



Volumetric image registration by template matching

Lijun Ding^a, Thisath Kularatna^a, Ardeshir Goshtasby^a, and Martin Satter^b

^aComputer Science and Engineering Dept., Wright State University, Dayton, OH 45435

^bNuclear Medicine/PET, Kettering Medical Center, Kettering, OH 45429

ABSTRACT

A template-matching approach to registration of volumetric images is described. The process automatically selects about a dozen highly detailed and unique templates (cubic or spherical subvolumes) from the target volume and locates the templates in the reference volume. The centroids of the “best” four correspondences are then used to determine the transformation matrix that resamples the target volume to overlay the reference volume. Different similarity measures used in template matching are discussed and preliminary results are presented. The proposed registration method produces a median error of 2.8 mm when registering Vanderbilt image data sets, with average registration time of 2.5 minutes on a 400 MHz PC.

Keywords: Image registration, volumetric image, template matching, mutual information, chamfer matching

1. INTRODUCTION

Image registration is the process of determining correspondence between all points in two images of the same scene. This operation is needed in many image analysis tasks. It is particularly valuable in medical imaging to fuse datasets from different imaging modalities to produce the maximum diagnostic sensitivity and specificity. For instance, registration of cerebral anatomic images from an MR scanner to the biochemical images from a PET scanner provides a unique data set useful in both diagnosis and therapy.

Images acquired from a single imaging modality (MR/MR, PET/PET, etc.) can be registered more easily than images from different modalities (MR/PET, MR/CT, etc.). To date, most work reported on volumetric image registration has determined registration parameters using information from the entire image volumes. The approach taken here automatically selects only similar information in the images for registration. We believe that a global optimization process that uses entire image volumes in the registration does not necessarily produce the best result when the images have intensity and/or geometric differences as is often the case. An approach that selects similar information from the images to determine the registration parameters should produce more accurate results than an approach that uses entire image information.

Consider the brain images shown in Figure 1. The far left column depicts the axial, sagittal, and coronal views from a contrast enhanced, T1-weighted MR scan adjacent to the same views of PET images of glucose metabolism of a patient with a brain tumor. Both images were acquired on the same day. Note the intensity pattern observed in the lesion from each modality does not correlate very well, while patterns in the normal brain are similar. In this instance, the MR scan reflects the high uptake of contrast material in the tumor (due to a leaky blood brain barrier) while the PET shows the tumor to be primarily low grade (correlating with relatively low glucose metabolism).

The far right two columns show the post-surgical and pre-surgical MR images of a patient with a brain tumor. In the post-surgical images, the 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 is similar while discarding information in dissimilar areas. Thus, templates are selected from one of the image volumes and their matches are located in the other. The template matching that generates the same registration parameters are kept and those that produce different parameters are discarded. In this manner, possible inaccuracies are detected and eliminated prior to calculation of the transformation function.

If the image volumes have small rotational differences, then the average distance between corresponding voxels in matching templates is much smaller than the average distance between corresponding voxels in the images themselves.

Send correspondence to A. Goshtasby, E-mail: agoshtas@cs.wright.edu

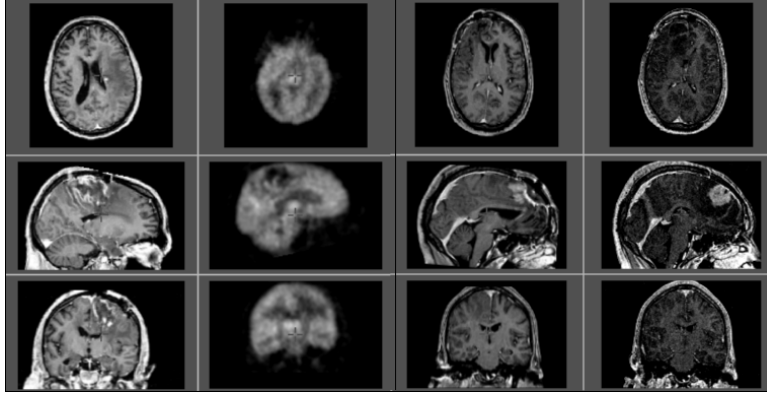


Figure 1. Rows 1, 2, and 3 respectively show axial, sagittal, and coronal cross sections of volumetric images. Columns 1 and 2 show MR and PET images of the a patient taken at about the same time, while columns 3 and 4 show MR images of a patient taken after and before a surgery.

