Precision Registration and Mosaicking of Multicamera Images

David J. Holtkamp and A. Ardeshir Goshtasby

***Abstract*—A method for registering and mosaicking multicamera images is described. Registration is achieved using control points and projective transformation, paying special attention to factors that contribute to the precision registration of images. Among the corresponding control points found in overlapping images, those that best satisfy the projective constraint are used to register the images. Experimental results show that a small number of correspondences that satisfy the projective constraint produce a more accurate registration than a large number of correspondences in least squares fashion. To achieve seamless mosaics, the intensities of all images captured by the cameras are transformed to the intensities of one of the images to minimize the intensity difference between the registered images. Mosaicking results on 4096 image sets acquired by a six-camera system are presented and discussed.**

***Index Terms*—Aerial image, image mosaicking, image registration, multicamera images, projective constraint.**

I. Introduction

IMAGE registration is the process of spatially aligning two or more images of a scene. This basic operation is often needed in remote sensing applications that involve two or more images of an area. Image registration accuracy profoundly affects the accuracy of applications that use it. Townshend *et al.* [44] and Bovolo *et al.* [5] have shown the influence of image registration accuracy on change detection accuracy, while Núñez *et al.* [31] and Aanæs *et al.* [1] have demonstrated the significance of image registration accuracy on image fusion accuracy.

Image mosaicking is the process of smoothly piecing together overlapping images of a scene into a larger image. This operation is needed to increase the area of coverage of an image without sacrificing its resolution. Due to the limited size of digital images, it is sometimes not possible to include an area of interest in an image. In such a situation, overlapping images are obtained, and the images are combined into a larger image through image mosaicking. An image mosaic is created from a set of overlapping images by registering and resampling all images to the coordinate space of one of the images. One of the oldest applications of image mosaicking is in map building [41].

An image mosaicking system has to take into consideration the relation between the cameras, distances of the cameras to the scene, the scene content, and the characteristics of the cameras. When the cameras are positioned close to each other in a camera rig and the distance of the camera rig to the scene is much larger than the changes in scene elevation, local geometric differences between overlapping images will be negligible, although global geometric differences may still exist between them. Images captured by a camera rig mounted on a high-flying aircraft represent such an example. If the changes in scene elevation are not negligible compared to the distance of the camera rig to the scene, overlapping images may have local geometric differences, and a transformation model that can represent local geometric differences between overlapping images is needed to register the images. Images of an indoor scene captured by an array of cameras represent such an example.

In the following sections, related work in image registration and image mosaicking is reviewed, details of an automatic mosaicking system are provided, and experimental results using 4096 image sets captured by a six-camera system are presented and discussed.

II. Background

Image registration is one of the most studied problems in image processing, image analysis, remote sensing, and medical imaging. Early methods registered images that had only translational differences. Anuta [2] used a fast Fourier transform algorithm as a means to carry out image correlation efficiently and register images with translational differences. Kuglin and Hines [22] separated phase from magnitude in the Fourier domain of the correlation matrix, and by using only the phase component of the transform enhanced the correlation peak and made the process more robust. Reddy and Chatterji [35] extended the phase correlation idea of Kuglin and Hines to registration of images with translational, rotational, and scaling differences. Methods that register images with more complex transformation functions have been proposed also. Flusser and Suk [12] developed a method for registering images by the affine transformation, Goshtasby [14] and Arévalo and González [3] used a combination of affine transformations, and Goshtasby [15] used surface splines to register images with nonlinear geometric differences.

Image registration methods can be categorized into intensity- and feature-based methods. Intensity-based methods use raw pixel intensities to register images. Feature-based methods, on the other hand, first process image intensities to identify unique landmarks or control points in images and then use corresponding control points in the images to register them.

Manuscript received November 26, 2008; revised February 7, 2009 and April 3, 2009. First published July 14, 2009; current version published September 29, 2009. This work was supported by the Air Force Research Laboratory under Grant FA8650-05-1-1914.

The authors are with the Department of Computer Science and Engineering, Wright State University, Dayton, OH 45435 USA (e-mail: agoshtas@wright.edu).

Digital Object Identifier 10.1109/TGRS.2009.2023114

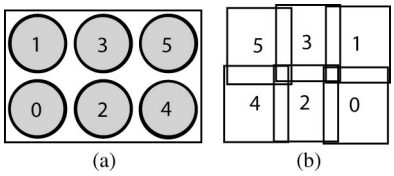

Fig. 1. (a) Front view of a six-camera system. (b) Overlapping images simultaneously captured by the cameras.

Excellent reviews of image registration methods are provided in [6], [17], [45], and [46]. Our method is feature-based and uses corresponding control points in images to register them.

Image mosaicking also has a long history. Earlier methods found the best seam between overlapping images, so that the images could be cut at the seam and pasted together to create a seamless mosaic [10], [21], [26], [27], [39]. When the camera parameters are known, the images can be standardized geometrically and radiometrically to create high-quality mosaics. When the camera parameters are not known, elaborate methods are needed to align overlapping images and seamlessly combine their intensities. Such methods typically convert the coordinate spaces of all images to that of one of the images through image registration. Then, from the correspondence between pixels in overlapping images, the intensities of the images are converted to those of one of the images to minimize the intensity difference between the overlapping images. The intensities of the registered images are then combined to create seamless mosaics.

Methods for creating mosaics from a sequence of images obtained by a moving camera [7], [20], [32], [33], [36] and by a camera rotating about an axis [25], [28], [40] have been developed. Two-stage algorithms that first align the images locally and then correct for registration errors through global alignment have been proposed [9], [24], [40]. Successful mosaicking of a large number of overlapping images using control points has been reported [7], [20]. Szeliski [43] has reviewed and classified existing mosaicking methods. Although the focus of past mosaicking methods has not been on precision registration of the images, the main focus of this paper will be on identifying ways to register the images as accurately as possible. A method to avoid registration failure and improve registration reliability is also described.

III. Problem Description and Assumptions

Overlapping images are captured by an array of cameras, as shown in Fig. 1(a). A triggering mechanism synchronizes the cameras so that they all capture the images simultaneously [Fig. 1(b)] at a rate of four frames per second. The camera rig, which is mounted on an unmanned aerial vehicle, takes the images during a flight. We would like to combine a set of simultaneously captured images into a seamless mosaic.

