Multi-modal Volume Registration Using Joint Intensity Distributions

Michael E. Leventon and W. Eric L. Grimson

Artificial Intelligence Laboratory,
Massachusetts Institute of Technology, Cambridge, MA
leventon@ai.mit.edu
http://www.ai.mit.edu/projects/vision-surgery

Abstract. The registration of multimodal medical images is an important tool in surgical applications, since different scan modalities highlight complementary anatomical structures. We present a method of computing the best rigid registration of pairs of medical images of the same patient. The method uses prior information on the expected joint intensity distribution of the images when correctly aligned, given *a priori* registered training images. We discuss two methods of modeling the joint intensity distribution of the training data, mixture of Gaussians and Parzen windowing. The fitted Gaussians roughly correspond to various anatomical structures apparent in the images and provide a coarse anatomical segmentation of a registered image pair. Given a novel set of unregistered images, the algorithm computes the best registration by maximizing the log likelihood of the two images, given the transformation and the prior joint intensity model. Results aligning SPGR and dual-echo MR scans demonstrate that this algorithm is a fast registration method with a large region of convergence and sub-voxel registration accuracy.

1 Introduction

Medical scans such as Magnetic Resonance (MR) and computed tomography (CT) are currently common diagnostic tools in surgical applications. The intensity value at a given voxel of a medical scan is primarily a function of the tissue properties at the corresponding point in space. Typically, various anatomical structures appear more clearly in different types of internal scans. Soft tissue, for example, is imaged well in MR scans, while bone is more easily discernible in CT scans. Blood vessels are often highlighted better in an MR angiogram than in a standard MR scan. Figure 1 shows three different acquisitions of MR scans. Notice that some anatomical structures appear with more contrast in one image than in the others. Anatomical structures in these various modalities can be segmented and displayed separately. However, it is most convenient for the surgeon to have information about all the structures fused into one coherent dataset. To perform the multi-modality fusion, the different volumetric images are automatically registered to a single coordinate frame.

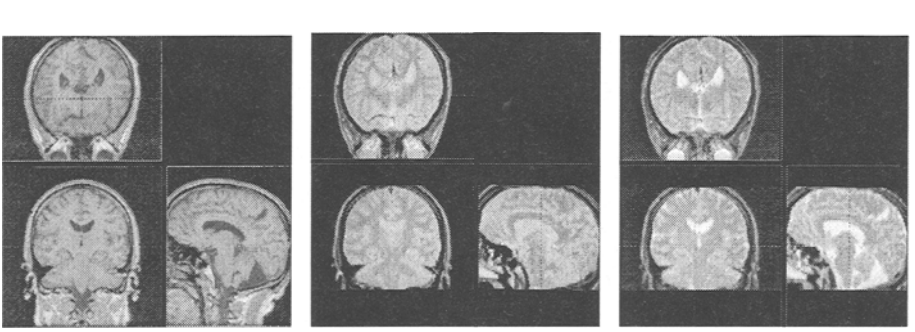

Fig. 1. SPGR, Proton Density, and T2 Weighted Magnetic Resonance images of a brain. Notice that some anatomical structures appear with more contrast in one image than the others.

1.1 Alignment of Features or Fiducials

One method of aligning two medical images is to extract features from the images, then compute the best alignment of the features [7]. This approach depends greatly on the ability to automatically and accurately extract reliable image features. In general, methods of feature extraction such as intensity thresholding or edge detection do not work well on medical scans, due to non linear intensity biases and highly textured structures. Without the ability to accurately localize corresponding features in the images, alignment in this manner is difficult.

A second registration method uses fiducial markers attached to a patient throughout the various acquisitions. If the markers can easily be located in the images, the volumes can be registered by computing the best alignment of the corresponding fiducials [8, 12]. The main drawback of this method is that the markers must remain attached to the patient throughout all image acquisitions.

1.2 Maximization of Mutual Information

Maximization of mutual information is a general approach applicable to a wide range of multi-modality registration applications [1, 2, 6, 11]. One of the strengths of using mutual information (and perhaps in some special cases, one of the weaknesses) is that MI does not use any prior information about the relationship between joint intensity distributions.

