# Adaptive Registration of Very Large Images

Brian P. Jackson & A. Ardeshir Goshtasby
*Computer Science and Engineering Dept.*
*Wright State University*
*Dayton, OH 45435, U.S.A.*
{jackson.212,ardy}@wright.edu

***Abstract*—The main difficulty in image registration is in determining the geometric difference between the given images. An image registration method is introduced that determines the geometric difference between two images in a coarse-to-fine fashion. The global transformation needed to register the images is defined in terms of a collection of local transformations, each describing the geometric difference between corresponding local neighborhoods in the images. Preliminary experimental results of the proposed method are presented and discussed.**

***Keywords*-image registration; adaptive registration; hierarchical registration; big image;**

## I. Introduction

The increased resolution of cameras in recent years has contributed to the increased size and complexity of captured images. To accurately and efficiently register such images, there is a need to accurately and efficiently find local geometric differences between the images.

Local geometric differences between two images of a scene depend on the 3-D geometry of the scene and the view angle difference between the cameras capturing the images. The simplest scene geometry is a plane. Since the imaging process is a projective process and the relation between two projections of a plane is a homography, the images of a flat scene can be registered by a single homography, independent of the view-angle difference between the images.

Natural scenes are rarely flat, so it is unlikely that images of a natural scene can be registered by a single homography. With a given error tolerance, however, a complex scene geometry can be approximated by a composite of planar patches. This suggests that images of a complex scene can be registered by a combination of homographies, each defining the geometric difference between corresponding local neighborhoods in the images.

An adaptive image registration method is introduced that determines the transformation function for registration of two images in terms of a number of local transformations, each aligning corresponding local image areas. The global transformation function is then created from a blending of the local transformation functions.

For the method to be practical on very large images, an effort is made to design the registration steps in such a way that the overall computational complexity of the method becomes a linear function of the number of pixels in the images.

### A. Problem Statement

Given two images of a 3-D scene captured from different views, find a transformation function that will map the geometry of one image to that of the other, spatially aligning the images. In order that the method can be applied to very large images, it is required that the time complexity of the method be a linear function of the number of pixels in the images.

### B. Assumptions

It is assumed that geometric distortions caused by lens nonlinearities do not exist in the images; therefore, captured images can be considered projective transformations of a 3-D scene onto planes. Methods for removing image distortions from lens nonlinearities exist in the literature [6].

It is also assumed that the images do not contain large occluded areas, and if occlusion is present, it only covers a small percentage of the overlap area between the images. In addition, it is assumed that the images are in the same modality. This is required so that similar feature points can be found in the images to calculate the transformation parameters.

## II. Background

Image registration is the process of finding correspondence between all points in two images of a scene. The process involves finding a transformation function that can change the geometry of one image to resemble that of the other so that the images can be spatially aligned. This alignment is needed to fuse information in the images or to find changes in the scene occurring between the times the images are obtained. Image registration is the first step in many medical, remote sensing, industrial, and defense-related imaging applications [9].

Various transformation functions for image registration have been developed [16]. Although these transformation functions contain information about the geometric difference between the images to be registered, the complexity of obtained transformations often depends more on the number of feature-point correspondences found in the images than on the geometric complexity of the scene. Feature points are more a function of the intensity pattern in the scene than the geometry of the scene. A simple planar scene with highly varying color or albedo can produce a large number of feature points, resulting in a complex transformation function despite a simple scene geometry. The objective in this work is to design a smart registration method that relies more on the scene geometry than on the scene color or albedo when determining the transformation function for registration of the images.

The proposed method subdivides the image domain into a number of local neighborhoods and for each neighborhood determines the parameters of an affine transformation (an approximation to homography) to register corresponding neighborhoods in the images. To ensure that adjacent local transformations join smoothly, the local transformations are blended to a globally smooth and continuous transformation using rational Gaussian weights.

Rational weights were proposed by Shepard [14] to estimate missing data from known data [3], [4]. The weight of a point with a known value on a point with an unknown value is set to the inverse distance between the two points. Arsigny et al. [1] used the inverse-distance weights of Shepard to nonrigidly register histologic slices in an iterative algorithm.

Inverse-distance weights are fixed and cannot not adapt to the size of local neighborhoods. To adapt the weights to the size of neighborhoods, we use rational Gaussian weights. The standard deviation of the Gaussians is adjusted to the size of local neighborhoods. We define a transformation function by a rational Gaussian weighted sum of affine functions, each function registering corresponding local neighborhoods in the images. The details of this method follow.

