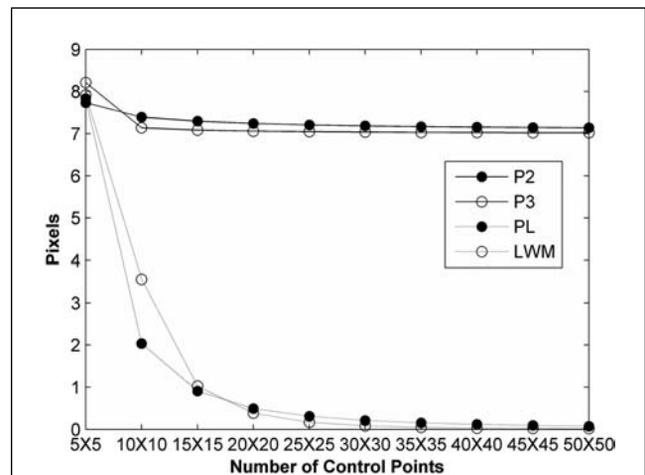


Figure 4. The relationship between RMSE and grid size for the four transformation models: P2—2<sup>nd</sup> order polynomial function; P3—3<sup>rd</sup> order polynomial function; PL—Piecewise Linear function; LWM—Local Weighted Mean function.

control points on the RMSE of local models is significant. The RMSE dramatically decreased by almost seven pixels when the number of control points increased from  $5 \times 5$  to  $25 \times 25$ . This may be attributed to (a) the effectiveness of local models to capture the complex local distortion, and (b) better representation of the complex local distortion by more control points. Meanwhile, the exponential shape of the RMSE curve shows that only marginal RMSE decreases are achieved when more than  $25 \times 25$  control points are used. This indicates that the local models perform well as long as sufficiently large numbers of control points are used. Therefore, for images with complex local distortion, both model type and the number of control points determine the registration accuracy. A local transformation model plus sufficient numbers of control points should be used to achieve good registration accuracy.

## Real Data Analysis

### Parameter Determination

When area-based methods are applied to two images with a predefined template window size on fixed grids, an appropriate window size  $n$  (refer to the Control Point Extraction Section) should be determined so that it is large enough to be statistically significant and stable but as small as possible to minimize the local geometric variation within the template. Window sizes of 31, 41, 51, 61, 81, 101 were tested with each similarity measure on the ground truth points, and the size of 61 was proved to be optimal to both measures in the sense that (a) the curves of CC and MI are smooth and stable, (b) the peaks are sharp and unique, and (c) the window size is minimum for (a) and (b). Some examples on known control points are shown in Figure 5. It showed that the curves of CC and MI became more stable and robust when the window size increased to 61. Another important parameter is the maximum possible translation  $m$  (as previously described) of the slave template with respect to the master template. Preliminary exploration of the two images indicates that the possible maximum displacement would not exceed 25 pixels, so the movement of the template is limited within 25 pixels in both horizontal and vertical directions.

The size of the regular grids  $N$  (refer to the Control Point Extraction Section) also should be determined so that the grid size is fine enough to represent the geometric variation but as coarse as possible to minimize the computational requirement and regularize the overall distortion. Without the knowledge of how many control points are needed, we generated 10 sets of control points using area-based methods (CC and MI) on regular grids ranging from  $5 \times 5$  to  $50 \times 50$ . So, the total number of the control points varies from 25 to 2,500.

### Control Point Extraction

CC and MI were applied to the real data with the parameters determined in the previous section (i.e.,  $n = 61$  pixels,  $m = 25$  pixels) for all the 10 sets of grids. For MI, the gray intensity level was reduced to 4-bit to have a more reliable histogram estimate. The master template was the  $61 \times 61$  sub-image of year 2001 centered on each grid. The slave template was the  $61 \times 61$  sub-image of year 2000 with the center moving within the specified range of each grid. The control point at each grid was determined as the position where the moving slave template maximizes the similarity measure. The maximum value was found by exhaustive search on the transformation space. To achieve sub-pixel level accuracy, a smooth surface was fitted to the  $5 \times 5$  similarity values around the maximum similarity value by using a cubic spline function. New similarity values on sub-pixel positions were interpolated from the  $5 \times 5$  original

similarity values. The control point with sub-pixel level accuracy was located at the position with the maximum similarity value among all the interpolated values. To show the most details, the deformation vector plots on a  $50 \times 50$  regular grid for the two similarity measures are plotted in Figure 6. (CC) and Figure 7. (MI).

