Volume image registration by template matching

Lijun Ding⋆, Ardeshir Goshtasby⋆, and Martin Satter†

⋆ *Department of Computer Science and Engineering,*
Wright State University, Dayton, OH 45435
† *Wallace-Kettering NeuroScience Institute,*
Kettering Medical Center, Kettering, OH 45429

Abstract

A template-matching approach to registration of volumetric images is introduced. The process automatically selects a dozen highly detailed and unique templates (cubic or spherical subvolumes) from one image and locates the templates in another image of the same scene. The centroids of the corresponding templates are used as corresponding control points and the best four correspondences that minimize an error criterion are used to determine the translational and rotational parameters needed to register the images. Use of different similarity measures in template matching and determination of coordinates of corresponding control points with subpixel accuracy are also discussed.

Keywords: Image registration, volumetric image, template matching, similarity measure, multimodal images

1 Introduction

Image registration is the process of determining the correspondence between all points in two images of the same scene. This process is required in many medical image analysis applications. For example, it is often needed to register multimodality images to enable maximal diagnostic sensitivity and specificity. Registration of cerebral anatomic images from an MR scanner with biochemical images from a PET scanner produces a unique dataset that is useful in both diagnosis and therapy.

To date, most work reported on volume image registration determines registration parameters using information from entire images. The approach taken here selects only similar areas in the images to determine the registration parameters. We believe that a global optimization that uses entire image volumes in the registration does not necessarily produce the best result when the images have partial intensity and/or geometric differences. For images

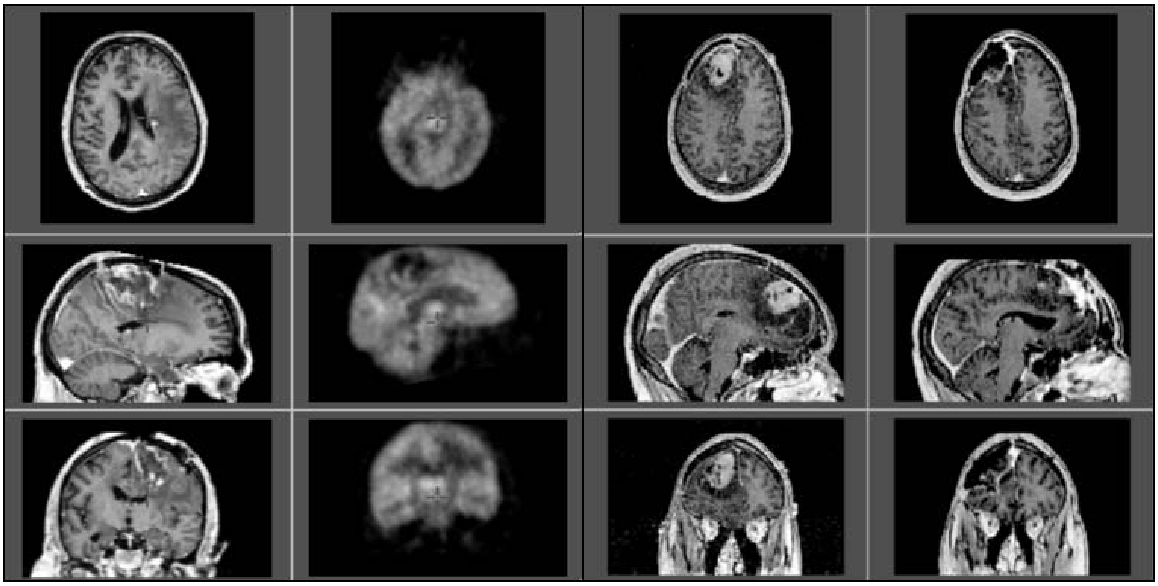


Figure 1: Rows 1, 2, and 3, respectively, show axial, sagittal, and coronal cross sections of volumetric images. Columns 1 and 2 show MR and PET images of a patient taken at about the same time, while columns 3 and 4 show MR images of a patient taken before and after a surgery.

with partial differences, an approach that uses image subvolumes in the matching should produce more accurate results than an approach that uses entire images in the matching.

Consider the brain images shown in Figure 1. The far left column depicts the axial, sagittal, and coronal views of a contrast-enhanced, T1-weighted MR scan of a patient with a brain tumor. The column to its right shows same views of a PET image of glucose metabolism of the same patient. Both images have been acquired on the same day. Note that the intensity patterns observed in the lesion in the two modalities do not correlate well, while patterns in the normal brain areas are similar.

The right two columns show the pre- and post-surgical MR images of a patient with a brain tumor. In the post-surgical image, local brain geometry has been considerably altered due to the resection. In this case, we do not want to use information at and near the tumor/resection site to determine the registration parameters. We would like to automatically select information from areas in the images that are similar and avoid selecting information from dissimilar areas. Thus, templates are selected from one of the image volumes and their matches are located in the other. A matching pair that reinforces a global registration are kept, while a pair that changes the parameters of the registration are discarded. In this manner, possible inaccuracies are detected and eliminated from the registration process.

If the centers of two image volumes correspond to each other and the images have small rotational differences, the average distance between corresponding voxels in the images will be larger than the average distance between corresponding voxels in two subvolumes of the images when the centers of the subvolumes coincide. Therefore, matching of image subvolumes (templates) is less sensitive to small rotational differences between images than matching of image volumes themselves. For that reason, we can expect to find template correspondences rather accurately even when the images to be registered have some rotational differences. Using a set of corresponding control points obtained from the centroids of corresponding templates in the images, we then determine the translational and rotational parameters that can bring the image volumes into registration.

The image volume "targeted" for re-orientation will be called the *target* and the image volume to which the target is aligned will be called the *reference*. We first describe the process of selecting a number of highly detailed and unique templates from the target image. We then outline the process of finding the corresponding templates in the reference image via template matching. Then, the selection criteria for finding the best four correspondences is given, and finally, a method for determining the transformation parameters from the four correspondences is described.

2 Template selection

To achieve highly accurate matches, the templates selected in the target image should represent highly detailed and unique regions. For a template to be highly detailed, it should contain a large number of high-gradient edges. This can be measured simply by computing the sum of gradient magnitudes in a template. To determine the highest detailed templates in an image, first, a small percentage (such as the top 5%) of the highest-gradient templates is selected and ordered in a list. Since many templates overlap each other, those that overlap by more than 50% of a selected template need not be included in the list. Among the remaining templates, the p most unique ones are selected and used in template matching.