## III. Approach

The method proposed for the registration of very large images has the following characteristics.

1) **High speed:** In order to achieve a fast processing speed, a multi-resolution approach is taken. Registration is first achieved at the lowest resolution where images are very small and have very little geometric differences. The result at the lowest resolution is then propagated to higher resolutions, converting a global registration problem to a collection of more manageable local registration problems.
   Guiding the correspondence process from coarse to fine has been attempted before. Likar and Pernus [12] and Jiang et al. [11] found corresponding points at the lowest resolution and used them to find correspondences at higher resolutions. The correspondences found at the highest resolution were then used to determine the parameters of a thin-plate spline [5] to register the images. Although thin-plate spline may be suitable for registration of images with up to a few hundred correspondences, when thousands of correspondences are available, not only computation of the transformation parameters becomes very time consuming, the system of equations to be solved may become ill-conditioned and impossible to solve.
2) **High precision:** In order to achieve high precision in registration, the global transformation is adapted to the local geometric differences between the images, ensuring that local neighborhoods in the images register accurately.
3) **Practical:** So that very large images can be efficiently registered, solution of very large systems of equations to find the transformation parameters is avoided. Instead, a method that requires the solution of only small equations to find local transformation parameters is employed. An efficient means to combine the local transformations into the global transformation is developed.

In the following, one of the images will be referred to as the *reference* and the second image will be referred to as the *sensed*. The reference image is kept unchanged and the sensed image is transformed to spatially align with the reference image. Points in the reference and sensed images will be denoted by $\mathbf{p} = (x, y)$ and $\mathbf{P} = (X, Y)$, respectively. The objective in image registration is to find a transformation function $\mathbf{F}$ such that for each point $\mathbf{p}$ in the reference image the corresponding point $\mathbf{P}$ in the sensed image can be determined from

$$\mathbf{P} = \mathbf{F}(\mathbf{p}). \tag{1}$$

Equation (1) makes it possible to scan the reference image and for each point $\mathbf{p}$ there find the corresponding point $\mathbf{P}$ in the sensed image, read the intensity/color there, and copy that at the point in the reference image. This process will, in effect, resample the sensed image point-by-point to a new image that has the geometry of the reference image.

Function $\mathbf{F}$ has two components, which we will denote by $F_1$ and $F_2$. Therefore, relation (1) can be written as

$$X = F_1(x, y), \tag{2}$$
$$Y = F_2(x, y). \tag{3}$$

The main objective in image registration is to find functions $F_1$ and $F_2$ that can map points in the reference image to the corresponding points in the sensed image.

To adapt the global transformation to local geometric differences between the given images, we define the global transformation by a weighted sum of a number of local transformations, each describing the geometric difference between corresponding local neighborhoods in the images. The simplest scene geometry is a plane and the relation between projections of a plane onto two image planes is a homography. When the neighborhoods are sufficiently small, local homographies can be approximated by affine transformations without loss of significant accuracy while gaining significant speed. Therefore, we define the global transformation by a weighted sum of local affine transformations.

A local affine transformation with components $f_1$ and $f_2$ mapping points in a neighborhood in the reference image to points in the corresponding neighborhood in the sensed image can be written as

$$X = f_1(x, y) = ax + by + c, \quad (4)$$
$$Y = f_2(x, y) = dx + ey + f. \quad (5)$$

In homogeneous coordinates, Eqs. (4) and (5) can be written as

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (6)$$

or

$$\mathbf{P} = \mathbf{f}\ \mathbf{p}. \quad (7)$$

If the global transformation is defined in terms of $N$ local transformations, and denoting the $i$th local transformation by $\mathbf{f}_i$, the relation between the global and local transformations can be written as

$$\mathbf{F} = \sum_{i=1}^{N} w_i \mathbf{f}_i. \quad (8)$$

Therefore, relation between points in the images can be written as

$$\mathbf{P} = \mathbf{F}\ \mathbf{p}. \quad (9)$$

Parameter $w_i$ in Eq. (8) is the contribution of the $i$th local transformation to the global transformation. If the center of the $i$th local neighborhood is $(x_i, y_i)$, $w_i$ is defined by

$$w_i(x, y) = \frac{W_i(x, y)}{\sum_{i=1}^{N} W_i(x, y)} \quad (10)$$

where $W_i(x, y)$ is proportional to the inverse distance of image point $(x, y)$ to the center of neighborhood $i$: $(x_i, y_i)$. We use a Gaussian to represent this proportionality. That is,

$$W_i(x, y) = \exp\left\{ -\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma_i^2} \right\}. \quad (11)$$