The visualization of deformation vectors clearly illustrated the local details of the image distortions. The plots of CC and MI showed very similar spatial pattern of deformation vectors: the distortions at most parts of the images looked very irregular with a local expression with the deformation vectors varying in length and towards every direction; only the distortion at the upper right portion of the image was relatively homogeneous and small. This is in accordance with the preliminary observation of the complex local geometric deformation found on the images. Moreover, a certain number of control point outliers were observed in the two deformation vector plots. The outliers detected by z-score test are shown as circles on deformation vector plots in Figures 6 and 7. A closer look at the original image reveals that the outliers are mainly due to (a) homogeneous areas such as water and grassland as shown in Plate 1a and 1b, (b) considerable actual changes between two images due to forest gap dynamics and shadow effects as shown in Plate 1c and 1d.

### Transformation Models

The 2001 image was transformed with respect to the 2000 image by applying PL and LWM with the 10 sets of control points extracted from the previous section. Second order and third order global polynomial models were also applied for the purpose of comparison. With an aim of making the accuracy assessment more objective and reliable, we used a large set of control points for the test so that more local distorted region can be assessed. In total, 500 independent control points were generated with the same area-based method at 500 random positions to test the registration results.

The curves of RMSE for all the four models with respect to 10 sets of control points generated by CC and MI are shown in Figure 8. Similar trends to those observed with simulated data for all the four models were observed in the real data. For the global models, RMSE decreased by about 0.5 pixels when the control points increased from  $5 \times 5$  to  $10 \times 10$ , but after that RMSE stabilized. For the local models, RMSE dramatically decreased by over three pixels as the control points increased. When more than  $40 \times 40$  control points were used, the RMSE decreases was small. Comparison between global models and local models shows that local models are much better than global models when sufficient control points are used, but if the number of control points is small, the difference in RMSE is small also. Sufficient control points are therefore necessary to embody the advantages of local models over global models. This confirmed the conclusions drawn in the simulated data analysis. For global models, the 3<sup>rd</sup> order polynomial showed consistent 0.5 pixels smaller RMSE than the 2<sup>nd</sup> order polynomial. For local models, the PL showed slightly smaller registration error as compared to LWM. However, with the simulated data, PL outperformed LWM at the first five grid sizes and underperformed it at the last five grid sizes. We will explain this contrast in the Discussion Section. Comparison between CC and MI showed that the RMSE of MI almost coincided with that of CC for all the four models except that MI was 0.1 pixel more than that of CC for two local models when more than  $30 \times 30$  control points were used. This can also be explained by the small differences between control points from CC and MI as shown in Figure 6 and Figure 7. The histogram of the difference between control points from