Therefore, the process of template matching is less affected by small rotational differences between images than matching of the image volumes directly. When the images have small rotational differences, we can expect to find corresponding templates accurately. Using a set of corresponding (homologous) points in the images defined by the centroids of the corresponding templates, we can then determine the translational and rotational parameters required to bring the image volumes into correct alignment.

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. Having a number of templates from the target, we then describe the process of finding the corresponding templates in the reference image via template matching. Then, the selection criteria for keeping the best four matches (homologous points) is given, and finally a method for determining the transformation parameters from the homologous points is outlined.

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, the templates are ordered according to their sum of gradient magnitudes. There is no need to actually keep all the templates in a list. A small percentage (such as the top 5%) of highest-gradient templates is sufficient. Since many templates may overlap each other, those that overlap by more than 50% are removed from the list. Therefore, once a template is selected, all templates that have more than 50% overlap with it are removed from the list. Among the remaining templates, the p most unique ones are identified 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 an evidence that the template is unique. The less sharp this peak, the more similar the template to windows in its neighborhood. When a template that is not locally unique is selected for matching, it may match rather well to many windows in its neighborhood, making distinction between the correct match and the incorrect ones difficult.

Another way to characterize uniqueness is to use the eigenvalues of the inertia matrix of the template.^{15,34} A template that has all large eigenvalues represents a locally unique region 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 along a plane in its neighborhood. Unique templates enable determination of their correspondences with a high degree of accuracy.

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 points to determine the registration parameters.

Template matching involves determining similarities between a given template and windows of the same size in an image, and identifying the window that produces the highest similarity measure. In the following, first different similarity measures are reviewed and then search strategies to find the best-match position of a template in an image are given.

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 used for template matching. There isn't a single similarity measure that can produce best results on all images. Depending on the types of images provided, one similarity measure may work better than all others. 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 $m \times m \times m$. 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 f_2 . Similarity image s will be of size $(m - n + 1) \times (m - n + 1) \times (m - n + 1)$, defining the search area size.

3.1.1. Sum of Absolute Differences

Sum of absolute intensity differences is the Minkowski metric of order one⁹ and is defined by

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

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 dissimilarity) measure requires in the order of n^3 additions for each search position. An algorithm proposed by Barnea and Silverman⁴ speeds 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 because further computations 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 $O(n^3 m^3)$ additions, 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 $O(n^3 d^3)$ additions.

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 in 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.*³³ found that image gradients produce more accurate matches than the raw images when using the sum of absolute differences as the similarity measure. This result was independently confirmed 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}

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

where $x, y, z = 1, \dots, m-n+1$ 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 $s = -1$, change in intensities of the template will be the reverse of change in intensities of the window.

The computational complexity of the numerator of formula (2) 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 window position. Therefore, computation of the denominator will require in the order of $n^3 m^3$ multiplications, when m is much larger than n . Overall, if m is much larger than n the computational complexity of equation (2) is $O(n^3 m^3)$ multiplications. If template is about the same size as the image, and $d = m - n + 1$, computational complexity of the template matching using cross-correlation coefficient will be $O(m^3 d^3)$ 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 the 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.*³² found that cross-correlation coefficient produced more accurate matches than the sum of absolute differences when rotational difference between the images was negligible. 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 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 to compute sum of absolute differences or cross-correlation coefficient in template matching. In such a situation, if the images can be segmented to obtain similar binary structures, the surfaces can be used in the matching. The process of matching surface structures is known as *chamfer matching* and was first proposed by Barrow *et al.* in matching of aerial images.⁵ The idea, however, is general and can be used to match volumetric images also.⁸

Chamfer matching is a technique that uses the average distance between two surface 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 corresponding points is used as the similarity measure. An efficient algorithm for determining this 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, it is sufficient to prepare a distance image whose entries show distances of voxels in the image to structure points closest to them in that image. Then, as the template is shifted over the distance image, at each shift position the sum of image values in the distance image that fall 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 image structures for chamfer matching may be longer than the time needed to carry out 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. Note that the template should typically contain fewer edges than the image over which it is searched. 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 in order to achieve higher accuracy in matching.

Chamfer matching is similar to surface matching methods^{6,26} that align two surfaces based on 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 Kularatna²² in registration of single modality images found that in almost all cases, matching of image intensities produced more accurate results than matching of surface structures. A similar conclusion was reached by Fitzpatrick and West¹⁴ in matching of multimodality images. This is expected since surface-based techniques discard information from inside of objects and use information only from their surfaces in the matching.