Given two random variables X and Y, mutual information is defined as [1]:

$$MI(X,Y) = H(X) + H(Y) - H(X,Y) \tag{1}$$

The first two terms on the right are the entropies of the two random variables, and encourage transformations that project X into complex parts of Y. The third term, the (negative) joint entropy of X and Y, takes on large values if X and Y are functionally related, and encourages transformations where X explains Y well. Mutual information does not use an *a priori* model of the relationships between the intensities of the different images. Our method not

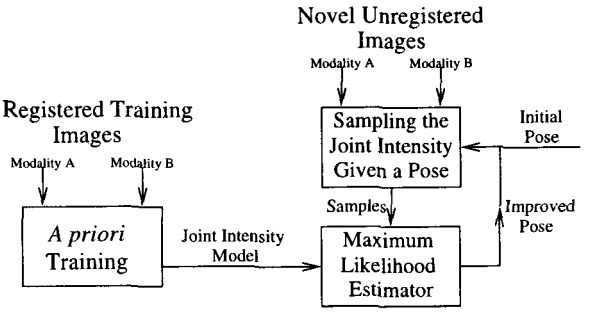

Fig. 2. Flowchart of the *a priori* training and the online registration.

only expects the relationship between the intensity values of registered images to be maximal in mutual information, but also to be similar to that of the pre-registered training data of the same modalities. The prior joint intensity model provides the registration algorithm with additional guidance which results in convergence on the correct alignment more quickly, more reliably and from further initial starting points.

1.3 Incorporating a Prior Model

The framework for our registration process is illustrated in Figure 2. The method requires a pair of registered training images of the same modalities as those we wish to register in order to build the joint intensity model. To align a novel pair of images, we compute the likelihood of the two images given a certain pose based on our model by sampling the intensities at corresponding points. We improve the current hypothesized pose by ascending the log likelihood function.

2 Learning the Joint Intensity Model

We consider two models of joint intensity: mixture of Gaussians and Parzen Window Density. In both methods, we seek to estimate the probability of observing a given intensity pair at the corresponding point in the two images.

2.1 Mixture of Gaussians Model

Given a pair of registered images from two different medical image acquisitions, we can assume that each voxel with coordinate $x = [x_1, x_2, x_3]^T$ in one image, I_1, corresponds to the same position in the patient's anatomy as the voxel with coordinate x in the other image, I_2. Further, consider that the anatomical structure S_k at some position in the patient will appear with some intensity value i_1 in the first image and i_2 in the second image with joint probability $P(i_1, i_2 \mid S_k)$. We also define $P(S_k) = \pi_k$ to be the prior probability that a random point in the medical scan corresponds to structure S_k.

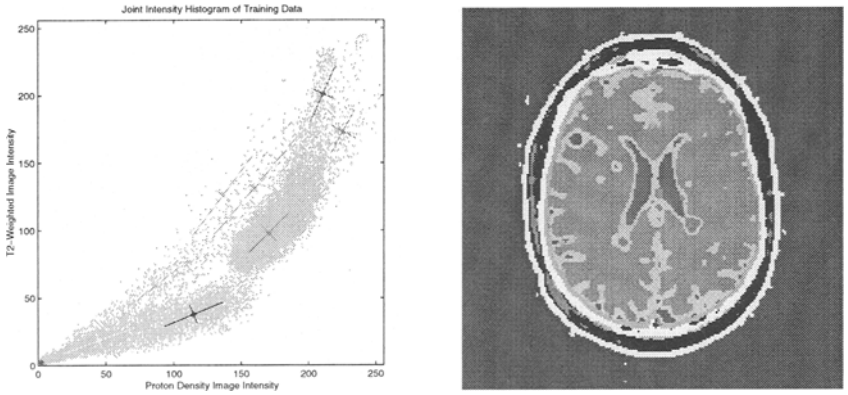

Fig. 3. LEFT: Joint intensity histogram of the registered MR PD/T2 training images used to fit a mixture of Gaussians. RIGHT: Rough segmentation of a registered image pair. Each voxel is classified based on which Gaussian it most likely belongs to, based on the mixture of Gaussian model.