One way to characterize uniqueness is to compute the correlation of a template with windows of the same size in its neighborhood. This process is called auto-correlation [18]. A sharp peak among the correlation coefficients at the template's position is evidence that the template is unique. The less sharp this peak, the more similar the template with windows in its neighborhood. When a template that is not locally unique is selected for matching, it may match rather well with many windows in its neighborhood, making distinction between the correct match and an incorrect one difficult.

Another way to characterize uniqueness is to use the eigenvalues of the inertia matrix of the template [14, 36]. A template that has all large eigenvalues represents a locally unique subvolume in an image. A template that has two large eigenvalues is not unique because it is similar to many templates in its neighborhood along a line. When only one of the eigenvalues is large, a template may be similar to many templates in its neighborhood along a plane. Unique templates enable determination of their correspondences with a high degree of accuracy. We select unique templates in an image using this second approach.

3 Template matching

Template matching is the process of finding the location of a subimage, called a *template*, inside an image. Once a number of corresponding templates are found, their centers are used as corresponding control points to determine the registration parameters.

Template matching involves comparing a given template with windows of the same size in an image and identifying the window that is most similar to the template. In the following sections, first, different similarity measures are reviewed, and then, search strategies to find the best-match position of a template in an image are discussed.

3.1 Similarity measures

Template matching requires comparison of a given template to windows of the same size in an image and identification of the window that is most similar to it. The accuracy of a template-matching process depends on the accuracy of the metric used to determine the similarity between a template and a window. The more accurate this metric, the more accurate the template-matching process.

Different metrics or similarity measures have been developed. There isn't a single similarity measure that is known to produce the best result in all situations. Depending on the types of images provided, one similarity measure may work better than another in template matching. In the following, we will review existing similarity measures and characterize their properties.

We assume a template denoted by f_1 and an image denoted by f_2 are given. We also assume the template is of size $n \times n \times n$ and the image is of size $m \times m \times m$ $(m > n)$. We will generate an intermediate image, called the *similarity image* and denote it by s. Entry (x, y, z) in the similarity image will show the similarity between the template and the window of the same size at location (x, y, z) in the image. Similarity image s will be of size $(m - n + 1) \times (m - n + 1) \times (m - n + 1)$.

3.1.1 Sum of absolute differences

Sum of absolute intensity differences is the Minkowski metric of order one [9] and is defined by

$$s(x,y,z)=\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}|f_1(i,j,k)-f_2(x+i-1,y+j-1,z+k-1)|;\quad x,y,z=1,\ldots,m-n+1. \tag{1}$$

Coordinates (x, y, z) represent the front-upper-left corner of a window in image f_2 where the template is being matched. Measure s shows the dissimilarity between f_1 and the window at location (x, y, z) in f_2. The smaller the value of $s(x, y, z)$, the more similar the template and the window. This similarity (or more precisely, dissimilarity) measure requires in the order of n^3 additions for each search position. An algorithm proposed by Barnea and Silverman [4] can speed up this computation further. Assuming the smallest value of s obtained so far is s_{min}, the algorithm keeps track of s_{min} and at each iteration in equation (1) compares the obtained sum to s_{min}. If the sum obtained so far is equal to or greater than s_{min}, further computation of the similarity measure at that position is abandoned. This is done because further computation at that position will only increase the value of s. If the sum obtained for all iterations is less than s_{min}, s_{min} is replaced with the new sum. Abandoning unnecessary computations speeds up the process by a factor of two to three. The computational complexity of the sum of absolute differences is still $O(n^3m^3)$ additions (but with a smaller coefficient), when m is considerably larger than n. If $n \approx m$, and assuming search area size is $d \times d \times d$, where $d = m - n + 1$, computational complexity of the algorithm will be in the order of n^3d^3 additions, which we denote by $O(n^3d^3)$.

Since this similarity measure finds the sum of absolute differences between raw image intensities, it is required that the two images that are given for registration be from the same modality. Even images in the same modality may have some intensity differences due to different scanning conditions. A study carried out by Svedlow *et al.* [35] found that image gradients produce more accurate matches than the raw image intensities when using this similarity measure. Similar result was observed by Penney *et al.* [25]

3.1.2 Cross-correlation coefficient

Cross-correlation coefficient is a distance metric of order two and is defined by [2, 27]

$$s(x,y,z)=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}f_1(i,j,k)f_2(x+i-1,y+j-1,z+k-1)}{[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}f_1^2(i,j,k)]^{\frac{1}{2}}[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}f_2^2(x+i-1,y+j-1,z+k-1)]^{\frac{1}{2}}} \tag{2}$$

where $x, y, z = 1, \ldots, m - n + 1$ are the coordinates of the front-upper-left corner of a window in the reference image. In this formula, the denominator is a normalization factor.

As the template and the window become more similar, $s(x, y, z)$ becomes larger. If the template and the window are normalized so that they have a mean of zero, s will have a value between -1 and +1. Positive values will show dependency (similarity) of intensities in the template and the window, and the larger the value of s the higher this dependency. When $s = 1$, intensities in the template and the window completely depend on each other and, therefore, they increase and decrease together. When $s = 0$, intensities in the two images are independent of each other, and when s is negative, intensities in the template and the window inversely depend on each other. When $s = -1$, changes in intensities of the template will be the reverse of changes in intensities of the window.

The computational complexity of the numerator of formula (2) for each search position is $O(n^3)$ multiplications. The first term in the denominator has to be computed once. The second term in the denominator has to be recomputed for each search position. Therefore, computation of (2) will require in the order of n^3m^3 multiplications, when m is much larger than n. If the template is about the same size as the image, and $d = m - n + 1$, the computational complexity of template matching using the cross-correlation coefficient as the similarity measure will be $O(m^3d^3)$ multiplications.

Cross-correlation coefficient is resistant to some intensity differences between images. Therefore, even when the template and the window have intensity differences, as long as intensities of corresponding voxels increase and decrease together, a high correlation coefficient will be obtained. This means that intensities in two images could be considerably different, but as long as they represent the same pattern, a high similarity will be obtained. This metric is, therefore, preferred over the sum of absolute differences in template matching. A study carried out by Svedlow *et al.* [34] found that cross-correlation coefficient produced more accurate matches than the sum of absolute differences when the images had no rotational differences. A study carried out by Penney *et al.* [25] found that sum of absolute differences produced more accurate matches than cross-correlation coefficient when the images had some rotational differences.