The following are assumed: 1) All cameras are of the same make and model; therefore, they all have the same optical characteristics; 2) image distortion due to lens distortion does not exist or is negligible; and 3) the cameras have the same internal parameters. That is, the focal lengths and zoom levels of all cameras are the same, and all pixels in the images are square and are of the same size.

Deviations from these assumptions may exist when working with two or more cameras. For instance, although efforts are made to set all cameras to the same zoom level, small differences may exist in scales of overlapping images. Moreover, a scene point may have different intensities in two images due to a difference in the gains of the cameras and the presence of image noise. We will therefore build into our mosaicking system steps that will compensate for small deviations from these assumptions.

Since the camera rig is relatively far from the scene and the images are captured from the same viewpoint, scene elevation will not play a role in the registration process. This is because two cameras see the same scene points in their overlap area. However, since images are obtained by cameras with slight view-angle differences, adjacent images represent projections of the scene onto two planes with a small orientation difference. Geometric distortions due to lens nonlinearities in an image can be corrected ahead of the time by a process known as image decalibration [16]. Therefore, in the following discussions, we assume that geometric distortions due to lens nonlinearities do not exist or are negligible in the images.

IV. Approach

Knowing that two overlapping images are related by the projective transformation, we first determine the parameters of the transformation using a number of corresponding control points in the images. One of the images will be used as the base, and the geometry of the second image will be transformed to that of the base image. Assuming $\mathbf{p} = (x, y)$ represents a point in the base image and $\mathbf{P} = (X, Y)$ represents the same point in the image overlapping the base image, the relation between corresponding points in the images in the homogeneous coordinate system can be written as

$$\begin{bmatrix} XW \\ YW \\ W \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{1}$$

or

$$X = \frac{ax + by + c}{gx + hy + 1} \tag{2}$$

$$Y = \frac{dx + ey + f}{gx + hy + 1}. \tag{3}$$

The 3×3 matrix with eight parameters in (1) is the projective transformation with components shown in (2) and (3). Having the coordinates of four corresponding points in the images, the eight unknown parameters $a-h$ of the transformation can be determined by substituting the coordinates of the four corresponding points into (2) and (3) and solving the obtained system of eight linear equations.

Images adjacent to the base image can be registered to the base image by the projective transformation. Since the images adjacent to the base image after registration will be in the coordinate space of the base image, the images can be combined

into a mosaic. We have chosen camera 2 as the base camera because its view angle has been closer to the nadir view than the view angles of other cameras. This choice will create mosaics with the least distortion when compared to a map.

A. Control Point Selection

The determination of the parameters of a projective transformation for registration of two overlapping images requires a minimum of four corresponding points in the images. We first select a number of control points in each image and then determine the correspondence between them.

Various methods for detecting control points in an image have been developed. Schmid *et al.* [37] have surveyed and compared various point detectors, finding the Harris detector [19] to be the most repeatable. A study [13] comparing ten different detectors has found that the repeatability of the Harris corner detector and that of scale-invariant feature transform (SIFT) detector [23], which finds blobs in an image, stand out among others, with the Harris detector performing slightly better than the SIFT detector.

Two adjacent cameras may have slightly different focal lengths, zoom levels, gains, and view angles. When two images are obtained at different focal lengths, one image will be blurred with respect to the other. When two cameras have different zoom levels, the obtained images will have different scales. When the gains of two cameras are different, the obtained images will have intensity differences. When the images are captured by cameras with different view angles, the images will have projective differences.

To detect control points that are stable under small differences in gain, zoom level, and view angle of the cameras, we choose the Harris detector [19] because of its high repeatability [13], [37] under radiometric and geometric changes. We choose the Harris detector over the SIFT detector [23] also because of the abundance of corners compared to blobs in urban scene images. To ensure that the same control points are obtained in images from cameras with slightly different focal lengths, we find control points at three different resolutions and keep only those that persist across the three resolutions.

To create three resolutions of an image, the image is smoothed with Gaussians of standard deviations $\sigma - \Delta\sigma$, σ, and $\sigma + \Delta\sigma$. A control point detected at resolution σ is considered stable if it does not move by more than $2\Delta\sigma$ at resolutions $\sigma - \Delta\sigma$ and $\sigma + \Delta\sigma$. Bergholm [4] has shown that most pixels in an image move less than a pixel when the standard deviation of the Gaussian smoother is changed by half a pixel. If a control point moves by more than a pixel when changing image resolution by half a pixel, the control point is discarded; otherwise, it is kept. Digital error and random noise can displace control points in images. This multiresolution processing limits the positional displacement of a control point under a slight change in resolution to a pixel, which can still be large when considering precision registration of very large images. Among all the obtained correspondences, we will show how to identify the most accurate ones using the projective constraint.

An example of control points detected by the multiresolution Harris detector is shown in Fig. 2. Image Fig. 2(a) depicts the control points detected by the original Harris detector [19], while image Fig. 2(b) shows the control points obtained by the multiresolution Harris detector as outlined earlier. Those Harris points that do not displace by more than a pixel as the standard deviation of the Gaussian smoother is increased from 1.5 to 2 pixels and decreased from 1.5 to 1 pixel are kept, and the remaining points are discarded. The lower right quadrants of Fig. 2(a) and (b) are shown again in Fig. 2(c) and (d) for a closer comparison of the original Harris points and the multiresolution Harris points.

Since only four accurate correspondences are sufficient to register overlapping images by the projective transformation, there is no need to detect thousands or even hundreds of control points in the images. About 50 to 100 control points are sufficient to register the images. We will therefore enter into a list the control points detected in an image in the descending order of their strengths, as determined by the Harris detector [19] at resolution σ, and select the top 50 to 100 control points that are widely dispersed over the image domain. It is anticipated that, from among the 50 to 100 control points, about a dozen or two will fall in the overlap area between adjacent images. Once a control point is selected, all control points below it in the list, which are within a distance threshold of it, are removed from the list. For the images used in this paper, a distance threshold equal to 1/64 of the diagonal length of the image was used to locate between 50 and 100 widely dispersed control points in an image.

