Hybrid Image Registration based on Configural Matching of Scale-Invariant Salient Region Features

Xiaolei Huang[1], Yiyong Sun[2], Dimitris Metaxas[1], Frank Sauer[2], and Chenyang Xu[2]

[1]Division of Computer and Information Sciences, Rutgers University, Piscataway, NJ, USA
{xiaolei, dnm}@cs.rutgers.edu

[2] Imaging and Visualization Department, Siemens Corporate Research, Princeton, NJ, USA
{yiyong.sun, frank.sauer, chenyang.xu}@scr.siemens.com

Abstract

We present a novel method for aligning images under arbitrary poses, based on finding correspondences between image region features. In contrast with using purely feature-based or intensity-based methods, we adopt a hybrid method that integrates the merits of both approaches. Our method uses a small number of automatically extracted scale-invariant salient region features, whose interior intensities can be matched using robust similarity measures. While previous techniques have primarily focused on finding correspondences between individual features, we emphasize the importance of geometric configural constraints in preserving global consistency of individual matches and thus eliminating false feature matches. Our matching algorithm consists of two steps: region component matching (RCPM) and region configural matching (RCFM), respectively. The first step finds correspondences between individual region features. The second step detects a joint correspondence between multiple pairs of salient region features using a generalized Expectation-Maximization framework. The resulting joint correspondence is then used to recover the optimal transformation parameters. We applied our method to registering a pair of aerial images and several pairs of single and multiple modality medical images with promising results. The preliminary results, in particular, showed that the proposed method has excellent robustness to image noise, intensity change and inhomogeneity, appearance and disappearance of structures, as well as partial matching.

1. Introduction

Image registration aims to spatially align one image to another. For that purpose, parameters of a global transformation model, such as rigid, affine or projective, are to be recovered to geometrically transform a *moving* image to achieve high spatial correspondence with a *fixed* image. The problem has been studied in various contexts due to its significance in a wide range of areas, including medical image fusion, remote sensing, recognition, tracking, mosaicing, and so on.

Existing methods for image registration can largely be classified into three categories: feature-based methods, intensity-based methods, and hybrid methods that integrate the previous two. Traditional feature-based methods use sparse geometric features such as points [15], curves, and/or surface patches [3, 12], and their correspondences to compute an optimal transformation. These methods are relatively fast. However, the main critiques of this type of methods in the literature are the robustness of feature extraction, the accuracy of feature correspondences, and the frequent need of user interaction. Intensity-based registration methods [17, 5] operate directly on the intensity values from the full image content, without prior feature extraction. These methods have attracted much attention in recent years since they can be made fully automatic and can be used for multi-modality image matching by utilizing appropriate similarity measures. However, these methods tend to have high computational cost due to the need for optimization on complex, non-convex energy functions. In addition, they require the poses of two input images be close enough to converge to a local optimum. Furthermore, they often perform poorly when partial matching is required. Recently, several hybrid methods are proposed that integrate the merits of both feature-based and intensity-based methods [13, 8, 10]. Most of them focus on incorporating user provided or automatically extracted geometric feature constraints into the intensity-based energy functionals to achieve smoother and faster optimization.

Despite the vast efforts, however, several hard problems in registration still remain. First, dealing with structure appearing/dissappearing between two images is still challenging. For instance, tumor growth/shrinkage in medical images acquired in the clinical tracking of treatment, trees/shadows or construction in aerial images taken at different times, and occlusion in other natural images often lead to significant differences in local image appearance (see Figs. 1, 7). Second, it is still difficult to match images acquired by sensors of different modalities in general, since

different sensors, such as MRI, CT or PET, may produce very dissimilar images of the same scene. The relationship between the intensities of the matching pixels is often complex and not known *a priori*. Image noise and intensity inhomogeneity also add to this complexity. Last, but not least, given two input images under arbitrary poses, recovering the globally optimal transformation efficiently is a hard problem due to the large parameter search space. To tackle these problems, the integration of both feature-based and intensity-based methods is very attractive since they are of complementary nature. While intensity-based methods are superior in multi-modal image matching and have better robustness to image noise and inhomogeneity, the feature-based methods are more natural to handle the structure appearing/dissappearing problem, occlusion, and partial matching as well as to align images despite of their initial poses.

In this paper, we propose a new hybrid image registration method, which is based on matching a small number of scale-invariant salient region features. Rather than using traditional geometric features such as curvature extrema points, curves/surface patches, the image alignment in our approach is driven directly by image intensities within automatically extracted salient regions. The overall approach is depicted in Fig. 1. First, on both the fixed and moving images, salient region features are selected, using an entropy-based detector, as those areas (each associated with a best scale) with the highest local saliency in both spatial and scale spaces (see Fig. 1, I.a-d). Then a *region component matching* (RCPM) step is used to determine the likelihood of each hypothesized fixed-moving pairing of two region features. The likelihood of each pairing is measured by the normalized mutual information between the two regions. The result of this step is a total ordering of the likelihoods of all hypotheses about individual feature matches. Due to image noise or intensity changes, the top matches from this result often contain an unpredictable portion of outliers (i.e., mismatches), whose effects can only be partially alleviated by the use of robust estimation techniques. In the literature, the global one-to-one correspondence constraint [2, 4] has been widely used. However, in the presence of unmatchable features or in the situation of partial matching, this global constraint is neither sufficient nor valid. To address these limitations, we emphasize the importance of the geometric configural constraints in preserving the global consistency of individual matches. Utilizing the top individual feature correspondence candidates from the RCPM step, we further design a *region configural matching* (RCFM) step in which we detect a joint correspondence between multiple pairs of salient region features (see Fig. 1, II.c-d). The strict geometric constraints imposed by the joint correspondence make the algorithm very effective in pruning false feature matches. The combinatorial complexity associated with detecting joint correspondences is addressed in an efficient manner by using one feature pair correspondence as a minimal base (see Fig. 1, II.a-b), then incrementally add to the base new feature pairs using an Expectation-Maximization algorithm. The likelihood of each hypothesized joint correspondence is always measured based on the global "alignedness" between the fixed image and the transformed moving image, given the transformation computed from the hypothesis. This allows convergence to the globally optimal transformation parameters. Various experiments on registering aerial images and medical images of single and multiple modalities demonstrate the effectiveness of the proposed method both quantitatively and qualitatively.