By making the assumption that voxels are independent samples from this distribution (and ignoring relative positions of voxels), we have

$$P(I_1, I_2) = \prod_{x \in I} P(I_1(x), I_2(x)) \tag{2}$$

$$= \prod_{x \in I} \sum_k \pi_k P(I_1(x), I_2(x) \mid S_k) \tag{3}$$

We model the joint intensity of a particular internal structure S_k to be a two dimensional (dependent) Gaussian with mean μ_k and full covariance matrix Σ_k. Letting i be intensity pair $[i_1, i_2]^T$,

$$P(i \mid S_k) = \left(\frac{1}{2\pi |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(i-\mu_k)^T \Sigma_k^{-1} (i-\mu_k)} \right) \tag{4}$$

This model of the intensities corresponds to a mixture of Gaussians distribution, where each 2D Gaussian G_k corresponds to the joint intensity distribution of an internal anatomical structure S_k. Thus, the probability of a certain intensity pair (independent of the anatomical structure) given the model, M is

$$P(i \mid M) = \sum_k \left(\frac{\pi_k}{2\pi |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(i-\mu_k)^T \Sigma_k^{-1} (i-\mu_k)} \right). \tag{5}$$

To learn the joint intensity distribution under this model, we estimate the parameters π_k, μ_k, and Σ_k using the Expectation-Maximization (EM) method [3].

The mixture of Gaussians model was chosen to represent the joint intensity distribution because we are imaging a volume with various anatomical structures that respond with different ranges of intensity values in the two acquisitions. We

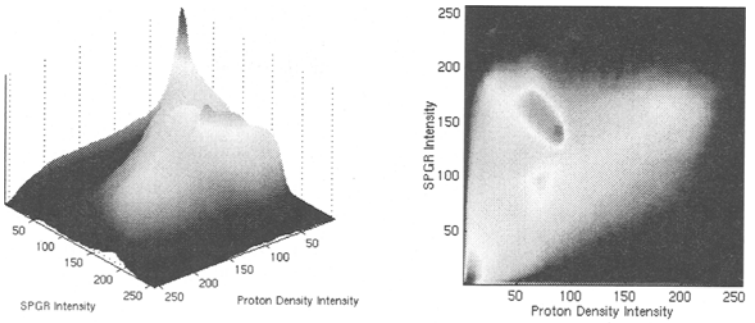

Fig. 4. Two views of the joint intensity distribution function computed using Parzen estimation with a Gaussian windowing function.

assume that those ranges of responses are approximately Gaussian in nature. Therefore, one might expect that each Gaussian in the mixture may correspond roughly to one type of anatomical structure. In other words, the model produces an approximate segmentation of the structures in the images. Figure 3b shows the segmentation of a registered pair of MR images using the Gaussian mixture model prior. Gerig, *et al.* [5] used similar methods of statistical classification to produce accurate unsupervised segmentation of 3D dual-echo MR data.

Segmentation of medical images based solely on intensity classification (without using position or shape information) is, in general, very difficult. Often different tissue types may produce a similar or overlapping range of intensity responses in a given medical scan, making classification by intensity alone quite challenging. MR images include nonlinear gain artifacts due to inhomogeneities in the receiver or transmit coils [4]. Furthermore, the signal can also be degraded by motion artifacts from movement of the patient during the scan.

The segmentation produced by this method shown in Figure 3b suffers from the difficulties described above. For example white matter and gray matter have overlapping ranges of intensities in both image acquisitions. Furthermore, note that the distinction between gray and white matter on the right hand side is not segmented clearly. This is most likely due to the bias field present in the image.

Despite these difficulties, the segmentation does a reasonable job of picking out the major structures, although it is inaccurate at region boundaries. Therefore, we do not intend to use this method alone to compute an accurate segmentation of the underlying structures. Instead, we could use the mixture model in combination with a more sophisticated algorithm to solve for segmentation, or for registration purposes, as described in section 3.