B. Control Point Correspondence

The problem of finding the correspondence between two sets of points is known as point pattern matching and has been studied extensively. The process involves removing outliers and determining the transformation that relates the inliers. RANdom SAmple Consensus (RANSAC) [11], clustering [42], and coherence matching [18] are some of the widely used techniques. RANSAC uses a distance tolerance to find correspondence between two point sets under a given transformation function. If the tolerance is too low, the process may miss finding the correspondences, and if the tolerance is too high, some of the correspondences may be inaccurate or incorrect. It is possible for RANSAC to find the correspondences at the first try, depending on the random subsets chosen from the two sets. Therefore, it has a very low lower bound computational complexity. However, it may not find the correspondences until after a very large number of tries; thus, its upper bound computational complexity is very high.

Coherence matching is similar to RANSAC except that, instead of taking random points from the two sets, the points are selected in a structured manner to reduce the number of four-point combinations needed to find the transformation parameters. To match two point sets that are related by the projective transformation, the convex hulls or the minimum-spanning trees (MSTs) [34] of the point sets are obtained, and points representing disjoint edge pairs in the convex hull or the MST of each set are used to find the transformation parameters. If the two sets contain M and N points, the number of possible combinations of convex hull or MST edge pairs is

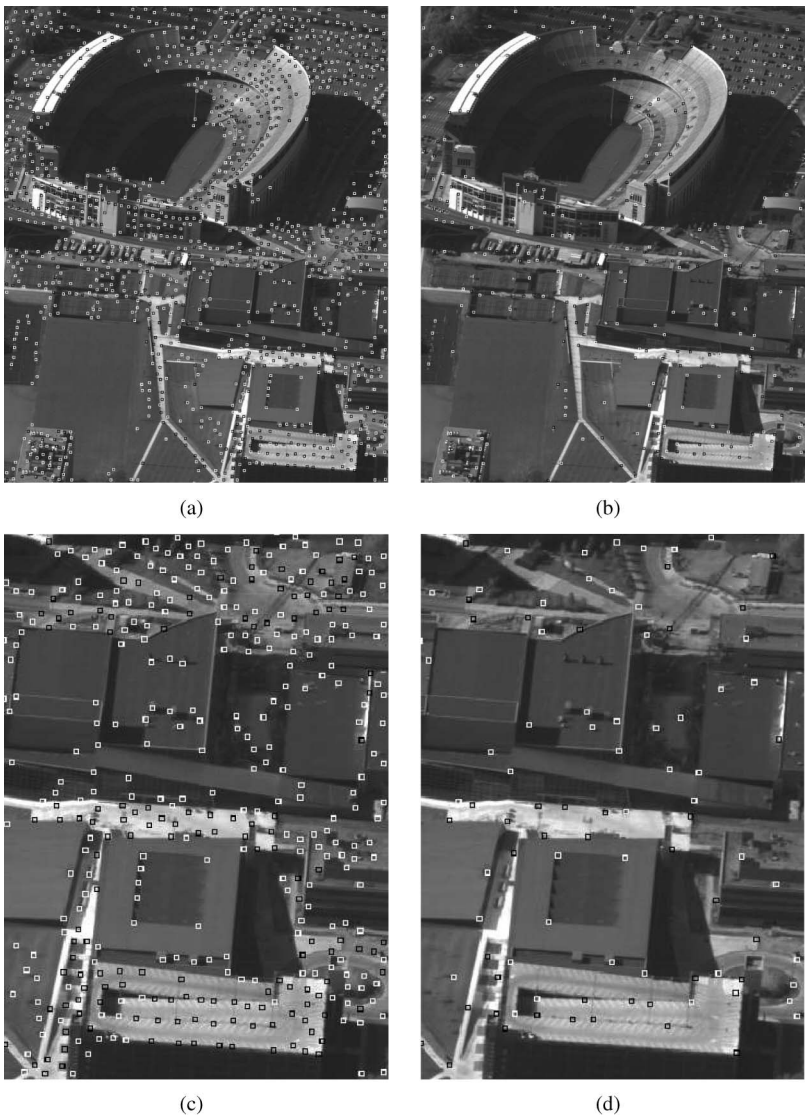

Fig. 2. (a) One thousand three hundred ninety-five Harris points detected in an aerial image when $\sigma = 1.5$ pixels. (b) Two hundred ninety-eight multiresolution Harris points found in the same image when $\sigma = 1, 1.5$, and 2 pixels. (c)–(d) Lower right quadrants of (a) and (b), respectively.

on the order of M^2N^2. This is considerably lower than the number of possible four-point combinations taken from the two point sets, which is on order of M^4N^4. The coherence matching algorithm can therefore find the correspondences much faster than RANSAC. Although rare, there is a possibility that, depending on the positioning of outliers with respect to inliers in the two point sets, there will not be sufficient common edge pairs in the convex hulls or the MSTs of the two point sets to find the correspondences. In such rare cases, the correspondences may be found using the four-point combinations by RANSAC.

Clustering has a fixed computational complexity. It requires a sufficient number of four-point combinations from the two sets to produce robust clusters in the parameter space. Here, also by selecting the points in a more structured manner, the clusters can be formed more quickly.

In this paper, we choose coherence matching using MST edges to find the correspondence between control points in overlapping images. More specifically, we 1) find MST edges of control points in each image; 2) match four points representing two disjoint MST edges in one image with four points representing two disjoint MST edges in another image, determining the eight unknown parameters of the projective transformation; 3) find the number of other control points that also match with the obtained transformation; and 4) determine the transformation that matches the most control points in the two images. In step 3), points (X, Y) and (x, y) in two images are considered corresponding if

$$\|(X, Y) - (f_x(x, y), f_y(x, y)\| < \epsilon \tag{4}$$

where f_x and f_y are the components of the projective transformation given by (2) and (3), and ϵ is a small distance tolerance. When $\epsilon = 1$ pixel, all corresponding control points will fall within a pixel of each other after the images are registered by the obtained transformation.