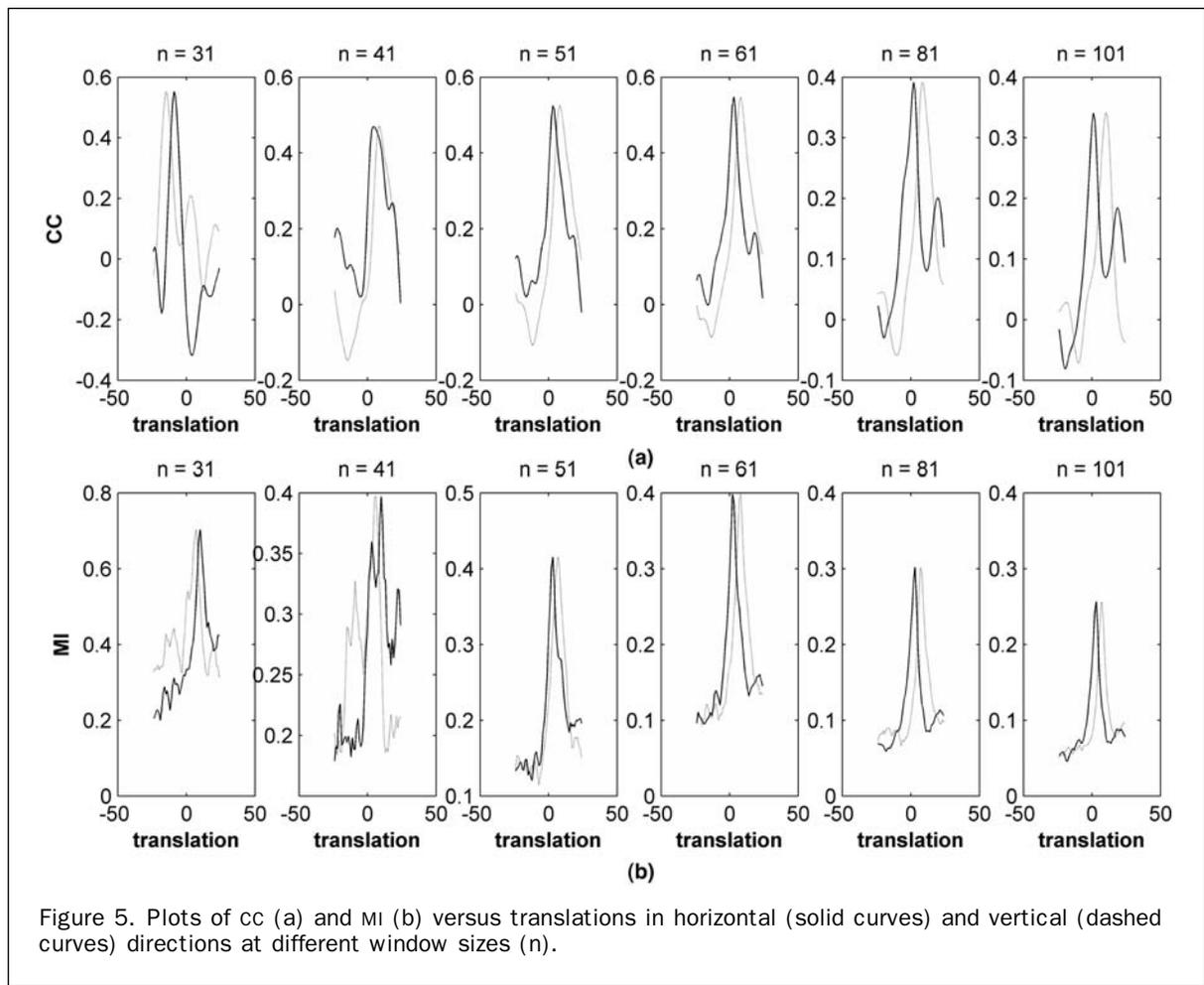


Figure 5. Plots of cc (a) and MI (b) versus translations in horizontal (solid curves) and vertical (dashed curves) directions at different window sizes (n).

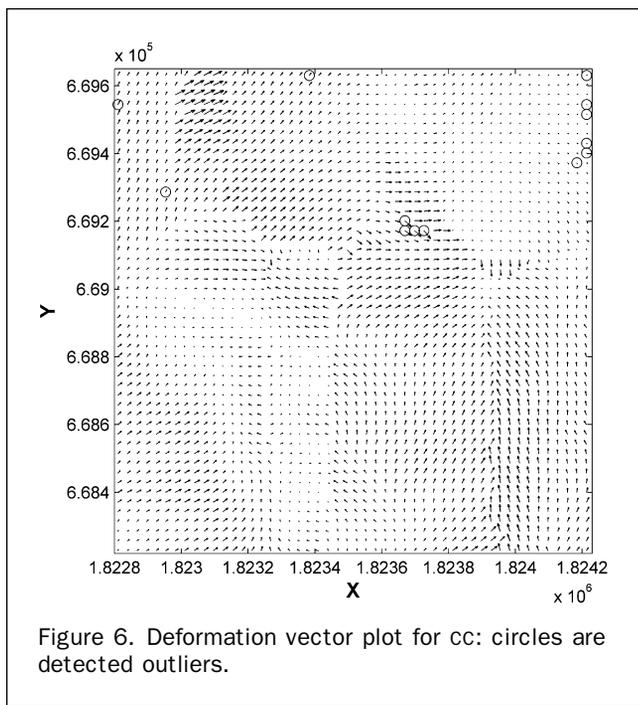


Figure 6. Deformation vector plot for cc: circles are detected outliers.