1.1. Previous Work

The proposed image registration method is largely inspired by the pioneering works from the object recognition literature [6, 7, 9]. From their works, we learned two important aspects that would be beneficial when used in image registration. The first aspect is the use of scale-invariant *region* features. In [6], objects are modeled as flexible constellations of regions (parts) in order to learn and recognize object class models. An entropy-based feature detector [9] is used to select region features that have complex intensity distributions and are stable in both spatial and scale spaces. When adapting the idea of region features to solve image registration problems, suitable and robust similarity measures need to be defined between region intensity values, to deal with multi-modal matching, image noise, and intensity inhomogeneity. The second aspect is the importance of geometric configural constraints in robust feature matching. In [7], the role of geometric constraints in object recognition is studied in depth using edge and other geometric features, and an *interpretation tree* (IT) algorithm is developed to search for globally consistent feature correspondences. In this paper, we present a new method of implementing the geometric configural matching. Compared to the interpretation tree search algorithms whose best-case and worst-case complexities can be significantly different, our method has a very predictable low computational cost and has the best-case and worst-case complexities on the same order.

The remainder of the paper is organized as follows. In section 2, we describe the salient region feature detector. In section 3, we present our region component and configural matching algorithms for registration. Experimental results on both aerial and medical images are demonstrated in section 4. We conclude with discussion in section 5.

2. Scale-Invariant Salient Region Features

The line of research on feature-based image matching has long been restrained by the question: what features to use? An interesting feature selection criterion was proposed for tracking under occlusion and disocclusion situations in [14].

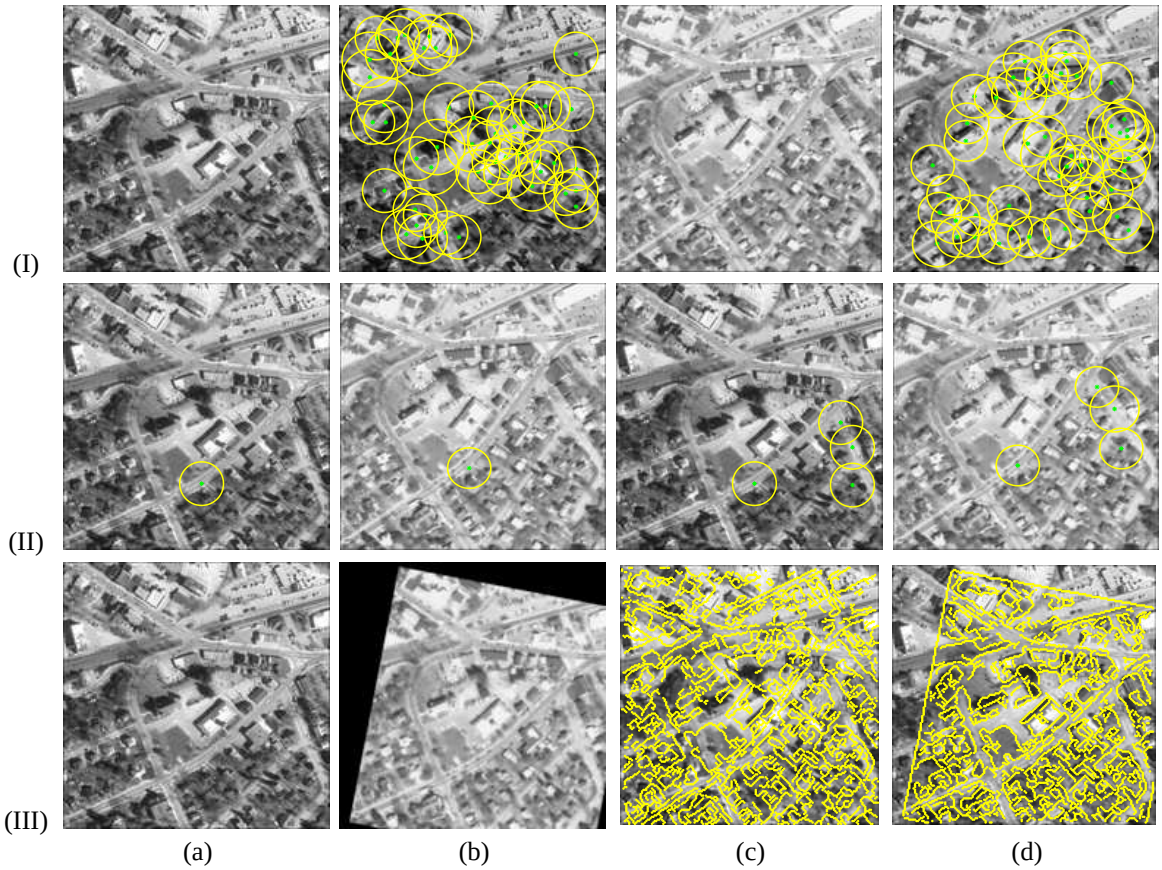

Figure 1: The registration method based on matching scale-invariant salient region features. (I.a) The fixed image I_f. (I.b) Salient region features (shown as yellow circles) detected on I_f. (I.c) The moving image I_m. (I.d) Salient region features detected on I_m. (II.a-b) The first corresponding feature pair chosen. (II.c-d) The corresponding feature pairs chosen by the algorithm upon convergence. (III.a-b) Registration result: (III.a) the fixed image I_f, and (III.b) the transformed moving image I_t based on the transformation parameters recovered using the chosen feature correspondences. (III.c-d) Comparison of the edge superimposed maps: (III.c) edges (in yellow) from the original moving image I_m superimposed on fixed image I_f, and (III.d) edges from the transformed moving image I_t superimposed on fixed image I_f.