Therefore, $X$- and $Y$-components of the global transformation can be described in terms of the local transformations by

$$X = \sum_{i=0}^{n-1} w_i(x, y) f_{1_i}(x, y),$$
$$Y = \sum_{i=0}^{n-1} w_i(x, y) f_{2_i}(x, y). \quad (12)$$

$f_{1_i}$ and $f_{2_i}$ are the components of the $i$th local transformation, both centered at $(x_i, y_i)$, and $w_i(x, y)$ is the weight function associated with the $i$th local function.

The rational Gaussian weights defined in Eq. (10) have been used to design or recover 2-D and 3-D shapes from scattered points [7], [8]. We use rational Gaussian weights here to represent the components of a transformation function for image registration. Note that as the distance between point $(x, y)$ and the center of a neighborhood increases, the influence of the local transformation associated with that neighborhood on the point decreases. The weights are normalized to have a sum of 1 everywhere in the approximation domain (the domain of the reference image).

The standard deviation of the Gaussian associated with the $i$th neighborhood, $\sigma_i$, shows the influence of the $i$th feature point on distant points. The smaller its value, the more local the influence of the feature point will be. Typically, parameter $\sigma_i$ is set to half the width of the local neighborhood. Since all local neighborhoods are the same size in our subdivision approach, we let $\sigma_i = \sigma$ and select $\sigma$ globally. For a $128 \times 128$-pixel neighborhood/block, $\sigma$ will be 64 pixels. As $\sigma$ is increased, the global transformation becomes smoother. As $\sigma$ is decreased, the transformation will more closely map corresponding feature points in the images to each other. A smaller $\sigma$ should be used when the localization accuracy of the feature points is high. When the correspondences are noisy, a larger $\sigma$ should be used to smooth the noise.

Note that defining a transformation function by rational Gaussian weights does not require the solution of a system of equations. This avoids the need to solve large systems of equations when a large number of point correspondences is available. The weighted mean approach also makes the process of determining the transformation function stable as the possibility of obtaining an ill-conditioned matrix of coefficients does not exist.

Also, since a weight function decreases from the center of the block or neighborhood over which it is defined, the influence of a local function remains mostly local to the block it belongs to and its influence at a certain distance from the center of the block becomes negligible and can be ignored. This makes the global transformation locally sensitive and for very large images with thousands of local functions, the process will, in effect, estimate the value at a point in the global transformation from a small number of local transformations near the point. Moreover, if the blocks are of a fixed size, the Gaussian defined in Eq. (11) can be precomputed, saved, and reused when needed.

### A. *Multiresolution Subdivision*

In order to find the local transformations needed to accurately register two images, a coarse-to-fine subdivision approach is taken. The reference image is scaled down to a sufficiently small size of $d_1 \times d_2$ pixels. Typically, $d_1$ and $d_2$ are between 128 and 256 depending on the geometric complexity of the scene and the dimensions of the reference image. $d_1$ and $d_2$ should be larger for a scene with a smoothly varying geometry and they should be smaller for a scene with a sharply varying geometry. The sensed image is scaled down by the same amount. At such low resolutions,

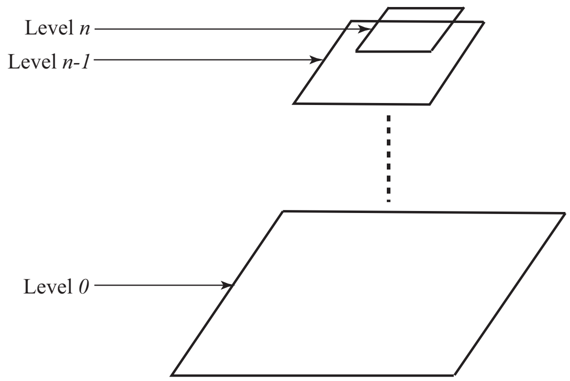


Figure 1. The image hierarchy used in the proposed image registration. The reference image at the top level is considered to have dimensions $d_1 \times d_2$, where $d_1$ and $d_2$ are typically between 128 and 256 pixels. The dimensions of the image at level $i-1$ are twice those of the image at level $i$. The same structural hierarchy is considered for the sensed image.

local geometric differences between the images become small and a single affine transformation can approximately align the images.