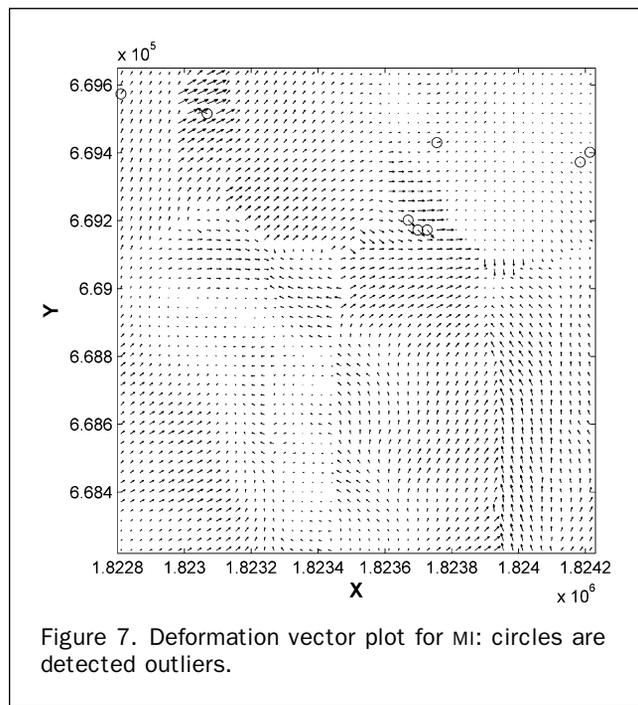


Figure 7. Deformation vector plot for MI: circles are detected outliers.

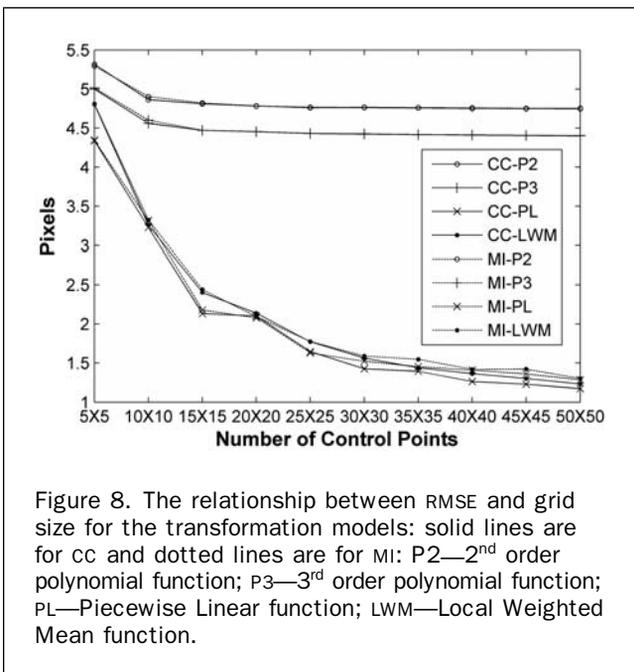
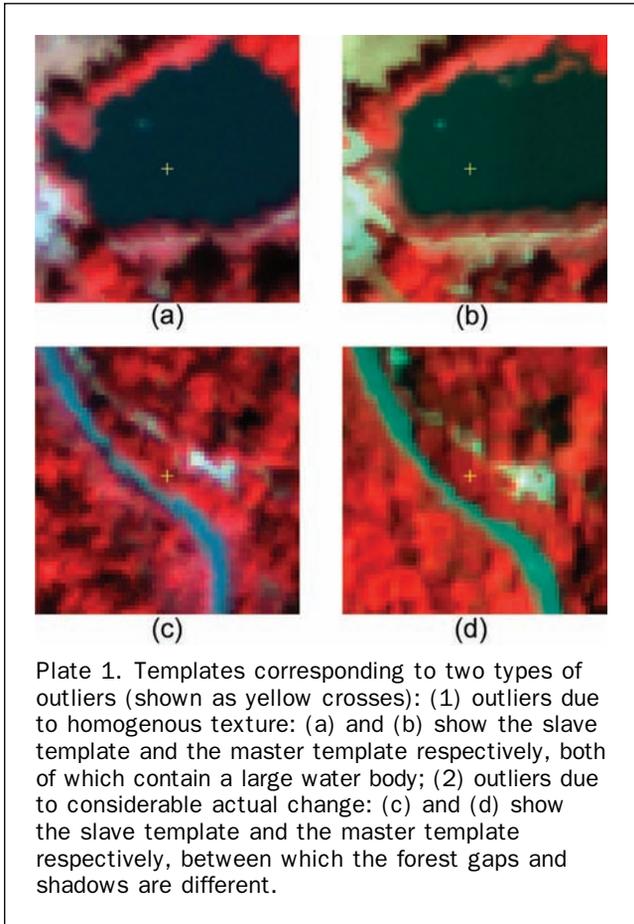