The criterion states that the right features for tracking are exactly those that make the tracker work best. Applying similar reasoning, we believe that good features for image registration should be those that are "unique" or "rare". The uniqueness or rarity of a feature we refer here is in the context of correspondence, i.e., given a feature from one image, whether the likelihood of having multiple corresponding features on the matching image is low, not in the context of its uniqueness or rarity in occurrence within the same image. For example, in the use of point features for image matching, the traditional intuition and argument is that pixels in homogeneous regions (similarly, points with low curvatures on curve or surface segments) tend to be ambiguous in correspondence and should either not be chosen as the preferred feature point or weighted less important during the matching process [16, 1]. We argue, however, that these popular beliefs are only correct in a relative sense and that the "uniqueness" of a feature is closely related to its associated scale. At a smaller scale, edge points, corner points, or points with high curvature appear to be more unique than others. At a medium or larger scale, points in homogeneous regions or with low curvature begin to appear unique as well. Medial axis points of a shape or a homogeneous region are examples of these type of points that are unique at the scale they are associated with. We believe that every point regardless of their local characteristics (edgeness, cornerness, medialness, curvature, etc.) in the image can be made unique if a proper scale and its neighborhood is selected to calculate the feature[1]. One pictorial example of this point of view is demonstrated in Fig. 2.

Thus motivated, we seek to use scale-invariant region features as the basis for our proposed registration method. In [9], a salient region feature detector is proposed. The salient regions are found using an entropy-based detector, which aims to select regions with highest local saliency in both spatial and scale spaces. For each pixel $\mathbf{x}$ on an image, a prob-

[1]Note that we are not the first to exploit this observation for image registration. For instance, in [13], promising results have been obtained recently for non-rigid brain image registration using an attribute vector of geometric moment invariants at different scales.

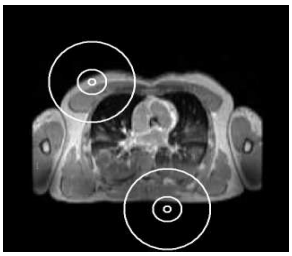

Figure 2: Demonstrating our belief that every point in the image can be made unique if a proper scale of its neighborhood is selected to calculate the feature. (Inner most circle) Locally at a small scale, the point neighborhood appear homogeneous. (Middle circle) At a larger scale, the point neighborhood begins to appear unique. (Large Circle) At a scale that is large enough, every point appears unique based on the characteristics of its neighborhood.

ability density function (PDF) $p(s, \mathbf{x})$ is computed from the intensities in a circular region of certain scale described by a radius s centered at $\mathbf{x}$. The local differential entropy of the region is defined by:

$$\mathcal{H}(s, \mathbf{x}) = -\int_{\mathcal{R}} p_i(s, \mathbf{x}) \log_2 p_i(s, \mathbf{x}) di$$

where i takes on values in the set of possible intensity values. The best scale $S_{\mathbf{x}}$ for the region centered at $\mathbf{x}$ is selected as the one that maximizes the local entropy: $S_{\mathbf{x}} = \text{argmax}_s \mathcal{H}(s, \mathbf{x})$. Then the saliency value, $\mathcal{A}(S_{\mathbf{x}}, \mathbf{x})$, for the region with the best scale is defined by the extrema entropy value, weighted by the best scale and a differential self-similarity measure in the scale space:

$$\mathcal{A}(S_{\mathbf{x}}, \mathbf{x}) = \mathcal{H}(S_{\mathbf{x}}, \mathbf{x}) \cdot S_{\mathbf{x}} \cdot \int_{\mathcal{R}} \left\| \frac{\partial}{\partial s} p_i(s, \mathbf{x}) \big|_{S_{\mathbf{x}}} \right\| di$$

Since the saliency metric is applicable over both spatial and scale spaces, the saliency values of region features at different locations and scales are comparable.

For the proposed registration method, we apply the following steps to pick a low number N ($N < 100$ for all our experiments) of salient region features (each defined by its center and the best scale):

- For each pixel location $\mathbf{x}$, compute the best scale $S_{\mathbf{x}}$ of the region centered at it, and its saliency value $\mathcal{A}(S_{\mathbf{x}}, \mathbf{x})$.
- Identify the pixels with local maxima in saliency values. Then the salient regions of interest are those that are centered at these pixels and have the best scales.
- Among the local maxima salient regions, pick the N most salient ones as region features for the image.

One of the main advantages of the salient region features is that they are theoretically invariant to rotation, translation and scale. We also quantitatively validate the invariance properties in section 4.1. Some examples on the extracted salient regions are shown in Fig. 1(I.a-d) and in Fig. 4(II.a-d).

3. The Salient Region based Registration Algorithm

Once we have extracted the salient region features from both the fixed and moving images, the alignment of the two images is achieved by finding a robust joint correspondence between multiple pairs of region features. This joint correspondence is then used to estimate the parameters of a desired transformation model. In this paper, we consider the 2D similarity transformation. This transformation can be described by four parameters: $(t_x, t_y, \sigma, \theta)$, where t_x, t_y are the translation along x and y directions respectively, σ is the isotropic scaling factor, and θ is the rotation angle.

Several notations are introduced as follows:

- I_f is the fixed image, I_m is the moving image, and I_t is the transformed moving image. We aim to recover the parameters of a similarity transformation that geometrically transforms the moving image to be aligned with the fixed image.
- Suppose N_f salient region features are detected on I_f, and N_m features on I_m.
- $C_{i,j}$ denotes the hypothesized correspondence between the ith region feature on I_f and the jth feature on I_m. Here $(i, j) \in [1, N_f] \times [1, N_m]$.
- $C_{i_1,j_1} \cap C_{i_2,j_2} \cap ... \cap C_{i_k,j_k} ...$ denotes a hypothesized joint correspondence between multiple region feature pairs: i_1th region on I_f corresponds to j_1th region on I_m, i_kth region on I_f corresponds to j_kth region on I_m, etc.

3.1. Region Component Matching (RCPM)

In the RCPM step, we measure the likelihood of each hypothesized correspondence between a region feature from I_f and a region feature from I_m, respectively. That is to say, we want to measure the likelihood $\mathcal{L}_{local}(C_{i,j})$ for each individual feature correspondence hypothesis $C_{i,j}$. We can then acquire a total ordering of these hypotheses according to their likelihoods.

We define the likelihood to be proportional to the similarity between the interior intensities of the two salient regions involved. Let us denote the ith region on I_f as A, and the jth region on I_m as B. Before measuring their intensity similarity, we first normalize their scales by supersampling (using bicubic interpolation) the smaller region to match the scale of the larger region. This also leads to scale-invariant matching. The translation invariance is intrinsic by aligning the two region centers. To further achieve rotation invariance, we sample the parameter space for rotation sparsely [2], and use the largest similarity value over all possible angles

[2] Typically, the rotation angles are sampled uniformly between $[-\pi, \pi)$ at an interval $\pi/36$.

as the similarity between the two regions. The similarity measure we use is a normalized form of mutual information, the Entropy Correlation Coefficient (ECC) [11]. Such metric has been proven robust in the literature in dealing with multi-modal image matching, image noise and intensity inhomogeneity.