3.1.4. Mutual Information

Suppose a template T and window W are available and we want to determine their similarity. Let's consider T and W are random variables and the probability that the intensity at a voxel in T be a is $P_T(a)$ and the probability that the intensity at a voxel in W be b is $P_W(b)$. Assuming that we overlay the template and the window, the probability that the 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 contain the same pattern, their intensities will be highly dependent and they will produce high joint probabilities. However, if the template and the window do not contain the same pattern, 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 is defined by^{24,35}

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

The joint probability $P_{TW}(a, b)$ for different values of a and b can be estimated from the histogram of intensities of corresponding voxels in the 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 position 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 with entry (a, b) showing the number of voxels in the template having intensity a while intensity at the 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 counts 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 a denotes the rows and b denotes the columns of $P_{TW}(a, b)$, we can estimate $P_T(a)$ and $P_W(b)$ from

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

and

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

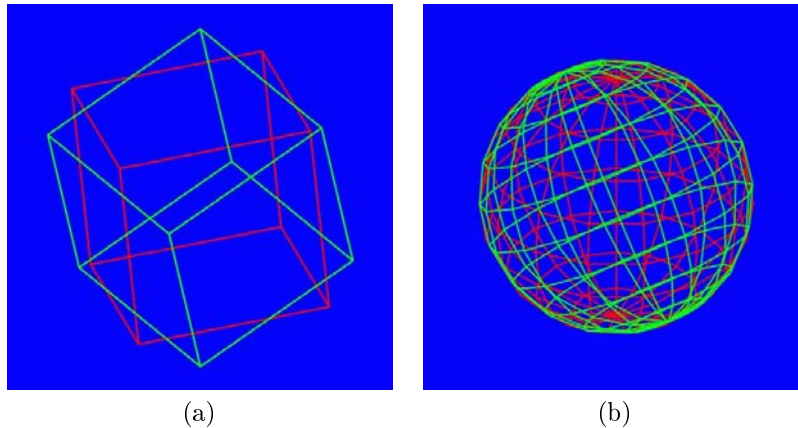


Figure 2. (a) A cubic template and a cubic window whose centers coincide. (b) A spherical template and a spherical window whose centers coincide.

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 one image correspond to homogeneous areas in the other image. However, noise in one or both images can quickly degrade the similarities. A study carried out by Penney *et al.*²⁵ in registration of same modality images found that mutual information does not perform as well as cross-correlation coefficient or the sum of absolute differences. We have found that slight image smoothing to reduce image noise improves the registration accuracy of single modality and multi-modality images when mutual information is used as the similarity measure.

The computational complexity of mutual information at each shift position is in the order of n^3 additions and 256^2 multiplications. 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 first.

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.^{19,23,30} 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, and mutual information ineffective in matching. Since invariant moments are independent of the orientation of a pattern, they can 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 exactly the same pattern in the template and a window in the image unless the rotational difference between the image and the template is a multiple of 90 degrees. This is demonstrated in Figure 2a. Therefore, even when the centers of a cubic template and a cubic window correspond to each other, since they cannot contain the same parts of a scene, the computed similarity between them may not be high enough to produce a correct match. To remedy this problem, we use spherical templates. If spherical templates are used, when the centers of the template and the 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, resulting in accurate matches.

The pqr th moment of template f_1 is defined by³⁴

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

and the pqr th order central moment of the template is defined by

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

where

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

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

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

The principal axes of a 3-D pattern is defined by the eigenvectors of the inertia matrix of the pattern, defined by¹⁵

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

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 with the principal axes of the window. To align the template and the 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 themselves. Once the rotational difference between the window and the template is removed, 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 in the same modality. This means use of invariant moments is limited to images of the same modality. Successful image alignment with invariant moments have been reported.¹

Assuming a highly detailed and unique template is selected in the target image, to find the window in the reference image that corresponds to it, at each shift position, the window in the reference image is aligned with the template using their principal axes. Once the template and the window are aligned, the similarity between them is determined using sum of absolute differences, cross-correlation coefficient, or mutual information. To uniquely align the template and the window, it is required that not only the template and the window have large eigenvalues, it is required that no two eigenvalues have about the same magnitude, otherwise wrong axes in the template and the window could be aligned, missing a 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 difference between the images, we expect to obtain rather accurate matches even when the template and the window are not perfectly aligned. Since the tolerance of these similarity measures to rotational difference between images increases as images become more blurred, the template-matching process can be carried out in two steps. First, by smoothing the template and the image they can be approximately aligned. Then, using the full-resolution template and image that are approximately aligned, a more accurate match can be obtained by search in a small neighborhood.