The transformation determined at the coarsest resolution is refined at the next higher resolution and new transformations are added as resolution is increased. By going from low to high resolution in this manner, the registration process is made to adapt the global transformation to the local geometric differences between the images.

The multiresolution image hierarchy used in the proposed method is depicted in Fig. 1. The resolution of each image at level $l$ is a factor of 2 higher than that of the same image at level $l+1$ in $x$ and $y$ directions. If $n$ is the top-level, the image at level $n$ is considered a single block and the image at level $n-1$ is divided into 4 equal blocks. Assuming the components of the transformation that register the images at the top level are represented by $f_1^{(n)}(x, y)$ and $f_2^{(n)}(x, y)$, the parameter of the transformation as described in Eqs. (4) and (5) can be computed by knowing three or more corresponding points in the images.

Since coordinates of points at level $n-1$ are a factor of 2 greater than the coordinates of the same points at level $n$, by multiplying the coordinates of feature points at level $n$ by 2 and using them to determine a new affine transformation with the components shown by Eqs. (4) and (5), the obtained transformation will register the images at level $n-1$ at least approximately. This step will transform the sensed image at level $n-1$ to the approximate geometry of the reference image at that level.

After the initial transformation at level $n-1$, the images are subdivided into 4 equal blocks and a new affine transformation is determined for each corresponding block. This process is the same as that used at the top level using Eqs. (4) and (5). Therefore, new corresponding points are found in corresponding blocks and the affine transformations are updated to more accurately register the images.

The process of going from a block at one resolution to blocks at one level higher resolution is the same as that going from images at level $n$ to subimages at level $n-1$. The block in each image is subdivided into 4 equal blocks and processed in the same manner. The global transformation obtained at the bottom level will be a blending of a large number of local transformations, each aligning corresponding blocks in the images at the bottom level (highest resolution).

To summarize, steps in the proposed adaptive algorithm are:

1) Register images at the top level: This step brings the lowest-resolution images into approximate alignment. Denoting the transformation that registers the images at the top level by $\mathbf{f}^{(n)}$, since there is only one transformation at the top level, $\mathbf{F}^{(n)} = \mathbf{f}^{(n)}$.
2) Compute an approximation to $\mathbf{F}^{(n-1)}$ from $\mathbf{F}^{(n)}$: If
$$\mathbf{F}^{(n)} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix}, \tag{13}$$
$$\tilde{\mathbf{F}}^{(n-1)} = \begin{bmatrix} a & b & 2c \\ d & e & 2f \\ 0 & 0 & 1 \end{bmatrix}, \tag{14}$$
will approximately register images at level $(n-1)$.
3) Refine the approximation: Subdivide the reference image and the resampled sensed image calculated in the preceding step into four equal blocks and register corresponding blocks in the images just like in the top level. Let's suppose the transformations obtained as a result are $\mathbf{f}_{00}^{(n-1)}$, $\mathbf{f}_{01}^{(n-1)}$, $\mathbf{f}_{10}^{(n-1)}$, and $\mathbf{f}_{11}^{(n-1)}$. These represent refinements to the approximately registered image blocks at level $n-1$.
4) Combine the local transformations into a global transformation: Combine the transformations obtained in the preceding step and $\tilde{\mathbf{F}}^{(n-1)}$ to obtain the global relation between the reference and sensed images at level $(n-1)$. That is, calculate
$$\mathbf{F}^{(n-1)} = w_{00}\mathbf{f}_{00}^{(n-1)}\tilde{\mathbf{F}}^{(n-1)} + w_{01}\mathbf{f}_{01}^{(n-1)}\tilde{\mathbf{F}}^{(n-1)} + w_{10}\mathbf{f}_{10}^{(n-1)}\tilde{\mathbf{F}}^{(n-1)} + w_{11}\mathbf{f}_{11}^{(n-1)}\tilde{\mathbf{F}}^{(n-1)}. \tag{15}$$
$w_{00}, w_{01}, w_{10}, w_{11}$ are weight functions as defined by Eq. (10) for the four blocks. Two consecutive affine transformations can be represented by a new affine transformation. Therefore, the combined transformation smoothly blends affine transformations from the four quadrants to create an overall nonlinear transformation due to the nonlinear nature of the weight functions.
5) Recursively repeat steps 1–4 treating corresponding blocks at a level as top-level images until reaching