Formally, the likelihood of a correspondence hypothesis $C_{i,j}$ is defined as:

$$\mathcal{L}_{local}(C_{i,j}) = \max_{\theta} ECC(A, B^{\theta})$$

where B^{θ} is the scale-normalized region B after rotating angle θ. The Entropy Correlation Coefficient (ECC) between the two regions is defined by:

$$ECC(A, B^{\theta}) = 2 - \frac{2\mathcal{H}(A, B^{\theta})}{\mathcal{H}(A) + \mathcal{H}(B^{\theta})}$$

where $\mathcal{H}$ indicates the joint or marginal differential entropy of the intensity value random variables of the two regions. Given two inputs u and v, the value of $ECC(u, v)$ has the following properties: $ECC(u, v)$ is scaled to $(0, 1)$, such that 0 indicates full independence and 1 complete dependence between the two inputs. Furthermore, $ECC(u, v)$ increases almost linearly when the relationship between u and v varies from full independence to complete dependence, which makes it an attractive measure of the likelihood that u corresponds to v.

Using this ECC definition, the likelihood values of all feature correspondence hypotheses $C_{i,j}$, where $(i, j) \in [1, N_f] \times [1, N_m]$, are comparable regardless of the scales of the region features. Thus we are able to sort these hypotheses in the order of descending likelihood. We then choose the top M such hypotheses to be used in the next configural matching step to extract a globally consistent joint correspondence. Here we make the assumption that there will be at least 2 valid feature correspondences among the top M candidates. From our extensive experiments, we found this assumption to be fairly reasonable with typical values of M between $20 \sim 40$.

The RCPM step also generates useful information regarding the transformation to align the two images, based on the purely local region-based matching. For instance, given a high likelihood correspondence between a region A on I_f and a region B on I_m, we can estimate the scaling factor by: $\sigma = \frac{s_A}{s_B}$, where s_A and s_B are the scales of the two regions respectively. The rotation angle can be estimated as: $\theta = \text{argmax}_{\theta'} ECC(A, B^{\theta'})$. And the translation can also be estimated by the displacement between the center of the region A and the center of the region B after rotation and scaling. These estimates are associated with the related feature correspondence hypothesis, to provide the initial estimate for the transformation in the next region configural matching step.

As a result of the RCPM step, we have a total ordering of the individual feature correspondence hypotheses. In addition, based on the top M hypothesis $C_{i,j}$, we have a transformation parameter estimate: $(t_x, t_y, \sigma, \theta)^{C_{i,j}}$. As an example, we show the top 5 region feature correspondence hypotheses for the pair of aerial images in Fig. 3. From the results, one can see that the RCPM step is able to extract good individual feature matches based on local region intensity-based matching. In the next configural matching step, we will demonstrate the use of geometric constraints to pick out the true correspondences (e.g., Fig. 3, I-III, V) and prune the outliers (e.g., Fig. 3, IV).

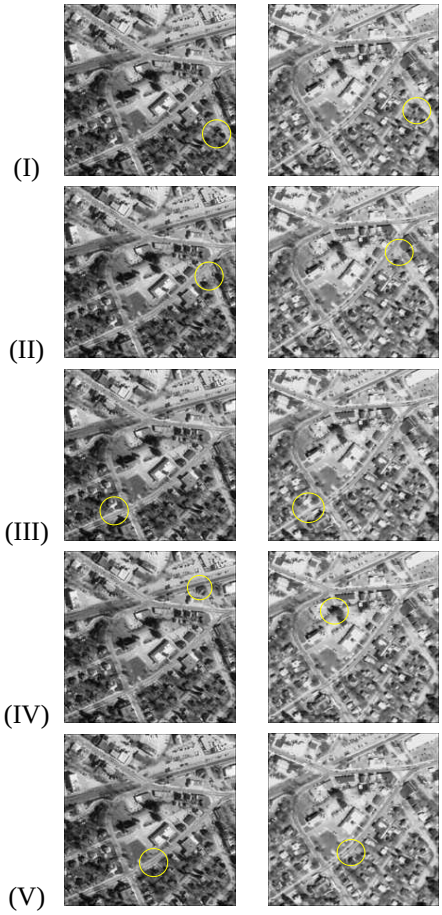

Figure 3: The top five candidate region feature correspondences computed by the region component matching (RCPM) step. The result is shown for the pair of aerial images in Fig. 1.

Note that the RCPM step has the most complexity of our entire registration method, since it has $N_f \times N_m$ hypothesis testings, and a total ordering of their likelihoods are pursued. However, because the number of region features, N_f and N_m, are low, the algorithm is still computationally efficient.

3.2. Region Configural Matching (RCFM)

In the RCFM step, we aim to detect a joint correspondence $C_{i_1,j_1} \cap C_{i_2,j_2} \cap ... \cap C_{i_k,j_k} ...$ between multiple pairs of region features, which results in the maximum likelihood in terms of global image "alignedness". The intuition behind the con-

figural matching is that, while false matches are very likely to arise when we search for individual local feature correspondences, the likelihood of a global geometrically consistent joint correspondence between multiple feature pairs being false is very low due to the strict geometric configuration constraints imposed by the joint correspondence.

We measure the likelihood of a hypothesized joint correspondence with n feature pairs using the ECC measure between the overlapping portions of the fixed image I_f and the transformed moving image $T_n(I_m)$. Here the transformation $I_t = T_n$ is estimated from all feature pairs contained by current hypothesis. This can be written as:

$$\mathcal{L}_{global}(C_{i_1,j_1} \cap C_{i_2,j_2} \cap ... \cap C_{i_n,j_n}) = ECC(T_n(I_m), I_f), \quad (i_k, j_k) \in [1, N_f] \times [1, N_m], k = 1, ..., n \tag{1}$$

This likelihood measures the global image "alignedness" under current hypothesis. In the end, we want to find a joint correspondence that has the maximum likelihood, while containing adequate number of feature pairs (typically a few) to recover the parameters of a similarity transformation.