2.2 Parzen Window Density Estimation

The prior joint intensity distribution can also be modeled using Parzen window density estimation. A mixture of Gaussians model follows from the idea that the different classes should roughly correspond to different anatomical structures

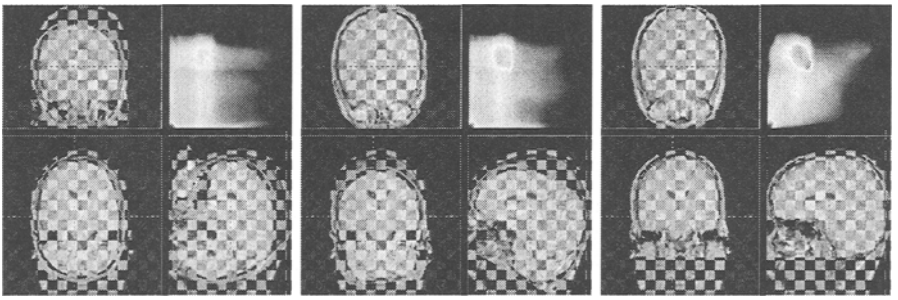

Fig. 5. Starting position, a middle position, and the final alignment computed by the registration gradient ascent algorithm. Each image shows the SPGR and PD overlayed in block format in the three orthogonal slices. The images in the upper right depict the histogram of the intensity pairs at that alignment. When the images are aligned, the histogram should resemble the distribution in Figure 4.

and thus provides an approximate segmentation into tissue classes. However, the EM algorithm for estimating the parameters of a mixture of Gaussians is sensitive to the initialization of the parameters and in some cases can result in an inaccurate prior model of the joint intensities.

We therefore also consider modeling the joint intensity distribution based on the Parzen window density estimation using Gaussians as the windowing function. In practice, this model defined by directly sampling the training data provides a better explanation of the intensity relationship than the Gaussian mixtures that require the estimation of various parameters.

Consider our registered training image pair $\langle I_1, I_2 \rangle$. We estimate the joint intensity distribution of an intensity pair $i = [i_1, i_2]^T$ given the prior model, M:

$$P(i \mid M) = \frac{1}{N} \sum_{\mu \in \langle I_1, I_2 \rangle} \left(\frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(i-\mu)^T(i-\mu)} \right) , \tag{6}$$

where the μ's are N samples of corresponding intensity pairs from the training images. Figure 4 illustrates this estimated joint intensity distribution.

3 Maximum Likelihood Registration

Given a novel pair of unregistered images of the same modalities as our training images, we assume that when registered, the joint intensity distribution of the novel images should be similar to that of the training data. When mis-registered, one structure in the first image will overlap a different structure in the second image, and the joint intensity distribution will most likely look quite different from the learned model. Given a hypothesis of registration transformation, T, and the Gaussian mixture model, M, we can compute the likelihood of the two images using Equation 2:

$$P(I_1, I_2 \mid T, M) = \prod_x P(I_1(x), I_2(T(x)) \mid T, M). \tag{7}$$

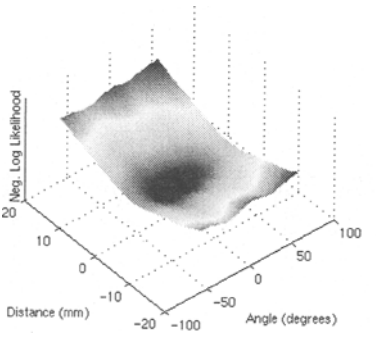

Fig. 6. Samples from the negative log likelihood function over various angles and x-shifts. Note that over this range, the function is very smooth and has one distinct minimum, which in this case occurs 0.86 mm away from the correct alignment.

We register the images by maximizing the log likelihood of the images, given the transformation and the model, and define the maximum likelihood transformation, T_{ML}, as follows:

$$\hat{T}_{ML} = \underset{T}{\operatorname{argmax}} \sum_{x} \log(P(I_1(x), I_2(T(x)) \,|\, T, M)) \tag{8}$$

The likelihood term in this equation can be substituted with either Equation 5 or 6, depending on which joint intensity model is chosen. For the results presented here, the Parzen model is used, as it better explains the intensity relationship between the two modalities. However, the mixture of Gaussians model encodes coarse tissue type classes and thus provides a framework for later incorporating into the registration process prior knowledge of the relative positions and shapes of the various internal structures.

To find the maximum likelihood transformation, T_{ML}, we use Powell maximization [9] to ascend the log likelihood function defined in Equation 8, finding the best rigid transformation. In both the mixture of Gaussian and Parzen window models of the distribution, the log likelihood objective function is quite smooth. Figure 6 illustrates samples from the negated objective function for various rotation angles (along one dimension) and x position shifts of the transformation. Over this sampled range of ± 60 degrees and ± 20 mm, the function is always concave and has one minimum which occurs within a millimeter of the correct transformation. Computing the registration by maximizing the likelihood of the image pair given the transformation and the model seems to be an efficient, accurate method of registration.