3.1.3 Geometric distance

If the template and the image contain binary structures, such as image edges, geometric distance is the preferred metric for template matching. If the template and the image are obtained by different sensors, image intensities cannot be used in sum of absolute differences or cross-correlation coefficient to determine the correspondences. In such a situation, if the images can be segmented to obtain similar surface structures, the surfaces can be used in the matching. The process of matching image structures is known as *chamfer matching* and was first proposed by Barrow *et al.* [5] in matching of aerial images [5]. The idea, however, is general and can be used in matching of volumetric images also [8].

Chamfer matching is a technique that uses the average distance between two binary structures as the similarity measure. To determine the similarity between a template and a window in an image, the window and the template are overlaid and for each structure point in the window the structure point closest to it in the template is determined and the average of the distances between the corresponding points is used as the similarity measure. An efficient algorithm for determining the average distance is given by Borgefors [7, 8]. When the template is shifted over an image, the computed average distance (similarity measure) will change also. The best match is the position where the average distance between the structures is minimum.

To carry out template matching using this similarity measure, it is sufficient to prepare a distance image for the reference image with entries showing their distances to structure points closest to them in the image. Then, as the template is shifted over the distance image, at each shift position the sum of the values in the distance image falling on structure points in the template is computed and divided by the number of structure points. The smaller the obtained average, the more similar the template and the window. An average of zero means that the template and the window match perfectly. As this average increases, the similarity between structures in the template and the window decreases.

Computation of the distance image takes in the order of m^3 additions. Assuming a structure with N voxels is present in the template, at each shift position, the process requires N additions to determine the similarity between the template and the window there. If $n \approx m$ and there are $d \times d \times d$ shift positions where the template can be shifted over the image ($d = m-n+1$), the time needed to determine the best match position is $O(Nd^3+m^3)$ additions. Typically, m^3 and Nd^3 are about the same. Therefore, computational complexity of chamfer matching is $O(m^3)$ additions when m and n are about the same. When m is much larger than n, computational complexity of chamfer matching is $O(m^3N)$ additions. Note that the time needed to segment the images before carrying out chamfer matching is not included in this computation. The time needed to determine the image structures used in chamfer matching may actually be longer than the time needed to carry out the template matching.

As long as structures obtained in the template are a subset of structures obtained in the image, chamfer matching will succeed. If image edges are used as structures in template matching, only very strong edges should be kept in the template. As long as edges in the template exist in the corresponding window, a high similarity will be obtained even when some edges in the window do not exist in the template. Therefore, when finding edges in the reference and target images, a smaller percentage of the edges should be kept in the target image than in the reference image in order to achieve a higher accuracy in template matching.

Chamfer matching is similar to surface matching methods [6, 26] that align two surfaces using an optimization method. The accuracy of surface matching techniques depends on the accuracy of the surface structures provided for matching. A study carried out by Kularatna [22] in registration of single modality images found that in almost all cases, image intensities produce more accurate matches than matching of the surface structures. A similar conclusion was reached by Fitzpatrick and West [13] in matching of multimodality images. This is expected since surface-based techniques discard information from the inside of objects and use information from only their surfaces to match images.

3.1.4 Mutual information

Suppose a template T and a window W are available and we want to determine their similarity. Assuming T and W are random variables and $P_T(a)$ is the probability that the intensity at a voxel in T is a and $P_W(b)$ is the probability that the intensity at a voxel in W is b, then, if we overlay the template and the window, the probability that intensity a in the template lies on top of intensity b in the window will be equal to their joint probability: $P_{TW}(a,b)$. If the template and the window truly correspond to each other, their intensities will be highly dependent and they will produce high joint probabilities. However, if the template and the window do not correspond to each other, they will produce small joint probabilities. If intensities in the template and the window are completely independent, the joint probabilities will be $P_T(a)P_W(b)$. Given these probabilities, mutual information can be computed from [24, 37]

$$I(T,W) = \sum_a \sum_b P_{TW}(a,b) \log \frac{P_{TW}(a,b)}{P_T(a)P_W(b)}. \tag{3}$$

The joint probabilities $P_{TW}(a,b)$ for different values of a and b can be estimated from the histogram of intensities of corresponding voxels in the matching template and window. To obtain the histogram, a 256×256 array is allocated and all its entries are initialized to zero. It is assumed that intensities in the template and the window vary between 0 and 255. If at a particular voxel the template shows intensity a and the window shows intensity b, entry (a,b) in the array is incremented by one. After processing all voxels in the template and the window, an array will be obtained whose entry (a,b) shows the number of voxels in the template with intensity a where corresponding positions in the window have intensity b. To obtain the probabilities, contents of the histogram array are divided by the sum of the entries. Note that the sum of the entries is n^3 if the template and the window are $n \times n \times n$. Entries of the obtained array correspond to the values of $P_{TW}(a,b)$ for different values of a

and b.

Assuming intensity a denotes the rows and intensity b denotes the columns in $P_{TW}(a, b)$, we can estimate $P_T(a)$ and $P_W(b)$ from

$$P_T(a) = \sum_{b=0}^{255} P_{TW}(a, b) \tag{4}$$

and

$$P_W(b) = \sum_{a=0}^{255} P_{TW}(a, b), \tag{5}$$

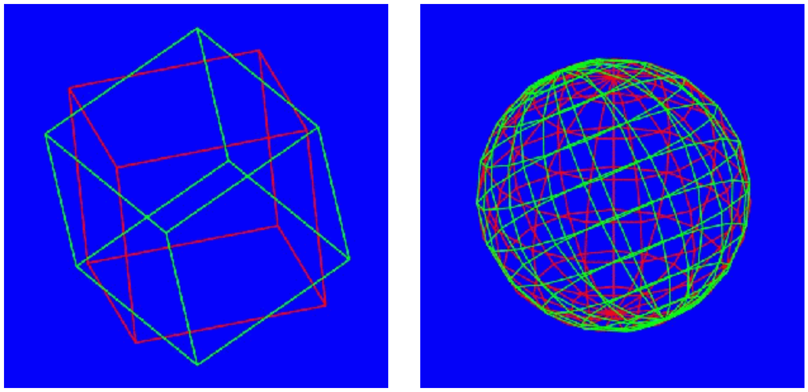
respectively. As long as intensities in the template and the window correlate, a high mutual information will be obtained. Compared to cross-correlation coefficient and sum of absolute differences, mutual information is more sensitive to image noise. Intensity a in the template should always correspond to intensity b in the window to produce a high similarity. This shows that images from different sensors can be registered using mutual information as long as, for example, homogeneous areas in the images correspond to each other. However, noise in one or both images will quickly degrade the similarity measure. A study carried out by Penney *et al.* [25] in registration of same modality images found that mutual information did not perform as well as the cross-correlation coefficient or the sum of absolute differences. We have found that slight image smoothing, which reduces image noise, improves the registration accuracy of mutual information.