To address the combinatorial complexity in detecting the joint correspondence, we first compute a minimal correspondence base of l feature pairs and get an initial estimation of the transformation. As shown in section 3.1, one correspondence between a pair of region features is sufficient to derive a transformation estimate, i.e., $l = 1$. To choose this first correspondence, we measure $\mathcal{L}_{global}(C_{i,j})$ for each individual feature match among the top M hypothesized correspondences resulted from the RCPM step. Using Equation 1, the parameters of T_l are $(t_x, t_y, \sigma, \theta)^{C_{i,j}}$ when measuring the likelihood of $C_{i,j}$. Then the first feature pair in the minimal correspondence base is the correspondence yielding the maximum likelihood, i.e.,

$$C_{i_1,j_1} = \underset{C_{i,j}}{\operatorname{argmax}}\, \mathcal{L}_{global}(C_{i,j})$$

To allow converging to a globally optimal solution, we further use a generalized Expectation-Maximization (EM) algorithm to incrementally add in new feature pairs to the joint correspondence base, while refining the center locations of the corresponding features. The generalized EM algorithm is described as follows:

1. Let current joint correspondence be $C = (C_{i_1,j_1} \cap ...C_{i_l,j_l})$. Locally refine the region feature centers in C in sub-pixel accuracy to achieve better matching, and use the refined corresponding region centers to estimate a current transformation T.
2. **E-step:** For each feature pair $C_{i,j}$ that is in the top M individual matches, but not in the current joint correspondence C, estimate the likelihood of this feature pair being a valid correspondence in terms of global consistency as $\mathcal{L}_{global}(C \cap C_{i,j})$, $C_{i,j} \notin C$.
3. **M-step:** Choose the new feature correspondence $C_{\hat{i},\hat{j}}$ that has the maximum likelihood. We also require the addition of $C_{\hat{i},\hat{j}}$ increasing the global image "alignedness".

   If $\mathcal{L}_{global}(C \cap C_{\hat{i},\hat{j}}) > \mathcal{L}_{global}(C)$

   Then

   a Let the new joint correspondence be $C = (C \cap C_{\hat{i},\hat{j}})$.

   b Locally refine the centers of the region features in the joint correspondence in sub-pixel accuracy to achieve better matching.

   c Re-compute the transformation T using the new joint correspondence.

   d Repeat EM steps 2-3.

   Else Output current transformation T as the converged transformation to align the fixed image I_f and the moving image I_m.

For the aerial image example, the feature pair shown in Fig. 1(II.a-b) is chosen to be the first feature pair in the minimal correspondence base. Note that this first pair in the RCFM step is not necessarily the same as the top feature pair resulted from the RCPM step, since different criteria are used to determine the likelihoods for ranking purposes. In fact, the first pair chosen by the RCFM step in Fig. 1(II.a-b) is the 5th feature pair in the RCPM step (see Fig. 3, V), because the transformation estimated from this feature pair gives rise to the maximum global image "alignedness". All the feature pairs in the final converged joint correspondence are shown in Fig. 1(II.c-d). Based on these correspondences, a similarity transformation is recovered and the registered image pair (i.e., the fixed image and the transformed moving image) is shown in Fig. 1(III.a-b).

Having at most M iterations, our RCFM step is very efficient. Two key points contribute to this efficiency: First, pick a minimal correspondence base with only one feature pair; Second, use the EM algorithm to add in new feature pairs incrementally, thus enabling the converged joint correspondence to include as many good feature pairs as possible, while keeping a minimal complexity.

4. Experiments

In this section, we present both the quantitative and the qualitative results of applying our image registration method on several simulated and real images.

4.1. Quantitative Results on Simulated Moving Images

In order to quantitatively validate the robustness, accuracy, and efficiency of the proposed method, we conduct a series of controlled experiments using a pair of brain images with

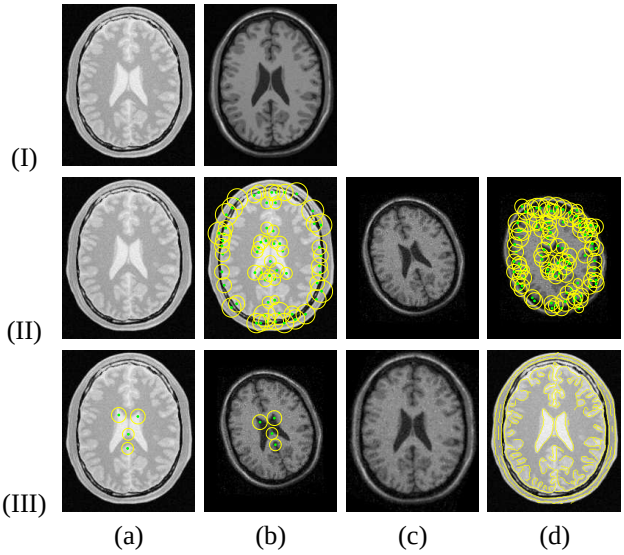

Figure 4: Registration on the pair of brain images used in the simulation experiment. (I.a) Original PD-weighted MR brain image. (I.b) Original T1-weighted MR brain image. (II.a) The fixed image I_f. (II.b) Salient region features on I_f. (II.c) The moving image I_m. (II.d) Salient region features on I_m. (III.a-b) The feature pairs in the joint correspondence chosen by the algorithm upon convergence. (III.c) The transformed moving image I_t. (III.d) The edge superimposed map after registration: edges from I_t (in red) superimposed on fixed image I_f.

the moving image simulated from a known transform. The first image is a PD (proton density) weighted MR brain image (see Fig. 4, I.a), and the second image is a T1 weighted MR brain image (see Fig. 4, I.b). The two images are originally registered, and the size of the images is 217×181.

In our first controlled experiment, we study the invariance properties of our method to scaling, rotation, and translation. We use the PD image as the fixed image, then simulate different moving images by artificially transform the T1 image with controlled parameters. The parameters are chosen according to the following four cases:

1. Case 1 studies the invariance to scaling. To this end, we fix the translation ($t_x = 0, t_y = 0$) and rotation ($\theta = 0$), but vary the scale factor σ in the range $[0.5, 1.5]$.