PD / T2 Registration

#	Error	#	Error	#	Error	#	Error
1	0.74	10	0.79	19	1.42	28	1.26
2	0.89	11	0.73	20	0.75	29	0.76
3	0.86	12	0.68	21	0.69	30	2.24
4	0.79	13	1.52	22	1.33	31	0.90
5	0.90	14	0.63	23	0.66	32	0.78
6	0.80	15	0.80	24	1.01	33	1.28
7	0.81	16	0.89	25	1.07	34	0.74
8	0.80	17	0.82	26	0.70	35	1.08
9	0.82	18	1.36	27	0.81	36	0.89

SPGR / PD Registration

#	Error	#	Error	#	Error	#	Error
1	4.87*	10	1.47	19	2.57	28	1.54
2	2.69	11	3.17	20	2.67	29	2.15
3	0.16	12	1.32	21	2.77	30	1.50
4	0.78	13	1.18	22	3.47	31	1.43
5	1.69	14	1.59	23	3.87	32	1.83
6	5.59*	15	1.25	24	6.03*	33	1.64
7	3.02*	16	1.26	25	9.72*	34	3.18
8	1.36	17	1.19	26	2.03	35	2.01
9	0.73	18	1.85	27	5.03*	36	2.46

Table 1. The results of registering 36 test images. The registration error (in mm) was computed by taking the average error of the eight vertices of the imaging volume. *The patient moved between scans and thus the ground truth is wrong. By inspection, our alignment looks better aligned than the "ground-truth" registration.

4 Results of Registration

The training and registration algorithms described above were tested using MR datasets from 37 patients[1]. Each patient was scanned using two protocols, a coronal SPGR scan of $256 \times 256 \times 124$ voxels with a voxel size of $0.9375 \times 0.9375 \times 1.5$ mm and a dual echo (proton density and T2-weighted) axial scan of $256 \times 256 \times 52$ voxels with a voxel size of $0.9753 \times 0.9375 \times 3.0$ mm. One of the patients' datasets was used for training purposes and the remaining 36 were used for testing. The two types of registration performed in these experiments were PD with T2 and SPGR with PD. The joint intensity distribution for each modality pair was modeled using a Parzen estimation. The initial pose in all cases was about 90° and a few centimeters away. Each registration converged in $1-2$ minutes on a Pentium Pro 200. Table 1 shows the results of the registrations.

4.1 Proton Density and T2-Weighted Volume Registration

We first consider the problem of registering PD images to transformed T2 as a means of testing the registration algorithm. Each T2-weighted scan was rotated by 90° and translated 5 cm and then registered with the corresponding PD scan. Since these T2 and PD images were acquired at the same time and are originally in the same coordinate system, by perturbing one scan by a known amount, we have accurate ground-truth upon which to validate the registration process. All 36 test cases registered with sub-voxel, and in most cases with sub-millimeter accuracy. Note that the method has a large region of convergence and thus does not require the starting pose to be very near the correct solution.

[1] The MR datasets were provided by M. Shenton and others, see acknowledgments.

4.2 SPGR and Proton Density Volume Registration

A more interesting and more challenging alignment problem consists of registering each patient's coronal SPGR scan with the same patient's PD image. This registration problem is more challenging given that the two modalities were not acquired simultaneously and also do not contain the same region of the head: the PD/T2 images are cropped at the chin (see Figure 1).

Since each patient had the scans performed during the same sitting in the scanner, the headers of the scans provide the ground-truth alignment between the various acquisitions, assuming the patient did not move between acquisitions. However, since the patients' heads were not fixed in the scanner, patients could move between acquisitions. Despite this issue, we use the scanner poses as ground truth, since in most cases it seems that the patient did not move significantly. By visually inspecting the SPGR registered with the PD images using the ground truth, one notices a discrepancy of as much as a centimeter in six of the patients' scans (marked with an $*$ in Table 1). In these cases, the error values reported are not valid, and our registration qualitatively appears better aligned than the "ground-truth" registration. Of the 36 cases we have used in our initial tests of this method, almost all of the cases automatically registered to within one voxel, from a starting position of about 90° and a few centimeters away.