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 template and image are about the same size and $d = m - n + 1$. The constant multiplication factor involved in these formulas is rather large due to considerable computations involved in finding the eigenvalues and eigenvectors of the inertia matrices of the template and the 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 these similarity measures, many patterns produce the same similarity measure, increasing the probability of a mismatch. In cross-correlation coefficient, invariant moments, and distance measure, more unique similarity measures are obtained since rearranging intensities in the matching template and window will change the similarity measure, enabling to distinguish more dissimilar patterns from each other than the sum of absolute differences or the mutual information.

The positions of windows selected in the reference image for matching with a given template are discrete locations in the search area $(m - n + 1) \times (m - n + 1) \times (m - n + 1)$. The best-match position, however, can be determined with subpixel accuracy using the similarities at the best-match position and at positions adjacent to it. Centered at the best-match position we take a $3 \times 3 \times 3$ array of similarity measures and fit a tri-quadratic function to them and locate the maximum (minimum) of the function. A minimum is obtained when sum of absolute differences or geometric distance is used as the similarity measure, while a maximum is obtained when cross-correlation coefficient or mutual information is used. A tri-quadratic function is defined by

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

Parameters A – J are determined by the least squares method using the $3 \times 3 \times 3$ array of similarities and their coordinates. Once A – J are determined, the maximum (minimum) 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 linear equations for x , y , and z .

3.2. Search Techniques

To find the best-match position of a template in an image, a search can be performed in a manner to reach the solution faster. Coarse-to-fine search and optimization algorithms have been used.

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 the full-resolution images are obtained. At each step, possible match positions are determined with associated confidence levels and search at a higher resolution is performed only at high confidence levels. This technique has been used in 2-D template matching using sum of absolute differences^{29,36} 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.¹² 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 can find the solution in a small number of iterations. Speeding the search by an optimization process may actually involve more overhead than the time it saves.

4. SELECTING THE BEST CORRESPONDENCES

Registration by template matching has advantages over methods that use entire images. When entire images are used, noise and dissimilarity between images average out, influencing the registration result. However, if a number of correspondences is established between images through template matching, the inaccurate matches 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 use distances between pairs of corresponding 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 point pairs are the same the point pairs correspond to each other, the probability that two matches shift by exactly the same amount and in exactly the same direction is extremely rare. Therefore, for an overwhelming majority of the cases, this test will find accurate corresponding points in the images. For rare cases where this test fails, further testing is needed to distinguish the wrong matches from the correct ones.

To choose the best p correspondences from among the available P matches, we find distances between point pairs in each image and list the corresponding point pairs in the ascending order of the difference between their distances, and from the top of the list take p of them. Only four corresponding points are sufficient to determine the registration parameters. Usually it is a good idea to take more than four points and use the additional points to remove the possible outliers.^{13,20,31} Care should be taken to ensure that the four points do not lie on or close to a plane. Otherwise, the registration parameters may be impossible to find, or they may be inaccurate. Therefore, if more than four points are initially selected, the four that make the tetrahedron with the largest volume should be chosen to determine the registration parameters.

An alternative method for distinguishing the accurate matches from inaccurate ones is to examine the geometries of the components of the transformation function that map one image to another. Typically, a transformation for registration of volumetric images has three components:

$$X = f_x(x, y, z) \quad (13)$$

$$Y = f_y(x, y, z) \quad (14)$$

$$Z = f_z(x, y, z) \quad (15)$$

where (x, y, z) and (X, Y, Z) are coordinates of corresponding points in the reference and target images, respectively. Given a point in the reference image, the components of a transformation function determine the 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 will be linear functions. If the point correspondences are accurate, we will obtain a plane representing each component of the transformation. However, if the correspondences are inaccurate, the functions will not be linear. To detect the inaccurate correspondences, we will represent each component of the transformation by an elastic function, such as the rational Gaussian function,¹⁶ and observe its gradients. Small or zero gradients will be evidence that all correspondences are accurate. Large gradients centered at a window position in the reference image is an evidence that the template is not accurately located there. In this manner, accurate and inaccurate matches can be distinguished from each other.