2. Case 2 studies the invariance to rotation. We fix the translation ($t_x = 0, t_y = 0$) and scaling factor ($\sigma = 1$), but vary the rotation angle in the range $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

3. Case 3 studies the invariance to translation. Here only the translation parameters t_x, t_y are varied in the range $[-50, 50]$.

4. Case 4 studies the combined effect of the transformation parameters by varying all parameters simultaneously: t_x, t_y in the range $[-50, 50]$, σ in the range $[0.5, 1.5]$, and θ in the range $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

	correctness	error	time
Case 1	98%	(0.9, 1.1, 0.027, 0.0)	138 s
Case 2	100%	(0.5, 0.6, 0.009, 1.5)	155 s
Case 3	100%	(0.2, 0.4, 0.000, 0.0)	155 s
Case 4	94%	(1.4, 1.7, 0.031, 2.1)	150 s

Table 1: Quantitative validation of the invariance properties of the method. For each case, the percentage of correct registration (correctness), the average error in recovered transformation parameters (error), and the average execution time for one trial (time) are given. The given errors are in the format: $(t_x, t_y, \sigma, \theta)$, where translation errors t_x and t_y are in pixels, rotation angle errors are in degrees, and the scaling errors are given relative to the original image scale. The times are given in seconds.

In each case, we generate 50 simulated moving images. Then we apply our registration algorithm to register the fixed image with each simulated moving image respectively. Since we know the ground truth transformation that was used to simulate each moving image, we can compare these ground truth with the recovered transformation parameters by our method. Three statistical performance measures are computed from the study and the results are listed in Table 1. The first measure is the *percentage of correctness* (correctness). In a registration trial, if the recovered transformation is sufficiently close to the ground truth [3], this trial results in a correct registration, otherwise, it is taken as a false registration case. The second measure is the *average error* (error). This measure gives the average error (i.e., difference) of the recovered transformation parameters from the ground truth. It reflects the accuracy and convergence property of our registration method. The last measure is the *average execution time* (time) for one trial of registering a pair of fixed and moving images. Note that our method is currently implemented in Matlab with several functions written in C++ and that all the experiments are conducted on a 2GHz PC workstation.

In the second controlled experiment, we study the robustness of the method to image noise. We use the original PD image as the fixed image, then generate test moving images by adding different levels of Gaussian noise to the original T1 image, and transforming the noise corrupted images according to random transformations. The Gaussian noise we add has zero mean with standard deviation λ. In Table 2, we show the three performance measures for three test cases. The three cases differ by the range of the standard deviation of the Gaussian noise added. (All possible values for the standard deviation are between $[0, 255]$). For each case,

[3] We consider the recovered transformation correct if its difference from the ground truth is less than a pre-defined error threshold. Typically, we set the threshold as follows: scale error less than 0.05, rotation angle error less than 5 degrees, translation error in x direction less than $D_x/50$, and translation error in y direction less than $D_y/50$, where D_x, D_y are the dimensions of the image along x and y directions, respectively.

range of λ	correctness	error	time
[5, 10]	100%	(0.3, 0.6, 0.007, 0.4)	142 s
[10, 20]	97%	(0.7, 0.9, 0.006, 1.2)	142 s
[20, 30]	90%	(0.9, 1.3, 0.009, 2.4)	144 s

Table 2: Quantitative simulation study of the performance of the method when images are corrupted by different levels of Gaussian noise. Three different cases are shown in three rows. The cases differ by the range of the standard deviation λ of the Gaussian noise added. For each case, three statistical measures are given in the same format as in Table 1.

30 noise corrupted T1 images are generated and randomly transformed, where the transformation parameters vary in the same ranges as in the first controlled experiment. From the results, one can see that, the method is quite robust to high levels of noise. This is partly due to the stability of the entropy-based region feature detector and the robustness of the intensity-based Entropy Correlation Coefficient (ECC) similarity measure. It is also due to the fact that our algorithm requires only a small number of good matched features to register the images. One pictorial example selected among all simulated experiments is shown in Fig. 4(II-III). In this example, the moving image I_m (see Fig. 4, II.c) is generated by adding Gaussian noise with zero mean, standard deviation 25 to the original T1-weighted image, then scaling down the image by 20%, and rotating by 20 degrees.

4.2. Qualitative Results on Real images

Experiments with the simulated moving images in the previous section provide a quantitative study on the performance of our registration method. Real world images often have significant levels of noise and intensity inhomogeneity. Furthermore, between the pair of images to be registered, structures may appear or disappear, and intensities for the same structure may change. We have shown the result of our algorithm on a pair of real aerial images in Fig. 1. In this section, we apply our method to more real world medical images from several domains. These results demonstrate the effectiveness of our method on image registration problems that could be difficult to be solved using either pure intensity-based or pure feature-based methods.

Figure 5 shows the result of registering two real brain images. This pair of images is from the Vanderbilt Database [18]. Note that the algorithm successfully picks up several distinctive region features, and is able to recover the large rotation between the two images.

Another example on registering two MR chest images is shown in Fig. 6. This pair of images is from the Visible Human Project database. The fixed image is a T1-weighted MR image, and the moving image is a PD-weighted MR image. Despite the different tissue intensity characteristics between the two images, the salient region feature pairs chosen by the

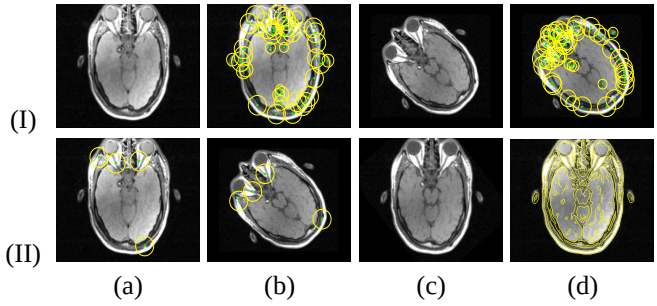

Figure 5: Registering a pair of real brain images from the Vanderbilt Database. (I.a) The fixed image. (I.b) Salient region features detected on the fixed image. (I.c). The moving image. (I.d) Salient region features on the moving image. (II.a-b) The corresponding feature pairs chosen by the algorithm upon convergence. (II.c) The transformed moving image. (II.d) The edge superimposed map after registration: edges (in yellow) from the transformed moving image superimposed on the fixed image.