Figure 8. The relationship between RMSE and grid size for the transformation models: solid lines are for CC and dotted lines are for MI: P2—2<sup>nd</sup> order polynomial function; P3—3<sup>rd</sup> order polynomial function; PL—Piecewise Linear function; LWM—Local Weighted Mean function.

CC and MI shows that MI and CC agreed with each other in most cases; only very few difference exceeded one pixel. However, MI is expected to be superior to CC in terms of more robustness to radiometric difference. The similar results may be due to the fact that the multi-temporal images are

TABLE 1. ACCURACY ASSESSMENT FOR CC AND MI

| Control Point | Transformation Model | Max X Residual | Max Y Residual | Total RMSE | SD of RMSE |
|---------------|----------------------|----------------|----------------|------------|------------|
| CC            | P2                   | 14.0057        | 16.0999        | 4.7477     | 1.8110     |
|               | P3                   | 15.248         | 12.4618        | 4.4002     | 1.8083     |
|               | PL                   | 5.8049         | 4.4878         | 1.1720     | 0.5422     |
|               | LWM                  | 6.1566         | 5.0695         | 1.2328     | 0.5944     |
| MI            | P2                   | 13.1218        | 16.3994        | 4.7519     | 1.8232     |
|               | P3                   | 14.2352        | 16.1380        | 4.4039     | 1.7985     |
|               | PL                   | 6.0833         | 5.0310         | 1.2853     | 0.6024     |
|               | LWM                  | 6.7867         | 7.1461         | 1.3018     | 0.7026     |

Note: P2: 2<sup>nd</sup> order Polynomial.  
P3: 3<sup>rd</sup> order Polynomial.  
PL: Piecewise Linear.  
LWM: Local Weighted Mean.

from the same sensor (ADAR) and the radiometric correlation between two images is close, thus MI did not show its advantage over CC in our example.

**Registration Error Analysis**

To further analyze the pattern of registration errors, we generated four statistics of registration errors using 50 × 50 control points with CC and MI. These statistics include the maximum x-residual, the maximum y-residual, total RMSE, and the standard deviation (SD) of the residuals (Table 1). The local models have much smaller maximum x-residual and maximum y-residual than the global models. This may be due to the fact that the global models lose the control for some specific distortions by averaging them across the whole image while local models can approximate those local details. Moreover, standard deviations of residuals for local models are much smaller than those for global models, which implies that the spatial errors of local functions are more evenly distributed than those of global functions which is varying across the whole image.

**Discussion**

**Performance of LWM and PL**

As mentioned in the Transformation Models Section, the simulated data and real data displayed opposite results. To understand the performance of LWM and PL more completely, we experimented with different transformation functions with simulated data and found that the performance of LWM as compared to PL depends on the specific forms of the function taken. For example, we modified the sine and cosine term denominators, and added additional terms to the transformation functions. The results are shown in the Figure 9. These results indicate that there is no evidence that LWM always outperforms PL, or vice versa. This suggests to us that both models should be compared in future studies to determine which model is better for a specific application.

**Texture Analysis of Area-based Methods**

It is worth mentioning that the performance of area-based algorithms depends on image texture around the grid points. In cases where a grid point is located in texture-less areas (for example water, desert, and grassland), area-based correlation algorithms may have difficulty in finding optimal matches because there is no unique and highly distinct local neighborhood. The example shown in Plate 1a and 1b is such a case. To investigate this issue in more detail, we tested the area-based algorithms on another part of our image. The subset image has a size of 380 × 380 pixels, and