5 Future Work

The promising initial registration results described herein provide various directions of future work. To date, we have tested this registration algorithm only on MR images, primarily due the availability of this particular set of data. There is interest in acquiring more datasets of different modalities, including MR Angiogram, CT, and PET to further examine this registration technique by means of the Retrospective Registration Evaluation Project [12].

Another area of further investigation is to include additional statistical models to the current framework. Non-linear bias fields present in the MR data can cause mismatches in intensity histograms between the training and test images. Registration using a prior on joint intensity information can be sensitive to these differences. Thus, there is interest in integrating the statistical intensity correction work of Wells, *et al.* [10] into this registration technique to both provide a more reliable intensity correspondence between training and test data, and perhaps assist in the process of segmenting the anatomical structures in the images.

Additional information such as prior knowledge of the shapes or relative positions of internal structures may help in the registration process. Such information certainly aids in segmentation, by offering guidelines such as what structures are most likely to appear next to other structures. Given that segmentation and registration are related in that the knowledge of one greatly assists in the computation of the other, this would imply that the addition of these types of priors may also assist in registration.

6 Acknowledgments

This article describes research supported in part by DARPA under ONR contract N00014-94-01-0994 and a National Science Foundation Fellowship. This work was developed using MRI data provided by Martha Shenton, Jane Anderson, and Robert W. McCarley, Department of Psychiatry and Surgical Planning Lab of Brigham & Women's Hospital, and Brockton VA Medical Center. The authors would like to acknowledge Dr. Shenton's NIMH grants, K02 M-01110 and R29 MH-50747, and Dr. McCarley's grant, R01-40799, and the Brockton Schizophrenia Center for the Department of Veterans Affairs.

References

1. A.J. Bell and T.J. Sejnowski. "An information-maximisation approach to blind separation." In *Advances in Neural Information Processing* **7**, Denver, 1994.
2. A. Collignon, D. Verndermeulen, P. Suetens, and G. Marchal. "3D multi-modality medical image registration using feature space clustering". In *First Conf. on Computer Vision, Virtual Reality and Robotics in Medicine*. Springer, 1995.
3. R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons.
4. T. Kapur, W.E.L. Grimson, R. Kikinis, "Segmentation of Brain Tissue from MR Images", in *First Conf. on Computer Vision, Virtual Reality and Robotics in Medicine*, Nice France, April 1995, pp. 429–433.
5. G. Gerig, J. Martin, R. Kikinis, O Kubler, M. Shenton, F. Jolesz. "Unsupervised tissue type segmentation of 3D dual-echo MR head data." Image and Vision Computing, IPMI 1991 special issue, **10**(6):349-360, 1992.
6. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. "Multimodality image registration by maximization of mutual information", in *Mathematical Methods in Biomedical Image Analysis*. IEEE Computer Society Press, 1996.
7. J.B. Antoine Maintz, P.A. van den Elsen, and M.A. Viergever. "Comparison of Feature-Based Matching of CT and MR Brain Images." *Computer Vision, Virtual Reality and Robotics in Medicine*, Nice France, April 1995, pp. 219–228.
8. V.R. Mandava, J.M. Fitzpatrick, C.R. Maurer, Jr., R.J. Maciunas, and G.S. Allen. "Registration of multimodal volume head images via attached markers." *Medical Imaging VI: Image Processing*, Proc. SPIE 1652:271-282, 1992.
9. W.H. Press, S.A. Teukolsky, S.T. Vetterling, B.P. Flannery, *Numerical Recipes in C, 2nd Edition*, Cambridge University Press, 1992.
10. W. Wells, R. Kikinis, W. Grimson, F. Jolesz. "Statistical Intensity Correction and Segmentation of Magnetic Resonance Image Data." *Third Conf. Visualization in Biomedical Computing*, Rochester MINN:SPIE 1994.
11. W.M. Wells III, P. Viola, H. Atsumi, S. Nakajima, R. Kikinis. "Multi-Modal Volume Registration by Maximization of Mutual Information". *Medical Image Analysis*, **1**(1):35–51, 1996.
12. J. West, J. Fitzpatrick, *et al.* "Comparison and evaluation of retrospective intermodality image registration techniques." In *Medical Imaging: Image Processing*, volume 2710 of *Proc. SPIE*, Newport Beach, California, February 1996.