We further reduce computation time in coherence matching as follows: 1) by using only the long MST edges (for example,

(a) (b) (c)

Fig. 3. (a)–(b) Images captured by cameras 2 and 0, respectively. (c) Intensities of the image from camera 0 are modified so that the histogram calculated for the area overlapping the image from camera 2 becomes similar to the histogram of the image from camera 2 calculated for the same overlap area.

using the longest 30 edges); 2) by carrying out matching of edge pairs from the two images if the ratio of the shorter edge over the longer one is greater than a prespecified value (such as 0.25); 3) by using edge pairs in images that have similar angles between them (for example, their difference being less than 10°); 4) by using edges with similar orientations (for instance, being within 10° of each other); and 5) by stopping the process when, by matching a pair of MST edges in two images, more than half of the control points in the images match.

The projective transformation obtained in this manner will resample an image overlapping the base image to the coordinate space of the base image. Since the images may have been obtained by cameras with slightly different gains, the images may show different intensities in their overlap area. In order to combine the registered images seamlessly, the intensities of an image overlapping the base image should be transformed to that of the base image. This is achieved by histogram modification [30], converting the intensities of the image overlapping the base image so that its histogram in the overlap area becomes similar to that of the base image. An example of intensity mapping by histogram modification is shown in Fig. 3. Two overlapping images are shown in Fig. 3(a) and (b), and Fig. 3(b), after histogram modification, is shown in Fig. 3(c). Fig. 3(a) and (c) now have similar intensities in their overlap area and, when combined, will produce a seamless mosaic.

C. Projective Constraint

The control points detected by the Harris detector [19] have discrete coordinates. Through interpolation of measures produced by the detector, the coordinates of the control points are determined with subpixel accuracy. Methods that use centers of gravity of regions also produce control points with subpixel positional accuracy [14], [38]. Due to the presence of noise, some control points displace even when their positions are determined with subpixel accuracy. As a result, some corresponding points will match more accurately than others. We would like to develop a methodology that can distinguish more accurate correspondences from the less accurate ones.

Using all the corresponding control points in the images to register the images by the least squares method will not produce as accurate a registration than when using a smaller number but more accurate correspondences. This is demonstrated in a simple example in Fig. 4. Fig. 4(a) is transformed by a known projective transformation to obtain Fig. 4(b). Therefore, the coordinates of exact correspondences in the images are known. The control points detected by the Harris detector [19] and the correspondences determined by coherence matching [17] are depicted within the images in Fig. 4(a) and (b). All correspondences are visually verified to be correct. If the images are registered using all the correspondences by the least squares method, a root-mean-squared intensity difference (RMSID) between registered images is 5.821, while RMSID between images registered using the best four correspondences obtained by the projective constraint is 5.674. The difference between the two is small, because most areas in the images are either homogeneous or gradually varying in intensities and also because the images do not contain noise. Nevertheless, the registration result obtained using the projective constraint is better than that obtained without it. Fig. 4(c) and (d) shows the absolute intensity difference of corresponding pixels in the registered images without and with the projective constraint, respectively.

A second example using images of an urban scene is given in Fig. 5. Fig. 5(a) and (b) shows the overlap area between the images from cameras 1 and 2 taken simultaneously. The point correspondences are also shown in these images. Again, the correspondences were visually verified to be correct. Registration result by the least squares method using all correspondences produced RMSID that is equal to 14.036. The RMSID obtained using the best four correspondences identified by the projective constraint was 9.996. The subtracted registered images using all the correspondences and the best four correspondences are shown in Fig. 5(c) and (d), respectively. The difference between the two methods using these real images is significant. The difference between the least squares method using all correspondences and the four best correspondences without least squares varies from image to image, but in the

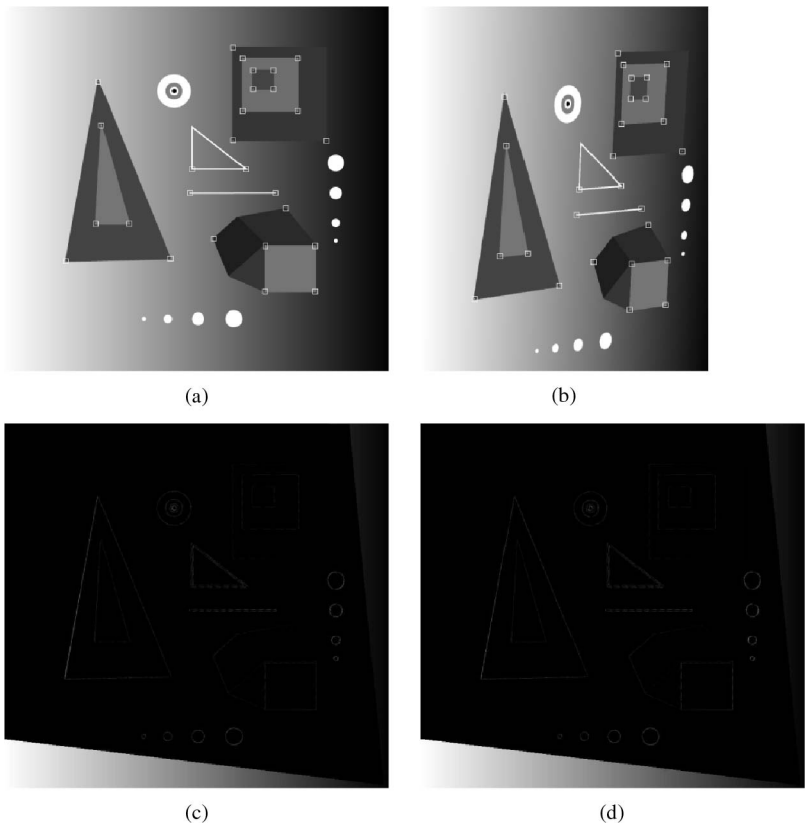

Fig. 4. (a)–(b) Two synthetically generated images with a known projective transformation. The corresponding control points in the images are also shown. These images do not contain noise and do not have local geometric differences. (c) Registration of the images using all control points and the least squares method. (d) Registration of the images using the four correspondences best satisfying the projective constraint. Shown are the absolute intensity differences of the corresponding pixels in the registered images. Registration results are close, although the one produced by the proposed method is slightly better.

images tested in this paper, the projective constraint consistently produced more accurate registration results than the least squares without projective constraint.