5. COMPUTING THE REGISTRATION PARAMETERS

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

$$\mathbf{p}_i = \mathbf{R}\mathbf{P}_i + \mathbf{T} + \mathbf{N}_i, \quad (16)$$

where \mathbf{N}_i is the unknown inaccuracy associated with the i th correspondences. The translation vector \mathbf{T} and the rotation matrix \mathbf{R} can be determined by minimizing

$$E^2 = \sum_{i=1}^p \|\mathbf{p}_i - (\mathbf{R}\mathbf{P}_i + \mathbf{T})\|^2. \quad (17)$$

Efficient algorithms for determining \mathbf{R} and \mathbf{T} have been given by Arun *et al.*³ In this method, the rotation matrix is first obtained by minimizing

$$E_R^2 = \sum_{i=1}^p \|\mathbf{Q}_i - \mathbf{R}_i\mathbf{q}_i\|^2, \quad (18)$$

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 centroids of points in the target and reference images, respectively. Knowing the rotation matrix, the translation vector \mathbf{T} is determined from

$$\mathbf{T} = \mathbf{P} - \mathbf{R}\mathbf{p}. \quad (19)$$

Determination of these parameters based on quaternions is given by Faugeras and Herbert.¹¹

Table 1. Root-mean-squared distances in mm between centers of corresponding templates in ten MR brain image pairs after registration.

Data Set	Chamfer Matching		Mutual Information	
	Iteration 1	Iteration 2	Iteration 1	Iteration 2
1	4.8	1.5	0.56	0.13
2	2.6	1.9	0.18	0.13
3	3.2	3.0	0.17	0.17
4	2.3	1.5	1.8	0.77
5	2.7	2.5	0.55	0.39
6	2.0	1.9	0.57	0.09
7	3.3	2.4	0.45	0.33
8	1.8	1.0	0.54	0.23
9	4.7	2.0	0.25	0.07
10	3.3	2.1	0.49	0.09

6. RESULTS

Accuracy of template matching on ten sets of MR brain images, some pre- and post-surgery and some tracking changes in brain tumors over time are shown in Table 1. Each image pair was approximately aligned by an interactive procedure via mouse control in 10-20 seconds. 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 then used as corresponding points, the inaccurate correspondences were removed as outlined above, and the best four correspondences were used to find the registration parameters.

Entries in the table show root-mean-squared error in template matching. After determining the registration parameters from the correspondences, the target image was resampled to overlay the reference image. The distance between centers of templates in the target image and centers of corresponding windows in the reference image were determined and the root-mean-squared distance for each image pair was determined and reported in the table.

Since template matching accuracy is influenced by rotational difference between images, resulting from the approximate registration step, after determining the registration parameters, the target image was replaced with the resampled target image. The new target image was assumed to approximately register the reference image and the automatic template selection and template matching process was repeated to determine the new registration parameters. The obtained results are summarized in Table 1.

Registration errors using the Vanderbilt data produced a median of 2.8 mm. For a case-by-case examination of the results, please refer to cswwww.vuse.vanderbilt.edu/image/registration/reg_eval_html/ding.html.

Computation time for each iteration of our image registration algorithm using images of size $256 \times 256 \times 256$ is about 2 minutes on a 400 MHz PC. If the images are rotationally aligned well by the interactive step, a single iteration is sufficient to bring the images into registration. However, if the images are not rotationally aligned well, a few iterations are required to find the optimal registration.

7. CONCLUSIONS

Matching templates (subvolumes) rather than the entire image volumes enables using information about areas in the images that are similar and discarding information about image areas that are dissimilar. Template matching also allows discarding the inaccurate matches and using only the accurate ones to determine the registration parameters. Only four corresponding points are necessary to determine the translational and rotational differences between brain images of the same 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 correspondences are determined by template matching. From among the dozen matches, the best eight are selected using distances between corresponding point pairs in the images. Finally, from among the eight, the best four are selected in such a way that the obtained tetrahedron (in the reference image) has the largest volume, and from the four correspondences, the registration parameters are determined.

We believe our method can be utilized to register multimodality medical images in the clinical setting due to it's a) speed (a few minutes per registration) and b) accuracy. Experimental results using the Vanderbilt data show a median registration error of 2.8 mm.

ACKNOWLEDGMENTS

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. O. D. Faugeras and M. Herbert, A 3-D recognition and positioning system algorithm using geometrical matching between primitive surfaces, *Proc. Int'l J. Conf. Artificial Intelligence*, pp. 996–1002, Aug. 1983.
12. 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.
13. M. Fischler and 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.
14. 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.
15. 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.
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-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. 1997.
30. 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.
31. 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.
32. 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.
33. 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.
34. L. Yang, F. Albrechtsen, 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.
35. I. Vajda, *Theory of Statistical Inference and Information*, Kluwer, 1989.
36. G. J. Vanderburg and A. Rosenfeld, Two-stage template matching, *IEEE Trans. Computers*, vol. 26, no. 4, pp. 384–393, 1977.