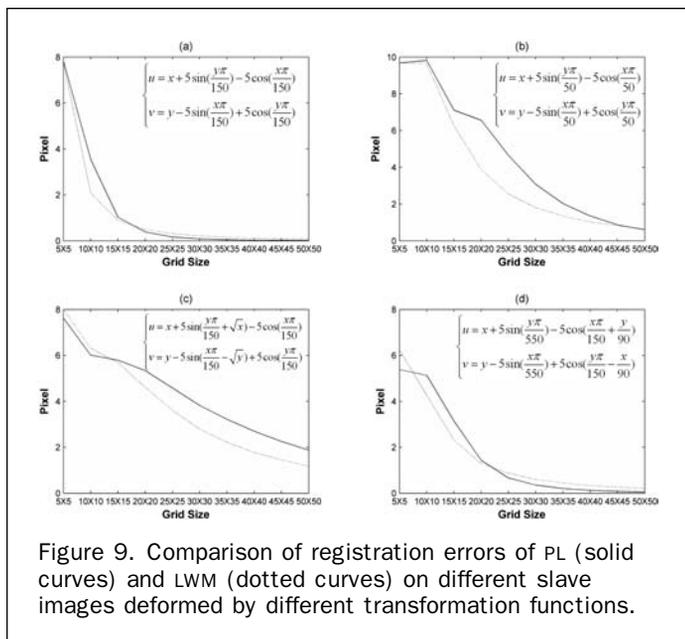


Figure 9. Comparison of registration errors of PL (solid curves) and LWM (dotted curves) on different slave images deformed by different transformation functions.

consists of water (low texture) on the upper portion and wetland (high texture) on the lower portion. The same set of parameters as used before was applied to this image and the deformation vector plots were overlaid on the image (Figure 10). It is interesting to observe that for both CC (Figure 10a) and MI (Figure 10b): (a) the deformation vectors corresponding to the wetland area show a regular pattern and reasonable local consistency, and (b) the deformation vectors corresponding to the water area are randomly oriented and show considerable local inconsistency. The deformation vectors in the water are meaningless and should be regarded as outliers. The failure of the area-based algorithm in bay water is obviously due to the extremely texture-less nature of the water, from which both CC and MI have difficulties to achieve the unique optimal matching.

The above results reveal one limitation of the area-based template matching algorithm on regular grids: it may fail with those grid points whose neighborhood are texture-less. In cases where texture-less areas are sparsely distributed over the entire image, one possible solution to this limitation is to relax the regular grids to constrained locally adaptive grids so that each grid is adaptively placed to textured local regions within the buffer zone of its regular position. Some preliminary texture analysis can be applied to aid the grid points to adapt to regions with more textures. However, in cases where large texture-less areas dominate one part of the entire image, area-based template matching will totally fail. Under such extreme circumstances, even human visual interpretation would be difficult, thus, setting up some man-made targets in the field before the image is taken might be helpful.

## Conclusions

In this paper, we proposed an algorithm for the automatic registration of two airborne mosaicked frame images with complex local deformation by combining area-based control point extraction methods with local geometric transformation models. Two similarity measures (CC and MI) were used and compared for the control point extraction. The area-based methods generated large amount of evenly distributed control points on regular grids. The control points with outliers removed were then applied to both global transformation