Some image properties do not change under projective transformation [29]. For instance, straight lines in one image map to straight lines in another image. Certain algebraic measures defined in terms of the cross-ratio of four colinear points, five coplanar points, and two lines and two points have been found to remain invariant under projective transformation [8]. Since we have corresponding points in two image planes, we will use the invariant of five coplanar points to distinguish accurate correspondences from inaccurate ones.

The five coplanar projective invariant is defined as follows [8]. Given points $\{(x_i, y_i) : i = 1, \ldots, 5\}$ in an image, the following two algebraic measures:

$$I_1(x,y) = \frac{\det[\mathbf{m}_{431}]\det[\mathbf{m}_{521}]}{\det[\mathbf{m}_{421}]\det[\mathbf{m}_{531}]} \tag{5}$$

$$I_2(x,y) = \frac{\det[\mathbf{m}_{421}]\det[\mathbf{m}_{532}]}{\det[\mathbf{m}_{432}]\det[\mathbf{m}_{521}]} \tag{6}$$

where

$$\det[\mathbf{m}_{123}] = \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{vmatrix} \tag{7}$$

remain invariant under the projective transformation [8]. Therefore, if images with coordinates (x, y) and (X, Y) are related by the projective transformation and control point (X_i, Y_i) corresponds to control point (x_i, y_i) for $i = 1, \ldots, 5$, then, by replacing (x_i, y_i) with (X_i, Y_i) in (5)–(7), we will obtain $I_1(x, y) = I_1(X, Y)$ and $I_2(x, y) = I_2(X, Y)$. If measures (I_1, I_2) from two images are not exactly the same, the smaller their distance

$$D = \sqrt{[I_1(x,y) - I_1(X,Y)]^2 + [I_2(x,y) - I_2(X,Y)]^2} \tag{8}$$

is, the higher the accuracy of the five correspondences will be. Note that, by using the five points in different order, other invariants will be obtained. Since our objective is to reject the inaccurate correspondences, I_1 and I_2 calculated with any ordering of the five points can tell whether the five points under consideration in the images correspond to each other or not. The likelihood that five points that do not correspond to each other displace accidentally by the right amounts in the images to satisfy the projective constraint is extremely rare. Therefore, the smaller the value for D in (8), the more accurate the five correspondences are expected to be.

Assuming that n corresponding control points are produced by coherence matching, distance D in (8) is calculated for the combinations of five corresponding control points out of n, and the combination producing the smallest distance is

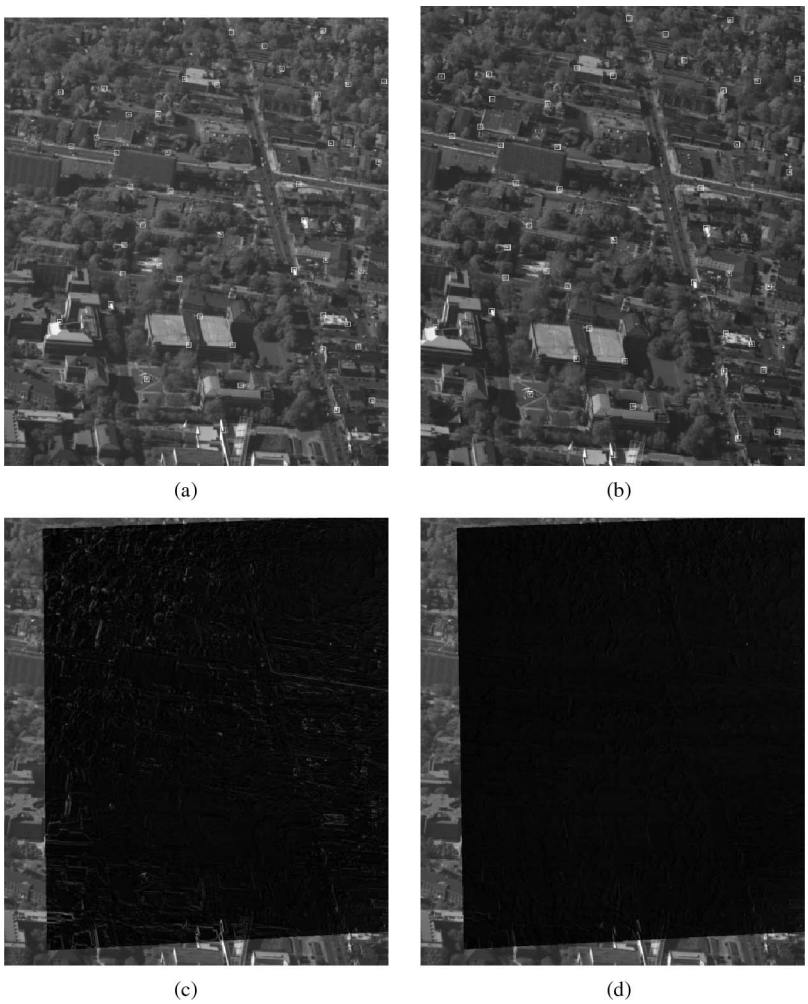

Fig. 5. (a)–(b) Two simultaneously captured images of an urban scene representing the overlap area between images from cameras 1 and 2. These images contain noise and have small local geometric differences due to a slight view-angle difference between the cameras. Registration of the images (c) using all the correspondences and (d) using the best four correspondences. Shown are the absolute intensity differences of the corresponding pixels in the registered images. Registration results by the two methods are visibly different.

chosen as the best combination. Since only four correspondences are sufficient to determine the parameters of the projective transformation, among the five correspondences, the four producing the smallest RMSID are used to register the images.

D. Robust Registration

Registration accuracy depends on image content. As the camera rig moves, contents of images change, changing the registration accuracy. Knowing that the cameras are fixed with respect to each other and parameters such as focal lengths and zoom levels of the cameras do not change during image acquisition, once the transformations to register a set of images are obtained, they can be used to register other image sets captured by the camera rig.