The computational complexity of mutual information at each shift position is in the order of n^3 additions and 256^2 multiplications. Usually, n^3 and 256^2 are about the same, and so the time needed to search for the best match position of a template of size $n \times n \times n$ in an image of size $m \times m \times m$ when m is much larger than n is $O(n^3 m^3)$ additions. If the template and the image are about the same size and the search area size is $d \times d \times d$, where $d = m - n + 1$, computation time will be $O(m^3 d^3)$ additions.

None of the similarity measures mentioned above can accurately measure the similarity between two images when they have rotational differences. If two images have rotational differences, invariant moments should be used to align them.

3.1.5 Invariant moments

Moments are features that characterize the geometry of a pattern. Moments can be normalized to become invariant of the position and orientation of a pattern [19, 23, 31]. Invariant moments are especially useful in template matching when the images have rotational differences. The rotational difference between images makes sum of absolute differences, cross-correlation coefficient, geometric distance, and mutual information ineffective in template matching. Since invariant moments are independent of the orientation of a pattern, they can



(a) (b)

Figure 2: (a) A cubic template and a cubic window whose centers coincide. (b) A spherical template and a spherical window whose centers coincide.

be used to determine the similarity between two subimages irrespective of their rotational differences.

A point to note is that if the template is rotated with respect to an image, and if the template is cubic, it is not possible to have a template and window with coinciding centers that can contain the same scene parts unless the rotational difference between them is a multiple of 90 degrees. This is demonstrated in Figure 2a. Even when the centers of a cubic template and a cubic window correspond to each other, since they contain somewhat different parts of a scene, the computed similarity may not be high enough to produce a correct match. To remedy this weakness, spherical templates are used. If spherical templates are used, when the centers of a template and a window correspond to each other, they will contain the same scene parts (see Figure 2b). When a template and a window contain the same pattern, they will have similar invariant moments irrespective of their orientations.

The pqrth moment of template f_1 is defined by [36]

$$m_{pqr} = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} i^p j^q k^r f_1(i,j,k), \tag{6}$$

and the pqrth order central moment of the template is defined by

$$\mu_{pqr} = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} (i-\bar{i})^p (j-\bar{j})^q (k-\bar{k})^r f_1(i,j,k), \tag{7}$$

where

$$\bar{i} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} i f_1(i,j,k)}{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} f_1(i,j,k)}, \tag{8}$$

$$\bar{j} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} j f_1(i,j,k)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} f_1(i,j,k)}, \tag{9}$$

$$\bar{k} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} k f_1(i,j,k)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} f_1(i,j,k)}. \tag{10}$$

The principal axes of a 3-D pattern are defined by the eigenvectors of the inertia matrix of the pattern [14], where the inertia matrix is defined by

$$M = \begin{pmatrix} \mu_{200} & \mu_{110} & \mu_{101} \\ \mu_{110} & \mu_{020} & \mu_{011} \\ \mu_{101} & \mu_{011} & \mu_{002} \end{pmatrix}. \tag{11}$$

Since M is a real and symmetric matrix, it will have real eigenvalues $\{\lambda_i : i = 1, 2, 3\}$ and orthogonal eigenvectors $\{\mathbf{u}_i : i = 1, 2, 3\}$. The rotational difference between the template and a window will be the amount of rotations needed to align the principal axes of the template and the principal axes of the window. To align a template and a window, first, the eigenvectors corresponding to the largest eigenvalues of the template and the window are aligned. Then, the eigenvectors corresponding to the second largest eigenvalues are aligned. This will automatically align their third eigenvectors. The transformation needed to align the principal axes of the template and the window can be used to align the template and the window. Once the rotational difference between the window and the template is eliminated, their similarity can be determined using any of the metrics mentioned earlier.

Note that eigenvalues and eigenvectors of the inertia matrix of a template are determined using intensities in the template: $f_1(i, j, k)$. For the template and its corresponding window to have similar eigenvalues and eigenvectors, the images provided for registration should be of the same modality. This means, use of invariant moments in image registration is limited to single-modality images. Successful registration of single-modality images using invariant moments has been reported [1].

Assuming a highly detailed and unique template is selected in the target image, the window in the reference image that corresponds to it can be obtained by 1) aligning windows in the reference image with the template using their principal axes, 2) determining the similarities between the template and the matched windows using sum of absolute differences, cross-correlation coefficient, or mutual information, and 3) selecting the window that produces the highest similarity measure. For the match to be unique, it is required that not only the template have large eigenvalues, but that no two eigenvalues have the same magnitude. Otherwise, wrong axes of the template and the window could be aligned, missing the correct match.

Small errors obtained in computation of the principal axes due to noise and intensity differences between images will result in small errors in alignment of the template and the window. Since sum of absolute differences, cross-correlation coefficient, and mutual information are tolerant to some rotational differences between images, we expect to register the images even when the template and the window are not initially aligned accurately. Since the tolerance of these similarity measures to rotational differences between images increases as images become more blurred, the template-matching process may be carried out in two steps. First, the smoothed template and image can be approximately aligned. Then, the approximately aligned template and image can be used in full resolution to find the accurate match by searching in a small neighborhood of the approximate match.

The computational complexity of invariant moments is $O(n^3m^3)$ multiplications if m is much larger than n, and $O(m^3d^3)$ multiplications if the template and the image are about the same size and $d = m - n + 1$. The constant multiplication factors involved in these formulas are rather large due to considerable computations involved in finding the eigenvalues and eigenvectors of the inertia matrices of the template and the matching windows. Therefore, a more realistic estimation of the computational complexity of template matching using invariant moments is $O(n^3m^4)$ or $O(m^4d^3)$.

Among the similarity measures mentioned above, cross-correlation coefficient, geometric distance, and invariant moments provide the most unique measures. The intensities in a template (and correspondingly in the matching window) can be rearranged without changing the similarity computed by the sum of absolute differences or the mutual information. In cross-correlation coefficient, invariant moments, and geometric distance, changing the pattern in the template and the window may change the similarity measure, thus resulting in more unique similarity measures in the matching. Mutual information is useful when registering multimodality images. As long as a mapping function exists that can map intensities in a template to intensities in a matching window, a high similarity measure will be obtained between the matching template and window. There is no need for intensities in a template and a window to correlate to produce a high similarity measure. It is only needed that a function exist that can uniquely intensities in the template and window. For this reason, we choose mutual information as the preferred similarity measure when registering multimodality images.