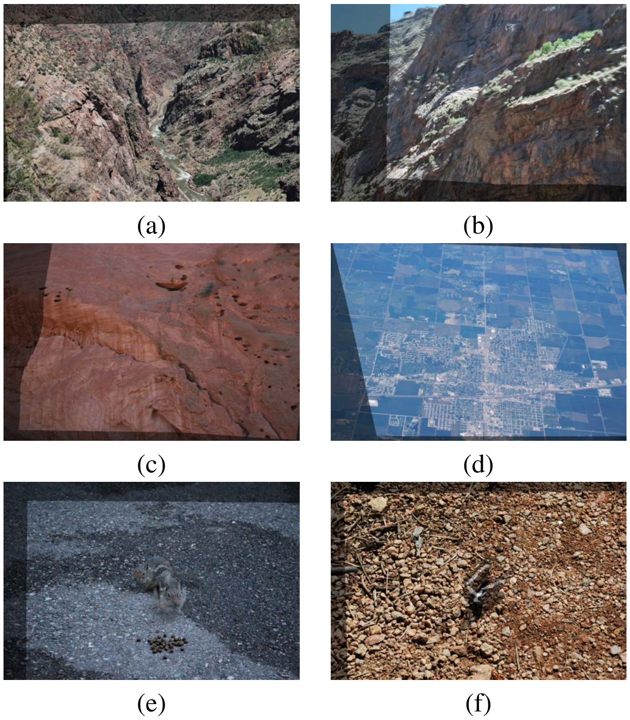


Figure 2. (a)–(c) Registration of images of terrain scenes containing sharp depth variations. (d) Registration of aerial images of a relatively flat scene captured at large off-nadir viewing angles. (e), (f) Registration of images containing moving objects/occlusions. These images are of size $3872 \times 2592$ pixels and are courtesy of *Image Registration and Fusion Systems*.

the bottom level. This will produce a global transformation at the bottom level that will be a weighted sum of local affine transformations, each registering corresponding blocks in the images at the highest resolution.

Determination of the global transformation function for registration of two images requires the ability to find corresponding feature points in the images. Various methods for determining feature points in images [10] and finding the correspondence between them [2] have been developed throughout the years that the readers may refer to.

## IV. Results

Example registrations by the proposed method are given in Fig. 2. These images are of dimensions $3872 \times 2592$. Frames (a)–(c) show registration of images captured from scenes with considerable depth variations, frame (d) shows registration of aerial images of a relatively flat scene but large off-nadir viewing angles, and frames (e) and (f) show registration of images containing occluded regions. These results demonstrate robustness of the method under highly varying geometric differences between images and presence of occlusion.

The images at the bottom-level (highest resolution) in Fig. 2 are subdivided into a grid of $16 \times 16$ subimages (blocks). A local affine transformation is determined to register corresponding reference and sensed subimages. The local transformations are then blended into a global transformation, which is used to resample the sensed image to the geometry of the reference image.

Computationally, registration of the images in Fig. 2 require about 30 seconds on a Windows PC with an Intel Core i7 processor, of which half of the time is spent on resampling the sensed image to the geometry of the reference image at the highest resolution by bilinear interpolation, a process that is a linear function of the number of pixels in the overlap area between the images.

The computational complexity of the method is on the order of $2^{2l}n < N_p$ simple operations, where $n$ is the number of feature points in each image block, $l$ is the number of levels in the hierarchy from low to high resolution, and $N_p$ is the number of pixels in the reference image. For the images in Fig. 2, $l = 4$ and $n = 40$. Therefore, 40 feature points are used in each block in the images to find the local transformation parameters. Local transformations from 256 blocks are then combined to create the global transformation, which is then used to resample the sensed image to overlay the reference image.

Considering blocks of size $256 \times 256$ and a traditional RANSAC algorithm to find the correspondence between $n$ feature points in corresponding blocks, and having $2^{2l}$ blocks at the bottom level in the coarse-to-fine hierarchy, correspondence can be established between points in the images on the order of $2^{2l}n$. As the number of feature points used in each image increases the fraction of correct putative correspondences decreases, exponentially increasing the number of iterations required by RANSAC to find the correspondence between points in the images when the images are related by a single homography or affine transformation. When the images have local geometric differences, traditional RANSAC cannot find the correspondences and a more elaborate and time consuming algorithm [13], [15], [17] is needed to find the correspondences.