Because registration accuracy depends on image content, rather than finding the transformations from only one set of images, different image sets are used to find the best set of transformations for the registration of the images. The quality of registration is measured using the RMSID between overlapping images after intensity standardization, as shown in Fig. 3. Finding the transformation parameters using more than one set of images also makes the mosaicking process more reliable. When more than one set of images is used, even when the registration of some sets fails, there will be registration of some sets that succeed, and among the successful registrations, the parameters of the one that produces the smallest RMSID are chosen to register all image sets.

In this paper, 4096 image sets sequentially captured by a six-camera rig were used. Transformations for the registration of overlapping images were determined using ten image sets that were uniformly spaced within the 4096 sequence. Sets $400i$ were chosen by changing i from one to ten. The mosaicking process involves five registration steps: 1) registering the image from camera 1 to the image from camera 3 to create mosaic A; 2) registering the image from camera 5 to mosaic A to create new mosaic B; 3) registering the image from camera 0 to the image from camera 2 to create mosaic C; 4) registering the image from camera 4 to mosaic C to create new mosaic D; and 5) registering mosaic B to mosaic D to create the final mosaic. The mosaicking process is shown in Fig. 6.

The use of multiple image sets to determine the best transformation parameters will not only improve the mosaicking

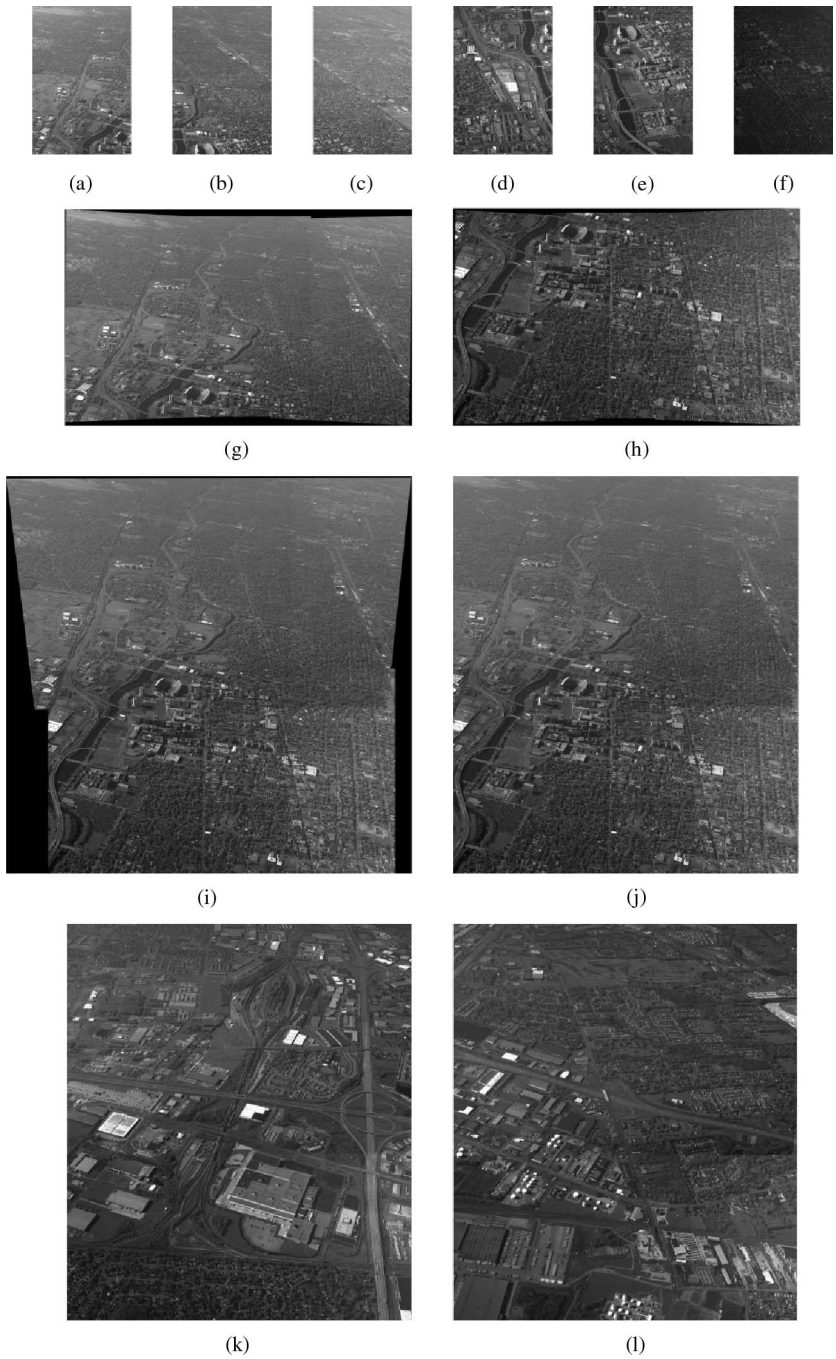

Fig. 6. (a)–(f) Images simultaneously captured by cameras 5, 3, 1, 4, 2, and 0, respectively, using the camera rig shown in Fig. 1. These images are of size 4008 × 2760 pixels. (g) Registering and mosaicking images 1, 3, and 5. (h) Registering and mosaicking images 0, 2, and 4. (i) Registering and mosaicking images (g) and (h). (j) Largest rectangular area contained in image (i). This is the final mosaicking result. (k)–(l) Two more mosaicking examples created using two other sets of images showing different scenes. The created mosaics are of size 7646 × 6628 pixels.

accuracy but will also improve the mosaicking reliability. Depending on the scene content, it may not always be possible to find a sufficient number of common control points in overlapping images to register them or to register them accurately. Repeating registration using images that contain different scenes will make it possible to find transformations that register overlapping images more accurately. As the number of image sets increases, the likelihood that registration of all image sets fails decreases exponentially. If there is only a 50% chance that the registration of a set of images succeeds, when the process is repeated on ten image sets, the success rate will increase to $100(1-(0.5)^{10})\%$, which is better than 99.9%.