3.2 Matching with subpixel accuracy

The positions of windows in the reference image that are matched with a template are discrete positions. The best-match position, however, could be somewhere between the discrete positions. To determine a best-match position with subpixel accuracy, centered at the discrete best-match position, similarities in a $3 \times 3 \times 3$ neighborhood are found and a tri-quadratic function is fitted to them. Then, the maximum or minimum of the function is determined. A minimum is determined when the sum of absolute differences or the geometric distance is used, while a maximum is determined when the cross-correlation coefficient or the mutual information is used. A tri-quadric function is defined by

$$s(x, y, z) = Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fxz + Gx + Hy + Iz + J. \tag{12}$$

Parameters A–J are determined by the least squares method using the $3\times3\times3$ array of similarities and their coordinates. Once A–J are determined, the maximum (for cross-correlation and mutual information) or minimum (for sum of absolute differences and distance measure) of $s(x, y, z)$ is determined by computing its derivatives with respect to x, y, and z; setting them to zero; and solving the obtained system of three linear equations for x, y, and z. In our implementation, after corresponding control points are determined, this method is used to determine the correspondences with subpixel accuracy.

3.3 Search techniques

To find the best-match position of a template in an image, different search techniques can be used. Coarse-to-fine search and optimization algorithms are the most popular ones.

In the coarse-to-fine approach, first, the images are reduced in scale and approximate match positions are determined. Then, the scales of the images are increased in two to three steps until full-resolution images are obtained. At each step, possible match positions are determined with associated confidence levels and a search at a higher resolution is performed only at high confidence matches. This technique has been used in 2-D template matching using sum of absolute differences [29, 38] and cross-correlation coefficient [17], but the idea is general and can be used in 3-D template matching also.

Optimization techniques require that an approximate match position be provided. Then, the accurate match position is determined by an iterative process while optimizing a criterion. Typical approaches are hill climbing [28], gradient descent [10], simulated annealing [21], and evolutionary algorithms [11]. Such algorithms often require that a good initial registration be provided. They then determine globally optimal parameters to register the images.

When the images are approximately registered, since the search area is small, an exhaustive search also can find the solution in a small number of steps. Trying to speed up the search using an optimization process may actually slow down the process because of the overhead involved in the optimization. In our implementation, the user is allowed to approximately align the images; therefore, the search area is small, and so we use an exhaustive search to locate the best match position of a template in the reference image.

4 Computing the registration parameters

Given a set of corresponding control points in two images, we would like to determine the translational and rotational differences between the images. Let's denote the translational and rotational differences between the image by $\mathbf{T}$ and $\mathbf{R}$, respectively, and the coordinates of corresponding control points by $\{\mathbf{p}_i = (x_i, y_i, z_i), \mathbf{P}_i = (X_i, Y_i, Z_i) : i = 1, \ldots, p\}$. Then, we can write

$$\mathbf{p}_i = \mathbf{R}\mathbf{P}_i + \mathbf{T} + \mathbf{N}_i, \tag{13}$$

where $\mathbf{N}_i$ is the unknown inaccuracy associated with the ith control point correspondence. The translation vector $\mathbf{T}$ and the rotation matrix $\mathbf{R}$ can be determined by minimizing

$$E^2 = \sum_{i=1}^{p} ||\mathbf{p}_i - (\mathbf{R}\mathbf{P}_i + \mathbf{T})||^2. \tag{14}$$

An efficient algorithm for determining $\mathbf{R}$ and $\mathbf{T}$ has been given by Arun *et al.* [3]. In this method, first, the rotation matrix is obtained by minimizing

$$E_R^2 = \sum_{i=1}^{p} ||\mathbf{Q}_i - \mathbf{R}_i\mathbf{q}_i||^2, \tag{15}$$

where $\mathbf{Q}_i = \mathbf{P}_i - \bar{\mathbf{P}}$, $\mathbf{q}_i = \mathbf{p}_i - \bar{\mathbf{p}}$, and $\bar{\mathbf{P}}$ and $\bar{\mathbf{p}}$ are the centers of gravity of the control points in the target and reference images, respectively. Then, knowing the rotation matrix, the translation vector $\mathbf{T}$ is determined from

$$\mathbf{T} = \mathbf{P} - \mathbf{R}\mathbf{p}. \tag{16}$$

If exact control point correspondences were known, the translational and rotational differences between the images could be determined from the coordinates of only three corresponding control points in the images by solving a system of linear equations. However, some quantization and matching errors are expected in the coordinates of corresponding control points. When control point correspondences contain errors, more than three control point pairs are needed to determine the registration parameters by the least-squares method as outlined above. Our experiments have shown that if the four most accurate correspondences are used, a better result will be obtained than using more than four correspondences, some of which are not as accurate. Therefore, we use the coordinates of the best four correspondences in the images to determine the registration parameters.

5 Determining the best four correspondences

Registration by matching templates has advantages over methods that match entire images. When entire images are matched, noise and dissimilarity between images average out, influencing the registration result. However, if a number of matches is established between images through template matching, the inaccurate ones can be discarded, and only the accurate ones can be used to find the registration parameters.

To distinguish the inaccurate matches from the accurate ones, we compare the lengths of edges connecting corresponding control points in the images. If two points in the target image truly correspond to two points in the reference image, distances between the points in the two images will be the same. Although there is no absolute guarantee that when distances between two corresponding point pairs in the images are the same the point pairs will correctly correspond to each other, but the probability that two matches shift by exactly the same amount and in exactly the same direction is extremely small. Therefore, for an overwhelming majority of the cases, this test is sufficient to find accurate corresponding control points in the images. To make the process more reliable, multiple tests are performed to identify an inaccurate match from the accurate ones.

To choose the best p correspondences from among the available P matches, we find distances between point pairs in each image and list corresponding point pairs in the ascending order of the difference between corresponding distances and take p of the points starting from the top of the list. In our case, $p = 4$ and, therefore, we use 4 pairs of corresponding points to determine the registration parameters. To further test the correctness of the selected 4 correspondences, we use 4 more correspondences in the list and by determining the transformation functions with combinations of the points, 4 at a time, determine whether all combinations produce similar parameters or not. If a combination produces a set of parameters different from other combinations, we find the control point that is not in the set of control points producing similar parameters and remove it. If that control point happens to be among the top 4, obviously, the control point will be removed and the next control point in the list will be selected to determine the registration parameters. As we go down the list, error between corresponding point pairs increases. Selecting the best 4 correspondences is a way of eliminating the less accurate ones as well as the outliers. Other methods to remove outliers have been proposed [12, 20, 32].