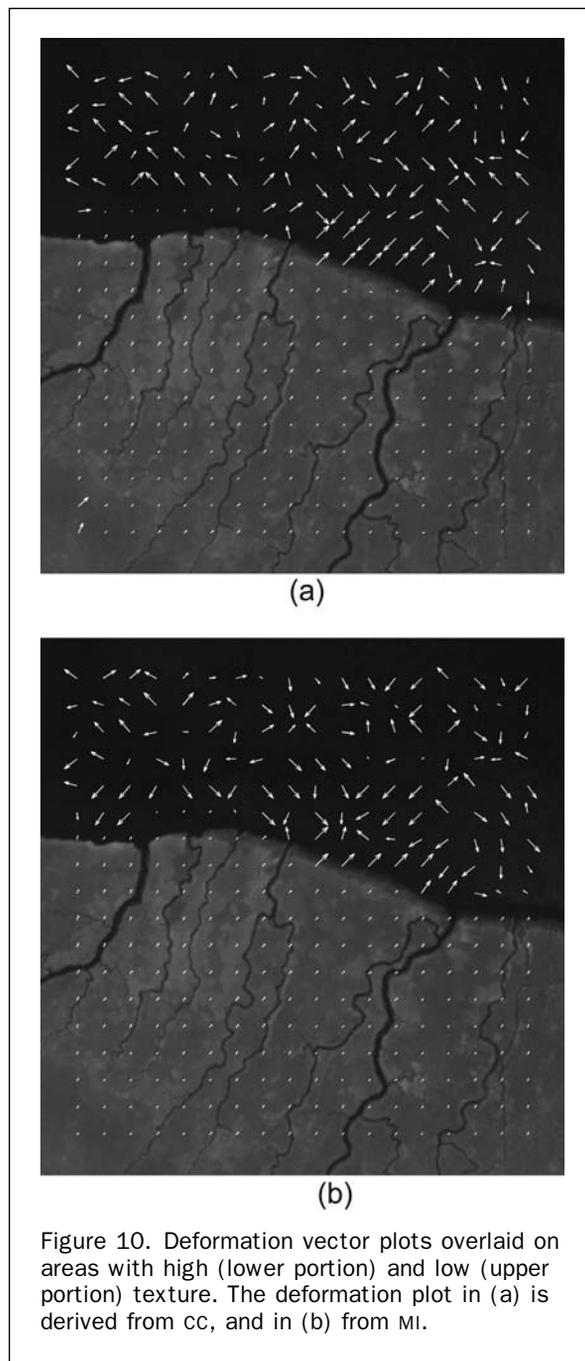


Figure 10. Deformation vector plots overlaid on areas with high (lower portion) and low (upper portion) texture. The deformation plot in (a) is derived from CC, and in (b) from MI.

models and local transformation models. Accuracy assessments showed that the local transformation models outperformed global transformation models in terms of total RMSE, standard deviation of residuals, and maximum x-residuals and y-residuals. For global models, the 3<sup>rd</sup> order polynomial outperformed the 2<sup>nd</sup> order polynomial. For local models, the piecewise linear model performed slightly better than the local weighted mean model. The difference in all the registration error statistics between CC and MI was slight, which may be due to the similar radiometric properties between two images. CC works well for the same type of images, whereas MI works better when multi-modal images are used.

To test the effect of the number of control points on transformation models, 10 sets of control points on grids of different density were generated by using area-based

methods based on both CC and MI. The registration results showed that more control points could dramatically reduce the registration errors for local transformation models whereas more control points did not improve the global transformation models. This indicates that: (a) local transformation models are appropriate and sufficient to register imagery with complex local distortion, and (b) more evenly distributed control points can better model the local distortion when local transformation models are used.

In summary, the combination of area-based control point extraction with local transformation models is more appropriate for the geometric registration of airborne images with complex local distortion. Specifically, we reveal two important findings here: first, the area-based methods generated larger amounts of evenly distributed control points compared to feature-based methods and manual methods; and second, local transformation models achieved better registration accuracy when larger amount of evenly distributed control points are used. These results are important for change detection research involving multiple sets of remotely sensed imagery; particularly for multi-temporal remote sensing research using high-spatial resolution imagery acquired from several aircraft missions.

### Acknowledgments

This work was supported by NASA Headquarters under the Earth System Science Fellowship Grant NGT5-30493 to D. Liu, and a NASA New Investigator Program award to M. Kelly. The authors would like to thank three anonymous reviewers who helped strengthen the paper.