V. Results and Discussion

A mosaicking example is shown in Fig. 6. First, images (a) and (c) are registered to image (b) to create mosaic (g). Then, images (d) and (f) are registered to image (e) to create mosaic (h). Finally, mosaic (g) is registered to mosaic (h) to create the overall mosaic (i). The largest rectangular area within the

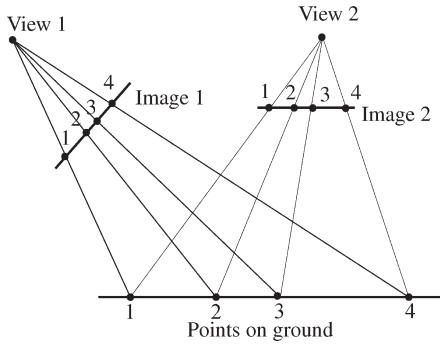

Fig. 7. Use of cross-ratio invariance in determining the geometric fidelity of a created mosaic compared to a Google map.

obtained mosaic as shown by (j) is used as the final result. Two more mosaicking examples of different scenes are shown in (k) and (l).

Ten sets of images were used to determine the transformation parameters that relate the coordinates of all images to those of the base image. Among the ten sets, failure was reported in one set due to an insufficient number of accurate corresponding control points in the overlap area between two of the images. From among the nine remaining sets, the set of transformations producing the smallest RMSID was chosen to register all 4096 image sets.

The determination of the transformation functions that register six overlapping images, each of size 4008×2760 pixels, takes about 15 min on a 3.2-GHz computer. After all transformations are found, mosaicking of each image set takes about 30 s on the same computer.

The geometric fidelity of the created mosaics was evaluated using Google maps as the gold standard. Since the images available to this paper were obtained at off-nadir views and Google maps represent the nadir view of the ground, rather than comparing actual distances between the mosaics and the Google maps, the cross-ratio of distances on straight lines [29] was used. A created mosaic has a correct image geometry if the same cross-ratio is obtained for four corresponding points on corresponding lines in the mosaic and in a map of the area.

The concept of cross-ratio invariance is shown in Fig. 7. If two images of a flat area are obtained from different views, the following cross-ratio of distances remains invariant between the images [8]:

$$CR = \frac{\overline{\mathbf{12}} \times \overline{\mathbf{34}}}{\overline{\mathbf{13}} \times \overline{\mathbf{24}}} \tag{9}$$

where $\overline{\mathbf{ab}}$ denotes the distance between points $\mathbf{a}$ and $\mathbf{b}$ in an image.

To determine the geometric fidelity of a created mosaic, a line is drawn in the Google map, and four ground points lying on the line are identified. The same points are identified in the mosaic. From four points in the map and three of the points in the mosaic, the location of the fourth point in the mosaic is calculated using the cross-ratio invariance given by (9). The distance between the estimated fourth point to the actual fourth point in the mosaic is used to quantify error between the map and the mosaic. Ten mosaics representing different scenes were used, and ten lines were drawn in each mosaic, making sure that the area inside where a line is drawn is flat and the selected points lie on the ground. The average distance between the estimated and actual points was found to be about three pixels, while the standard deviation of the distances was about one pixel. Error was noticed to increase away from the base image and toward mosaic borders. Errors as high as seven pixels were observed between the created mosaics and the corresponding Google maps.

VI. Concluding Remarks

The design of a computational method for creating image mosaics from multicamera images has been described. The primary focus of the design was on the precision registration of the images. Although subpixel accuracy was achieved in the registration of image pairs, it was not possible to achieve similar accuracy when creating image mosaics.

When a mosaic of a large area is created and the images represent off-nadir views of the area, the created mosaic will represent the projection of the globe to a plane that is not horizontal. Such a mosaic will be geometrically distorted with respect to a map that represents the projection of the globe to a horizontal plane. To improve the geometric fidelity of a mosaic, each image should be first orthorectified to appear as if obtained from the nadir view. This, however, requires information about the view angles of the cameras and a digital elevation model of the area, which were not available to this study.

Due to the limited resolution and/or field of view of digital cameras, an aerial image covers only a limited area on the ground. To increase the area of coverage of an image without reducing its resolution, overlapping images may be obtained and pieced together to create a larger image. This paper has concluded that, if information about camera and scene parameters is not available, although seamless mosaics can be created, which are useful for viewing and image analysis, the geometric fidelity of the mosaics will not be high enough to be used as maps. Information about the location of the camera rig with respect to the scene and a digital elevation model of the scene are needed to produce map-quality mosaics.

Acknowledgment

The authors would like to thank the Air Force Research Laboratory for the images, the Image Registration and Fusion Systems (imgfsr.com) for making its projective registration software library available to this project, the reviewers for their insightful comments, and Libby Stephens for her editorial assistance.

References

[1] H. Aanæs, J. R. Sveinsson, A. A. Nielson, T. Bøvith, and J. A. Benediktsson, "Model-based satellite image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1336–1346, May 2008.

[2] P. E. Anuta, "Spatial registration of multispectral and multitemporal digital imagery using fast Fourier transform techniques," *IEEE Trans. Geosci. Electron.*, vol. GE-8, no. 4, pp. 353–368, Oct. 1970.

[3] V. Arévalo and J. González, "Improving piecewise linear registration of high-resolution satellite images through mesh optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3792–3803, Nov. 2008.