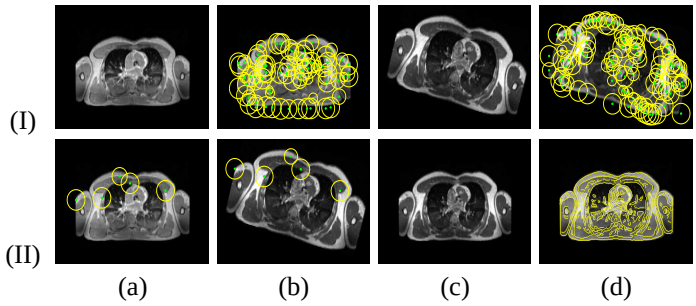

Figure 6: Registering a pair of chest MR images from the Visible Human project database. The layout of the images is the same as those in Fig. 5.

method to recover the transformation parameters correspond very well both in scale and location (see Fig. 6, II.a-b).

To demonstrate the performance of our algorithm on images with appearing and disappearing structures, we use a pair of brain images, with one of which contains a tumor. The two images are from two different subjects, and the tumor in one of the images changes its appearance significantly. The results produced by our method are shown in Fig. 7. Here the feature-based aspect of our algorithm enables it to focus on regions of similar appearance within a natural scale, thus being robust to the appearance and disappearance of local structures.

Last, but not least, we show the effectiveness of the proposed method on robust partial matching and mosaicing applications. We use a pair of curved human retinal images, as in [3]. The results are shown in Fig. 8. In this experiment, we also demonstrate the importance of the EM procedure in incrementally selecting good feature correspondences that increase the matching similarity and guaranteeing convergence. In Fig. 8, row II, we handpicked the feature pairs that seem to correspond to each other well. This results in seven feature pairs, and we transform the moving image using the transformation recovered by these feature pairs (see

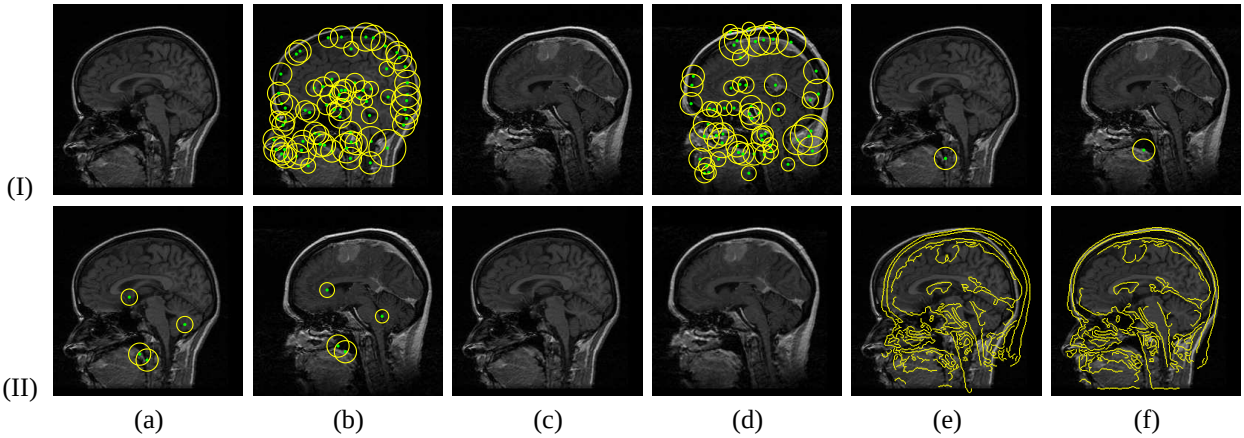

Figure 7: Registering brain images with tumor. (I.a) The fixed image. (I.b) Salient region features detected on the fixed image. (I.c). The moving image. (I.d) Salient region features on the moving image. (I.e-f) The first corresponding feature pair chosen. (II.a-b) The corresponding feature pairs chosen by the algorithm upon convergence. (II.c-d) The registration result: (II.c) the fixed image, and (II.d) the transformed moving image. (II.e-f) Comparison of the edge superimposed maps: (II.e) edges from the original moving image superimposed on the fixed image, and (II.f) edges from the transformed moving image superimposed on the fixed image.

Fig. 8, II.c-e). In the last row III of Fig. 8, we show the feature correspondences automatically chosen by the method. There are only three best feature pairs chosen, and the transformation result can be seen in Fig. 8(III.c-e). Comparing the edge superimposed map in Fig. 8(II.d) and that in Fig. 8(III.d), one can see that the three feature pairs chosen by the method in fact produce better transformation than using all the seven handpicked feature pairs. The comparison can be seen more clearly from the two zoom-in views of the edge superimposed maps: Fig. 8(II.e) *vs.* Fig. 8(III.e).

5. Discussions and Conclusions

In our current implementation of the geometric configural constraints, it is worth noting that the measure for the "goodness" of a candidate feature correspondence is based on its likelihood value and whether its addition will increase the global image "alignedness". On one hand, this permits us to efficiently recover the few best feature correspondences and to detect a convergence without explicitly setting hard thresholds. On the other hand, the strictness of the constraint also eliminates feature pairs that essentially correspond to each other individually (e.g. some handpicked good feature pairs in Fig. 8, row II, which are not chosen by the algorithm), but could deteriorate the overall global image alignment once added to the joint correspondence.

To conclude, we have presented a novel image registration method based on the region component and configural matching using scale-invariant salient region features. The proposed method possesses characteristics of both feature-based and intensity-based methods. While the overall framework is based on finding correspondences between features, all the feature correspondence likelihoods and decisions are made according to intensity-based similarity measures between region features and images. The method is efficient in that it recovers a transformation using sparse salient region feature correspondences. It is also very robust because it exploits strict global geometric constraints when finding a joint correspondence between multiple feature pairs.

In our future work, we will extend our algorithm to deal with more complicated transformation models such as affine, projective transformations as well as non-rigid deformations in both 2D and 3D. It is also interesting to investigate schemes to couple the joint correspondence detection and transformation model prediction. The goal is to identify as many good feature correspondences as possible, and fully utilize these correspondences to predict an appropriate transformation model for registration, then to estimate the transformation parameters.