### References

- Brown, L.G., 1992. A survey of image registration techniques, *ACM Computing Surveys*, 24(4):325–376.
- Chen, H.M., M.K. Arora, and P.K. Varshney, 2003. Mutual information-based image registration for remote sensing data, *International Journal of Remote Sensing*, 24(18):3701–3706.
- Chen, H.M., P.K. Varshney, and M.K. Arora, 2003. Performance of mutual information similarity measure for registration of multitemporal remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, 41(11):2445–2454.
- Cole-Rhodes, A.A., K.L. Johnson, J.L. Moigne, and I. Zavorin, 2003. Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient, *IEEE Transactions on Image Processing*, 12(12):1495–1511.
- Collignon, A., F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, 1995. Automated multi-modality image registration based on information theory, *Information Processing in Medical Imaging*, Norwell, Massachusetts, Kluwer, 263–274.
- Coulter, L.L., D.A. Stow, and S. Baer, 2003. A frame center matching technique for precise registration of multitemporal airborne frame imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 41(11):2436–2444.
- Dai, X., and S. Khorram, 1998. The effects of image misregistration on the accuracy of remotely sensed change detection, *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1566–1577.
- Devereux, B.J., R.M. Fuller, L. Carter, and R.J. Parsell, 1990. Geometric correction of airborne scanner imagery by matching Delaunay triangles, *International Journal of Remote Sensing*, 11(12):2237–2251.
- Flusser, J., 1992. An adaptive method for image registration, *Pattern Recognition*, 25(1):45–54.
- Fonseca, L.M.G., and B.S. Manjunath, 1996. Registration techniques for multisensor remotely sensed imagery, *Photogrammetric Engineering & Remote Sensing*, 62(9):1049–1056.
- Gong, P., E.F. Ledrew, and J.R. Miller, 1992. Registration-noise reduction in difference images for change detection, *International Journal of Remote Sensing*, 13(4):773–779.
- Goshtasby, A., 1986. Piecewise linear mapping functions for image registration, *Pattern Recognition*, 19(6):459–466.
- Goshtasby, A., 1987. Piecewise cubic mapping functions for image registration, *Pattern Recognition*, 20(5):525–533.
- Goshtasby, A., 1988a. Registration of images with geometric distortions, *IEEE Transactions on Geoscience and Remote Sensing*, 26(1):60–64.
- Goshtasby, A., 1988b. Image registration by local approximation methods, *Image and Vision Computing*, 6(4):255–261.
- Inglada, J., and A. Giros, 2004. On the possibility of automatic multisensor image registration, *IEEE Transactions on Geoscience and Remote Sensing*, 42(10):2104–2120.
- Jensen, J.R., 2004. *Introductory Digital Image Processing*, 3<sup>rd</sup> edition, Upper Saddle River, New Jersey, Prentice Hall.
- Kelly, M., D. Shaari, Q. Guo, and D. Liu, 2004. A comparison of standard and hybrid classifier methods for mapping hardwood mortality in areas affected by sudden oak death, *Photogrammetric Engineering & Remote Sensing*, 70(11):1229–1239.
- Kelly, M., and R. Meentemeyer, 2002. Landscape dynamics of the spread of Sudden Oak Death, *Photogrammetric Engineering & Remote Sensing*, 68(10):1001–1009.
- Kennedy, R.E., and W.B. Cohen, 2003. Automated designation of tie-points for image-to-image coregistration, *International Journal of Remote Sensing*, 24(17):3467–3490.
- Knoll, K.F. and E.J. Delp, 1986. Adaptive gray scale mapping to reduce registration noise in difference images, *Computer Vision, Graphics and Image Processing*, 33:129–137.
- Moigne, J.L., W.J. Campbell, and R.F. Crompt, 2002. An automated parallel image registration technique based on the correlation of wavelet features, *IEEE Transactions on Geoscience and Remote Sensing*, 40(8):1849–1864.
- Richards, J.A., and X. Jia, 1999. *Remote Sensing Digital Image Analysis: An Introduction*, 3<sup>rd</sup> edition, New York, Springer-Verlag.
- Roche, A., G. Malandain, and N. Ayache, 2000. Unifying maximum likelihood approaches in medical image registration, *International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging*, 11(1):71–80.
- Schowengerdt, R.A., 1997. *Remote Sensing: Models and Methods for Image Processing*, 2<sup>nd</sup> edition, San Diego, Academic Press.
- Shekhar, S., C. Lu, and P. Zhang, 2003. A unified approach to detecting spatial outliers, *Geoinformatica*, 7(2):139–166.
- Stow, D., L. Coulter, and S. Baer, 2003. A frame center matching approach to registration for change detection with fine spatial resolution multi-temporal imagery, *International Journal of Remote Sensing*, 24(19):3873–3879.
- Townshend, J.R.G., C.O. Justice, C. Gurney, and J. McManus, 1992. The impact of misregistration on change detection, *IEEE Transactions on Geoscience and Remote Sensing*, 30(5):1054–1060.
- Viola, P., and W.M. Wells III, 1997. Alignment by maximization of mutual information, *International Journal of Computer Vision*, 24(2):137–154.
- Zitova, B., and J. Flusser, 2003. Image registration methods: A survey, *Image and Vision Computing*, 21(11):977–1000.

(Received 05 May 2005; accepted 22 July 2005; revised 15 August 2005)