An alternative method for distinguishing the accurate matches from the inaccurate ones is to examine the geometries of the components of the transformation function that map one image to another [30]. A transformation function for registering volumetric images has three components:

$$X = f_x(x, y, z) \quad (17)$$

$$Y = f_y(x, y, z) \quad (18)$$

$$Z = f_z(x, y, z) \quad (19)$$

where (x, y, z) and (X, Y, Z) are coordinates of corresponding control points in the reference and target images, respectively. Given a point in the reference image, the components of a transformation determine coordinates of the corresponding point in the target image. The target image can be resampled in this manner to overlay the reference image. Since we are considering registration of brain images of the same patient and we expect only translational and rotational differences between the images, the components of the transformation are linear. If the control point correspondences are accurate, we will obtain a plane representing each component of the transformation. However, if the control point correspondences are inaccurate, the functions will not be linear. To detect the inaccurate correspondences, we will represent each component of the transformation by a nonlinear function, such as the thin-plate spline [15] or the rational Gaussian surface [16], and observe the gradients of the surfaces. Constant gradients in all three components of the transformation will be evidence that all correspondences are accurate. Variations in gradients in the neighborhood of a match in any of the components of the transformation is evidence that the match is not accurate. In this manner, accurate and inaccurate matches can be distinguished from each other. In this work, the former method is used to select the best four matches from among a dozen matches obtained by template matching.

6 Results

We have implemented an image registration method that uses template matching to determine a dozen corresponding control points in the images with subpixel accuracy and uses the best four correspondences to determine the registration parameters. Mutual information is used as the similarity measure.

The accuracy of the proposed registration method on eight sets of single-modality MR brain images from the Vanderbilt dataset [39, 40] are shown in Table 1. Each image pair was approximately aligned using the mouse device within a 15-second time limit. Then, the automatic template selection and template matching process was initiated to find a number of corresponding templates in the images. The centers of the corresponding templates were used as corresponding control points, the inaccurate correspondences were removed as outlined above, and the best four correspondences were used to compute the registration parameters.

Entries in Table 1 show mean, median, and maximum errors for each dataset. After determining the registration parameters from four corresponding control points in the images, the target image was resampled to overlay the reference image. Since template matching accuracy is influenced by rotational differences between images caused by the approximate registration, after the initial registration, the target image was replaced with the resampled target image and the new target image was assumed to approximately register with the

Table 1: Mean, median, and maximum errors (in mm) of the proposed image registration method using 8 single modality MR images from the Vanderbilt dataset [39, 40]. Errors achieved by the best method are shown in parentheses.

Data Set	Mean	Median	Maximum
1	3.23 (2.78)	3.20 (2.80)	3.46 (2.90)
2	1.75 (1.81)	1.77 (1.82)	1.92 (1.88)
3	2.01 (1.65)	2.07 (1.61)	2.61 (2.12)
4	4.03 (2.58)	3.99 (2.62)	4.47 (2.12)
5	2.79 (3.06)	2.62 (2.99)	3.71 (3.36)
6	3.52 (3.46)	3.48 (3.36)	4.12 (4.19)
7	2.64 (2.54)	2.67 (2.41)	3.74 (3.45)
8	2.10 (2.62)	2.10 (2.66)	2.84 (3.31)

reference image. The automatic template selection and template matching process was repeated to determine the new registration parameters. The results summarized in Table 1 are, therefore, after two iterations of our method. The brain images used originally had external fiducial markers. The markers were visible in the images and so they were used by West *et al.* [39] to register the images. We used the registration results obtained with the fiducials as the gold standard. The images we received did not have the fiducials and therefore we had to automatically find a set of corresponding control points in the images by template matching. Our results in Table 1 are comparable to the best results available today [33, 40]. Our overall mean and median errors are only 0.2 mm and 0.15 mm, respectively, larger than those of the best method, and our overall worst error is no worse than that of the best method by 0.27 mm. The errors for the best method are shown in parentheses in this table.

To determine the accuracy of our registration method on multimodality images, 13 datasets, again prepared by West *et al.* [39, 40], were used. The first 7 datasets contain MR and CT images and the last 6 datasets contain MR and PET images. Each dataset contains from a few to a dozen image pairs, resulting in a total of 106 image pairs. Mean, median, and maximum registration errors in each dataset are computed and listed in Table 2. The registration results using the fiducials as produced by West *et al.* [39] were used as the gold standard to determine the registration accuracy reported here. Again, since the images we used did not contain fiducials, corresponding control points were determined by template matching. The results shown in Table 2 are also after two iterations of the template-matching process. First, the images were aligned approximately through an interactive step. Then, the approximate registration was refined through template matching. The obtained registration

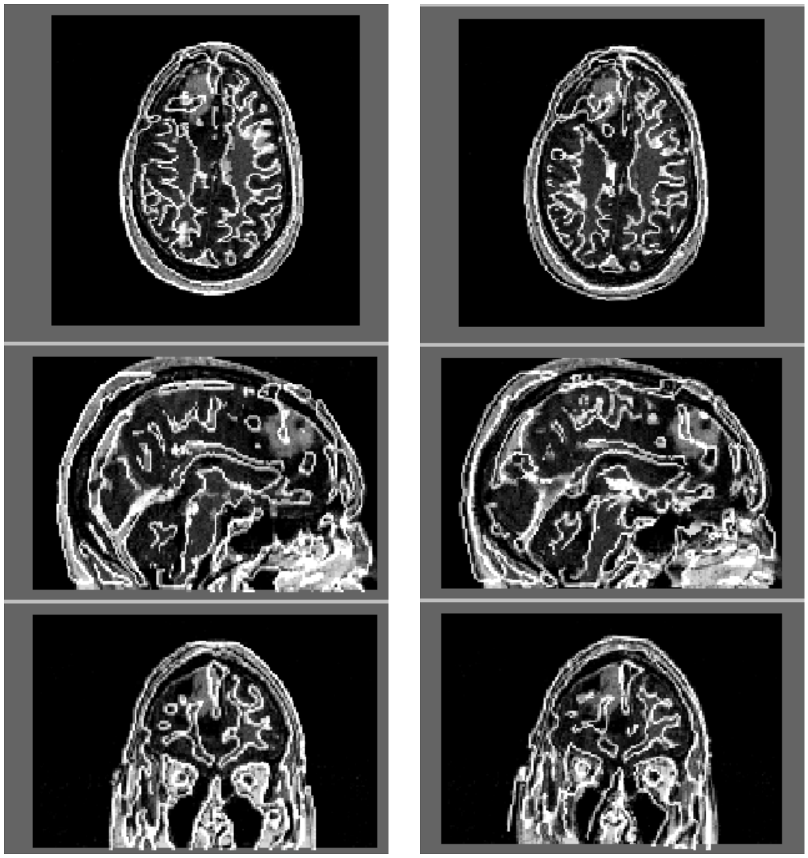
Table 2: Mean, median, and maximum errors (in mm) of the proposed image registration method when using the 13 multimodality images in the Vanderbilt dataset [39, 40]. Corresponding errors by the best available method [33, 40] are shown in the parentheses.