[4] F. Bergholm, "Edge focusing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 6, pp. 726–741, Nov. 1987.
[5] F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2070–2082, Jul. 2008.
[6] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, Dec. 1992.
[7] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, Aug. 2007.
[8] C. Coelho, A. Heller, J. L. Mundy, D. A. Forsyth, and A. Zisserman, "An experimental evaluation of projective invariants," in *Geometric Invariance in Computer Vision*, J. L. Mundy and A. Zisserman, Eds. Cambridge, MA: MIT Press, 1992, pp. 87–104.
[9] G. M. Cortelazzo and L. Lucchese, "A new method of image mosaicking and its application to cultural heritage representation," *Comput. Graph. Forum*, vol. 88, no. 3, pp. 265–276, Sep. 1999.
[10] E. Fernández and R. Marti, "GRASP for seam drawing in mosaicking of aerial photographic maps," *J. Heuristics*, vol. 5, no. 2, pp. 181–197, Jul. 1999.
[11] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
[12] J. Flusser and T. Suk, "A moment-based approach to registration of images with affine geometric distortion," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 2, pp. 382–387, Mar. 1994.
[13] F. Fraundorfer and H. Bischof, "Evaluation of local detectors on nonplanar scenes," in *Proc. 28th Workshop Austrian Assoc. Pattern Recog.*, 2004, pp. 125–132.
[14] A. Goshtasby, "Piecewise linear mapping functions for image registration," *Pattern Recognit.*, vol. 19, no. 6, pp. 459–466, 1986.
[15] A. Goshtasby, "Registration of images with geometric distortions," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 60–64, Jan. 1988.
[16] A. Goshtasby, "Correction of image deformation from lens distortion using Bézier patches," *Computer Vis., Graph., Image Process.*, vol. 47, no. 3, pp. 385–394, Sep. 1989.
[17] A. Goshtasby, *2-D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications*. Hoboken, NJ: Wiley, 2005.
[18] A. Goshtasby and G. C. Stockman, "Point pattern matching using convex-hull edges," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 5, pp. 631–637, 1985.
[19] C. Harris and M. K. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, pp. 147–152.
[20] J. Kopf, M. Uyttendale, O. Deussen, and M. F. Cohen, "Capturing and viewing gigapixel images," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 1–10, Jul. 2007.
[21] M. Kreschner, "Seamline detection in color ortho-image mosaicking by use of twin snakes," *ISPRS J. Photogramm. Remote Sens.*, vol. 56, no. 1, pp. 53–64, Jun. 2001.
[22] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *Proc. IEEE Inf. Conf. Cybern. Soc.*, New York, 1975, pp. 163–165.
[23] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th Int. Conf. Comput. Vis.*, Kerkyra, Greece, 1999, pp. 1150–1157.
[24] R. Marzotto and V. Murino, "High resolution video mosaicing with global alignment," in *Proc. Comput. Vis. Pattern Recog.*, 2004, pp. I:692–I:698.
[25] P. Mclauchlan and A. Jaenicke, "Image mosaicing using sequential bundle adjustment," *Image Vis. Comput.*, vol. 20, no. 9/10, pp. 751–759, Aug. 2002.
[26] D. L. Milgram, "Computer methods for creating photomosaics," *IEEE Trans. Comput.*, vol. C-24, no. 11, pp. 1113–1119, Nov. 1975.
[27] D. L. Milgram, "Adaptive techniques for photomosaicking," *IEEE Trans. Comput.*, vol. C-26, no. 11, pp. 1175–1180, Nov. 1977.
[28] B. S. Morse, D. Gerhardt, C. Engh, and M. A. Goodrich, "Application and evaluation of spatiotemporal enhancement of live aerial video using temporally local mosaics," in *Proc. Comput. Vis. Pattern Recog.*, 2008, pp. 1–8.
[29] J. L. Mundy and A. Zisserman, *Geometric Invariance in Computer Vision*. Cambridge, MA: MIT Press, 1992.
[30] H. R. Myler and A. R. Weeks, *The Pocket Handbook of Image Processing Algorithms in C*. Englewood Cliffs, NJ: PTR Prentice–Hall, 1993, pp. 112–113.
[31] J. Núñez, X. Otazu, O. Fors, A. Prades, V. Palà, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.
[32] S. Peleg, "Elimination of seams from photomosaics," *Comput. Vis., Graph., Image Process.*, vol. 16, no. 1, pp. 1206–1210, 1981.
[33] S. Peleg, B. Rousso, A. Rav-Acha, and A. Zomet, "Mosaicing on adaptive manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1144–1154, Oct. 2000.
[34] F. P. Preparata and M. I. Shamos, *Computational Geometry: An Introduction*. New York: Springer-Verlag, 1985, p. 95. 226.
[35] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.
[36] H. S. Sawhney and R. Kumar, "True multi-image alignment and its application to mosaicing and lens distortion correction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 3, pp. 235–243, Mar. 1999.
[37] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *Int. J. Comput. Vis.*, vol. 37, no. 2, pp. 151–172, Jun. 2000.
[38] C. A. Shah, Y. Sheng, and L. C. Smith, "Automated image registration based on pseudoinvariant metrics of dynamic land-surface features," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3908–3916, Nov. 2008.
[39] Y. Shiren, L. Li, and G. Peng, "Two-dimensional seam-point searching in digital image matching," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 1, pp. 49–53, Jan. 1989.
[40] H.-Y. Shum and R. Szeliski, "Construction of panoramic image mosaics with global and local alignment," *Int. J. Comput. Vis.*, vol. 36, no. 2, pp. 101–130, Jul. 2000.
[41] C. C. Slama, *Manual of Photogrammetry*, 4th ed. Falls Church, VA: Amer. Soc. Photogramm., 1980.
[42] G. Stockman, S. Kopstein, and S. Benett, "Matching images to models for registration and object detection via clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 3, pp. 229–241, May 1982.
[43] R. Szeliski, "Image alignment and stitching: A tutorial," in *Handbook of Mathematical Models in Computer Vision*. New York: Springer-Verlag, 2005, pp. 273–292.
[44] J. R. G. Townshend, C. O. Justice, and C. Gurney, "The impact of misregistration on change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 1054–1060, Sep. 1992.
[45] L. Zagorchev and A. Goshtasby, "A comparative study of transformation functions for nonrigid image registration," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 529–538, Mar. 2006.
[46] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

David J. Holtkamp received the B.S. degree in computer engineering from Wright State University, Dayton, OH, in 1988, where he is currently working toward the M.S. degree in the Department of Computer Science and Engineering.

His research focuses on the design of tools for image analysis and, in particular, image registration. He is also currently involved in the design and implementation of a portable whole-body laser range scanner with self-calibration capability. Some of his other interests are field-programmable gate array design, mobile computing, and network/computer security.

A. Ardeshir Goshtasby received the B.E. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1974, the M.S. degree in computer science from the University of Kentucky, Lexington, in 1975, and the Ph.D. degree in computer science from the Michigan State University, East Lansing, in 1983.

He is currently a Professor of computer science and engineering with the Department of Computer Science and Engineering, Wright State University, Dayton, OH. His areas of research are image registration, image fusion, and curves and surfaces. He is the author of *2-D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications* (Wiley, 2005).