Acknowledgments

We would like to thank for the stimulating discussions with C. Chefd'hotel, Dr. Guehring, Dr. Williams and other members in the Interventional Imaging Program at the Siemens Corporate Research. We would also like to thank Prof. Stewart and Prof. Fizpatrick for providing the retinal images and the brain images, respectively. This work is funded by the Siemens Corporate Research, Inc.

References

[1] B. Avants and J. Gee. Comparison and evaluation of retrospective intermodality image registration techniques. In *Proc. of second Int'l Workshop on Biomedical Image Registration, PA, USA*, 2003.

[2] S. Belongie, J. Malik, and J. Puzicha. Matching Shapes. In *Proc. of IEEE International Conf. on Computer Vision*, pages 456–461, 2001.

[3] A. Can, C.V. Stewart, B. Roysam, and H.L. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering

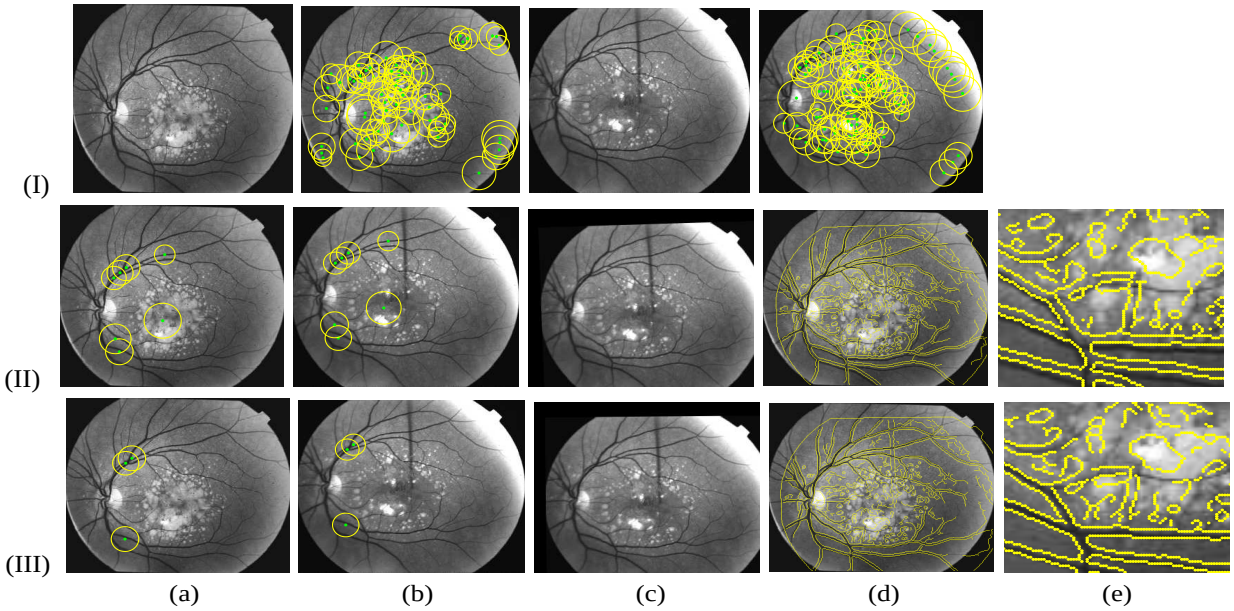

Figure 8: Registering two curved human retinal images. (I.a) The fixed image. (I.b) Salient region features on the fixed image. (I.c). The moving image. (I.d) Salient region features on the moving image. (II.a-b) The hand picked feature pairs that seem to correspond well. (II.c) The transformed moving image using the seven hand-picked feature correspondences. (II.d) Edges of the transformed moving image (in yellow) superimposed on the fixed image. (II.e) Zoom in view of II.d. (III.a-b) The corresponding feature pairs automatically chosen by the algorithm upon convergence. (III.c) The transformed moving image. (III.d) Edges of the transformed moving image superimposed on the fixed image. (III.e) Zoom in view of III.d.

pairs of images of the curved human retina. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):347–364, 2002.

[4] H. Chui and A. Rangarajan. A New Algorithm for Non-Rigid Point Matching. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages II: 44–51, 2000.

[5] A. Collignon, F. Maes, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multimodality image registration using information theory. In *Proc. of Information Processing in Medical Imaging*, pages 263–274, 1995.

[6] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages II:264–271, 2003.

[7] E. Grimson. *Object Recognition by Computer: the Role of Geometric Constraints*. MIT Press, Cambridge, MA, 1990.

[8] T. Hartkens, D.L. Hill, A.D. Castellano-Smith, D.J. Hawkes, C.R. Maurer, A.J. Martin, W.A. Hall, H. Liu, and C.L. Truwit. Using points and surfaces to improve voxel-based nonrigid registration. In *Proc. of International Conf. on Medical Imaging Copmuting and Computer-Assisted Intervention*, pages II:565–572, 2002.

[9] T. Kadir and M. Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

[10] Y. Keller and A. Averbuch. Implicit similarity: A new approach to multi-sensor image registration. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages II:543–548, 2003.

[11] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multi-modality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.

[12] C.R. Maurer, R.J. Maciunas, and J.M. Fitzpatrick. Registration of head CT images to physical space using a weighted combination of points and surfaces. *IEEE Transactions on Medical Imaging*, 17(5):753–761, 1998.

[13] D. Shen and C. Davatzikos. HAMMER: Hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):1421–1439, 2002.

[14] J. Shi and C. Tomasi. Good features to track. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 593–600, 1994.

[15] J.-P. Thirion. New feature points based on geometric invariants for 3D image registration. *International Journal of Computer Vision*, 18(2):121–137, May 1996.

[16] S.J. Timoner. *Compact Representations for Fast Nonrigid Registration of Medical Images*. Ch.6, Ph.D. dissertation, AI Tech Report 2003-015, MIT, 2003.

[17] P. Viola and W. Wells. Aligment by Maximization of Mutual Information. In *Proc. of IEEE International Conf. on Computer Vision*, pages 16–23, 1995.

[18] J. West, J. Fitzpatrick, M. Wang, B. Dawant, C. Maurer, R. Kessler, and R. Maciunas. Comparison and evaluation of retrospective intermodality image registration techniques. In *Proc. of the SPIE Conf. on Medical Imaging, vol. 2710*, pages 332–347, 1996.