Data Set	**Mean**	**Median**	**Maximum**
1	3.00 (2.16)	2.81 (2.09)	6.28 (3.75)
2	2.52 (1.49)	2.29 (1.47)	6.28 (2.68)
3	2.93 (2.02)	2.41 (2.01)	5.24 (4.47)
4	1.83 (0.86)	1.70 (0.88)	4.11 (1.84)
5	1.59 (0.84)	1.27 (0.66)	4.45 (1.97)
6	2.22 (1.08)	2.11 (0.78)	5.56 (3.89)
7	2.17 (1.02)	2.12 (0.98)	4.65 (2.04)
8	3.05 (7.35)	3.12 (3.27)	6.74 (62.17)
9	3.76 (4.09)	3.74 (2.62)	9.44 (23.45)
10	3.26 (8.41)	3.13 (2.50)	8.99 (64.24)
11	4.37 (9.71)	4.16 (2.96)	10.90 (72.01)
12	2.77 (2.18)	2.82 (2.13)	4.81 (5.03)
13	3.52 (3.55)	3.72 (2.22)	7.63 (14.03)

was assumed an approximate registration and again refined through template matching.

Overall mean, median, and maximum errors of our method are 2.85 mm, 2.72 mm, and 6.54 mm, respectively, while those of the best method are 3.44 mm, 1.89 mm, and 20.12 mm, respectively. We see that our median error is worse than that of the best method, but our mean and max errors are better than those of the best method. Our method seems to be more stable in producing the results because it selects similar image areas in the matching while avoiding use of dissimilar areas. However, in cases where images have no partial intensity and/or geometric differences, use of whole images tends to produce more accurate results than use of templates.

It should be mentioned that the images prepared by West *et al.* [40] represented all pre-surgical images. Therefore, the image pairs used in the registration did not have partial geometric differences. We believe, if the images to be registered represent pre- and post-surgical images, then, due to partial geometric differences between images, our image registration method will perform better than methods that use whole images in the registrations. A simple demonstration of this is made below by using the pre- and post-surgical images shown in columns 3 and 4 of Figure 1. Using template matching, we obtain the registration result shown in Figure 3a, while using entire images, we obtain the registration result depicted in Figure 3b. Mutual information was used as the similarity measure in both cases. Since a gold standard is not available in this case, quantitative comparison of the



(a) (b)

Figure 3: (a) Registration of the images shown in columns 3 and 4 of Figure 1 using template matching. (b) Registration of the same images using the entire images. To compare the quality of registration by the two methods, the edges of the target image were overlaid with the reference image .

errors is not possible. After determining the registration parameters by the two methods, 3-D edges of the target image were determined and instead of resampling the target image to overlay the reference image, the target image edges were resampled to overlay the reference image. The cross-sections of the overlaid images are shown in Figure 3. Same cross-sections of the reference image are shown in Figures 3a and 3b. Corresponding cross-sections of the target image are also shown in these images. We observe differences in registration by the two methods. We see that edges in Figure 3a are more accurately aligned with structures in the reference image than edges in Figure 3b.

Computation time for each iteration of the proposed image registration on images of size $256 \times 256 \times 256$ voxels is about 2 minutes on a 400 MHz PC. If the images are rotationally aligned well, a single iteration is sufficient to bring the images into registration. However, if the images are not rotationally aligned well, a few iterations are needed to bring the images into registration.

7 Conclusions

Matching templates (subvolumes) rather than entire image volumes allows use of information in similar areas in images for matching. Template matching makes it possible to eliminate the inaccurate matches and use only the accurate ones in determination of the transformation parameters. Only four corresponding control points are sufficient to determine the translational and rotational differences between brain images of a patient. It is better to use four correspondences that are accurate than more than four correspondences, some of which are inaccurate. Therefore, in our method, first, about a dozen control point correspondences are determined by template matching. From among the dozen correspondences, the best four are selected in such a way that corresponding distances between corresponding point pairs in the images are the closest, and from the four correspondences, registration parameters are determined.

Our method is currently being used in a clinical setting to register multimodality brain images at Wallace-Kettering Neuroscience Institute, Kettering, Ohio. Our software requires only a few minutes to register multimodality images and its accuracy is a few millimeters for images in the Vanderbilt dataset [39].

Finally, it should be mentioned that if the images to be registered do not have partial intensity and/or geometric differences, use of entire images in the registration may produce more accurate results than use of templates. This is because the former uses entire image information to determine the registration parameters while the latter uses partial image information to do the same. The additional accuracy gained, however, will be at the cost of a slower registration speed.

Acknowledgements

The authors would like to thank Wallace-Kettering Neuroscience Institute, Kettering Medical Center, for partially funding this work and for providing some of the images used in this study. The images and the standard transformations used to determine the registration accuracy of our algorithm were provided as part of the project, "Evaluation of Retrospective Image Registration," National Institutes of Health, Project Number 1 R01 NS33926-01, Principal Investigator: J. Michael Fitzpatrick, Vanderbilt University, Nashville, TN.

References

[1] N. M. Alpert, J. F. Bradshaw, D. N. Kennedy, and J. A. Correia, The principal axis transformation: A method for image registration, *Journal of Nuclear Medicine,* vol. 31, no. 10, pp. 1717–1722, 1990.

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

[3] K. S. Arun, T. S. Huang, and S. D. Blostein, Least-squares fitting of two 3-D point sets, *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 9, no. 5, pp. 698–700, 1987.