If some information about the geometric relation between the images is available, the number of levels needed to register the images with sufficient accuracy can be selected. For instance, if it is known that the images represent aerial images of a relatively flat scene, $l = 1$ or 2 should be sufficient. If the scene is known to contain highly varying depth values, an appropriately larger $l$ is required to allow the process to adapt to local geometric differences between the images.

## V. Conclusions

Image registration is the first and often the most important step in many image analysis applications. An inaccurate registration can lead to an inaccurate or incorrect decision. Image registration should be sufficiently fast to be useful. A registration that takes hours to complete is of no value to an application that has to render a decision on an incoming image in a matter of minutes. With images of increased size, it is important that the method used to register images of size 10 megapixels take less than a minute on a personal computer to complete the process.

To achieve a high computational speed without sacrificing accuracy, a coarse-to-fine subdivision approach is proposed. Through image subdivision, not only a highly localized registration is achieved, a fast computational speed is gained as well. By ignoring points that do not fall in corresponding neighborhoods when searching for point correspondences, a computational complexity that is a linear function of the number of pixels in the images is reached.

The proposed method has some limitations. It requires that the given images be in the same modality. This is required so that many of the same feature points can be detected in the images. If the images are in different modalities but have the same global orientation and scale, registration can be achieved by finding feature points in the reference image and searching for them in the sensed image by template matching. This may slow down the computations considerably; however, the coarse-to-fine subdivision approach can still be used to limit search for the correspondences. Also, the proposed rational Gaussian weighted mean can still be used to blend the local transformations to a global transformation.

## ACKNOWLEDGMENT

The authors would like to thank *Image Registration and Fusion Systems* for the images used in this study.

## REFERENCES

[1] V. Arsigny, X. Pennec, and N. Ayache, "Polyrigid and polyaffine transformations: A novel geometrical tool to deal with non-rigid deformations – application to the registration of histologic slices," *Medical Image Analysis*, vol. 9, pp. 507–523, 2005.

[2] O. Chum and J. Matas, "Optimal randomized RANSAC," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1472–1482, 2008.

[3] R. Franke and G. M. Nielson, "Scattered data interpolation and applications: A tutorial and survey," *Geometric Modeling,* Springer, pp. 131–160, 1991.

[4] R. Franke and L. L. Schumaker, "A bibliography of multivariate approximation," *Topics in Multivariate Approximation*, pp. 275–335, 1987.

[5] A. Goshtasby, "Registration of images with geometric distortions," *IEEE Trans. Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 60–64, 1988.

[6] ——, "Correction of image deformation from lens distortion using Bézier patches," *Computer Vision, Graphics, and Image Processing*, vol. 47, no. 3, pp. 385–394, 1989.

[7] ——, "Design and recovery of 2-D and 3-D shapes using rational Gaussian curves and surfaces," *Int'l J. Computer Vision*, vol. 10, no. 3, pp. 233–256, 1993.

[8] ——, "Geometric modeling using rational Gaussian curves and surfaces," *Computer-Aided Design*, vol. 27, no. 5, pp. 363–375, 1995.

[9] ——, *2-D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications,* Wiley Press, 2005.

[10] ——, *Image Registration: Principle, Tools, and Methods,* Springer, 2012.

[11] J. Jiang, S. Zheng, A. W. Toga, and Z. Tu, "Learning based coarse-to-fine image registration," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–7, 2008.

[12] B. Likar and F. Pernus, "A hierarchical approach to elastic registration based on mutual information," *Image and Vision Computing*, vol. 19, pp. 33–44, 2001.

[13] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Processing*, vol. 23, no. 4, pp. 1706–1721, 2014.

[14] D. Shepard, "A two-dimensional interpolation function for irregularly spaced data," *Proc. 23rd Natl Conf. ACM*, pp. 517–524, 1968.

[15] E. Vincent and R. Laganière, "Detecting planar homographies in an image pair," *Proc. 2nd Int'l Sym. Image and Signal Processing and Analysis*, pp. 182–187, 2001.

[16] L. Zagorchev and A. Goshtasby, "A comparative study of transformation functions for nonrigid image registration," *IEEE Trans. Image Processing*, vol. 15, no. 3, pp. 529–538, 2006.

[17] M. Zuliani, C. S. Kenney, and B. S. Manjnath, "A multiransac algorithm and its application to detect planar homographies," *Proc. Int'l Conf. Image Processing*, vol. 3, pp. 153–156, 2005.