[4] D. I. Barnea and H. F. Silverman, A class of algorithms for fast digital image registration, *IEEE Trans. Computers,* vol. 21, no. 2, pp. 179–186, 1972.

[5] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, Parametric correspondence and chamfer matching, *Proc. 5th Int'l J. Conf. Artificial Intelligence,* pp. 659–663, 1977.

[6] P. J. Besl and N. D. McKay, A method for registration of 3-D shapes, *IEEE Trans. Pattern Analysis Machine Intelligence,* vol. 14, no. 2, pp. 239–256, 1992.

[7] G. Borgefors, Hierarchical chamfer matching: A parametric edge matching algorithm, *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 10, no. 6, pp. 849–865, 1998.

[8] G. Borgefors, On digital distance transforms in three dimensions, *Computer Vision and Image Understanding,* vol. 64, pp. 420–433, 1996.

[9] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach,* Prentice-Hall International, p. 232, 1982.

[10] A. K. Dewdney, Analysis of steepest-descent image-matching algorithms, *Pattern Recognition,* vol. 10, pp. 31–39, 1978.

[11] D. Fischer, P. Kohlhepp, and F. Bulling, An evolutionary algorithm for the registration of 3-D surface representations, *Pattern Recognition,* vol. 32, no. 1, pp. 53–70, 1999.

[12] M. Fischler and R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[13] J. M. Fitzpatrick and J. B. West, A blinded evaluation and comparison of image registration methods, in *Empirical Evaluation Techniques in Computer Vision,* K. W. Bowyer and P. J. Phillips (eds.), IEEE Computer Society Press, pp. 12–27, 1998.

[14] J. M. Galvez and M. Canton, Normalization and shape recognition of three-dimensional objects by 3-D moments, *Pattern Recognition,* vol. 26, no. 5, pp. 667–682, 1993.

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

[16] A. Goshtasby, 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.

[17] A. Goshtasby, S. H. Gage, and J. F. Bartholic, A two-stage cross correlation approach to template matching, *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 6, no. 3, pp. 374–378, 1984.

[18] B. K. P. Horn, *Robot Vision,* McGraw Hill, pp. 130–132, 1986.

[19] M. K. Hu, Visual pattern recognition by moment invariants, *IRE Trans. Information Theory,* vol. 8, pp. 179–187, 1962.

[20] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, Comparing images using the Housdroff distance, *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 9, pp. 850–863, 1993.

[21] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Optimization by simulated annealing, *Science,* vol. 220, pp. 671–680, 1983.

[22] T. Kularatna, *Volumetric Image Registration Using Chamfer Matching,* M.S. Thesis, CSE Department, Wright State University, Dec. 1999.

[23] C-H Lo, H-S Don, 3-D moment forms: Their construction and application to object identification and positioning, *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 11, no. 10, pp. 1053–1055, 1989.

[24] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, Multimodality image registration by maximization of mutual information, *IEEE Trans. Medical Imaging,* vol. 16, no. 2, pp. 187–198, 1997.

[25] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. Hawkes, A comparison of similarity measures for use in 2-D and 3-D medical image registration, *IEEE Trans. Medical Imaging,* vol. 17, no. 4, pp. 586–595, 1998.

[26] C. A. Pelizzari, G. T. Y. Chen, D. R. Spelbring, R. R. Weichselbaum, and C. T. Chen, Accurate three-dimensional registration of CT, PET and/or MR images of the brain, *J. Computer Assisted Tomography,* vol. 13, pp. 20–26, 1989.

[27] W. K. Pratt, Correlation techniques of image registration, *IEEE Trans. Aerospace and Electronic Systems,* vol. 10, no. 3, pp. 353–358, 1974.

[28] W. H. Press, *Numerical Recipes in C,* Cambridge University Press, 1994.

[29] A. Rosenfeld and G. J. Vanderburg, Coarse-fine template matching, *IEEE Trans. Systems, Man, and Cybernetics,* pp. 104–107, Feb. 1977.

[30] M. Satter and A. Goshtasby, Registration of deformed images, *Image Registration Workshop,* Greenbelt, Maryland, pp. 221–231, Nov. 21–22, 1997.

[31] F. A. Sadjadi and E. L. Hall, Three-dimensional moment invariants, *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 2, no. 2, pp. 127–136, 1980.

[32] G. C. Stockman, S. Kopstein, and S. Bennet, Matching images to models for registration and object detection via clustering, *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 4, pp. 229–241, 1982.

[33] C. Studholme, D. L. G. Hill, and D. J. Hawkes, An overlap invariant entropy measure of 3-D medical image alignment, *Pattern Recognition,* vol. 32, pp. 71–86, 1999.

[34] M. Svedlow, C. D. McGillem, and P. E. Anuta, Experimental examination of similarity measures and preprocessing methods used for image registration, *Proc. Symp. Machine Processing of Remotely Sensed Data,* pp. 4A-9–17, 1976.

[35] M. Svedlow, C. D. McGillem, and P. E. Anuta, Image registration: Similarity measure and preprocessing method comparisons, *IEEE Trans. Aerospace and Electronic Systems,* vol. 14, no. 1, pp. 141–149, 1978.

[36] L. Yang, F. Albregtsen, and T. Taxt, Fast computation of three-dimensional geometric moments using a discrete divergence theorem and a generalization to higher dimensions, *Graphical Models and Image Processing,* vol. 59, no. 2, pp. 97–109, 1997.

[37] I. Vajda, *Theory of Statistical Inference and Information,* Kluwer, 1989.

[38] G. J. Vanderburg and A. Rosenfeld, Two-stage template matching, *IEEE Trans. Computers,* vol. 26, no. 4, pp. 384–393, 1977.

[39] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer, Jr., R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. A. van den Elsen, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. G. Hill, D. J. Hawkes, C. Studholme, J. B. A. Maintz, M. A. Viergever, G. Malandain, X. Pennec, M. E. Noz, G. Q. Maguire, Jr., M. Pollack, C. A. Pelizzari, R. A. Robb, D. Hanson, and R. P. Woods, Comparison and evaluation of retrospective intermodality brain image registration techniques, *J. Computer Assisted Tomography,* vol. 21, no. 4, pp. 554–566, 1997.

[40] J. West and J. M. Fitzpatrick,
http://cswww.vuse.vanderbilt.edu/∼image/registration/results.html
http://cswww.vuse.vanderbilt.edu/∼image/registration/reg_eval_html/ding.html
http://cswww.vuse.vanderbilt.edu/∼image/registration/reg_eval_html/rohlfing_1.html