IEEE Transactions on Medical Imaging, 22(7):865-874, 2003

Elastic Registration in the Presence of Intensity Variations

Senthil Periaswamy[1] and Hany Farid[1,2]
[1]Department of Computer Science and
[2]Center for Cognitive Neuroscience
Dartmouth College
Hanover, NH, 03755

We have developed a general purpose registration algorithm for medical images and volumes. This method models the transformation between images as locally affine but globally smooth. The model also explicitly accounts for local and global variations in image intensities. This approach is built upon a differential multiscale framework, allowing us to capture both large- and small-scale transformations. We show that this approach is highly effective across a broad range of synthetic and clinical medical images.

1. Introduction

Image registration is the process of finding a transformation that aligns one image to another. Within this broad area of research, medical image registration has emerged as a particularly active field (see e.g., [4, 8] for general surveys). This activity is due in part to the many clinical applications including diagnosis, longitudinal studies, and surgical planning, and to the need for registration across different imaging modalities (e.g., MRI, CT, PET, X-RAY, etc.). Medical image registration, however, still presents many challenges. Several notable difficulties are 1.) the transformation between images can vary widely and be highly nonlinear in nature 2.) the transformation between images acquired from different modalities may differ significantly in overall appearance and resolution; 3.) each imaging modality introduces its own unique challenges, making it difficult to develop a single generic registration algorithm.

We have developed a general registration algorithm that contends with both large overall transformations, and with more localized nonlinear transformations. This algorithm also explicitly models and estimates changes in overall appearance, allowing for effective registration within and across imaging modalities. Before describing this algorithm, we think it is helpful to first outline various decisions that are typically made when designing a registration algorithm. Within this context, we will discuss a number of existing registration techniques in order to highlight the similarities and differences of our approach to these techniques. This review is by no means exhaustive; the techniques cited are only representative examples.

In estimating the transformation between two images we must choose: 1.) to estimate the transformation between a small number of extracted landmarks or features, or between the unprocessed intensity images; 2.) a model that describes the geometric transform; 3.) whether to and how to explicitly model intensity changes; 4.) an error metric that incorporates the previous three choices; and 5.) a minimization technique for minimizing the error metric, yielding the desired transformation.

Feature-based approaches extract a (typically small) number of corresponding landmarks or features between the pair of images to be registered. The overall transformation is estimated from these features. Common features include corresponding points, edges, contours or surfaces. These features may be specified manually or extracted automatically. Feature-based approaches have the advantage of greatly reducing computational complexity. Depending on the feature extraction process, these approaches may also be more robust to intensity variations that arise during, for example, cross modal registration. These

Correspondence should be addressed to Hany Farid: 6211 Sudikoff Laboratory; Dartmouth College; Hanover NH 03755. Tel: 603.646.2761; Fax: 603.646.1672; Email: farid@cs.dartmouth.edu

approaches can be, however, highly sensitive to the accuracy of the feature extraction. Intensity-based approaches, on the other hand, estimate the transformation between the entire intensity images. Such an approach is typically more computationally demanding, but avoids the difficulties of a feature extraction stage.

Independent of the choice of a feature- or intensity-based technique, a model describing the geometric transform is required. A common and straight-forward choice is a model that embodies a single global transformation. For example, a zeroth-order polynomial limits the transformation to simple translations, a first-order polynomial allows for an affine transformation, and, of course, higher-order polynomials can be employed yielding progressively more flexible transformations. For example, the registration package Automated Image Registration (AIR) employs upto a fifth-order polynomial consisting of 168 parameters (for 3-D registration) [13, 14]. Such a global approach has the advantage that the model consists of a relatively small number of parameters to be estimated, and the global nature of the model ensures a consistent transformation across the entire image. The disadvantage of this approach is that estimation of higher-order polynomials can lead to an unstable transformation. Alternatives to these global approaches are techniques that model the global transformation as a piecewise collection of local transformations. For example, the transformation may be modeled with a local low-order polynomial, and global consistency is enforced via some form of a smoothness constraint. The advantage of such an approach is that it is capable of modeling highly nonlinear transformations without the numerical instability of high-order global models. The disadvantage is one of computational inefficiency due to the significantly larger number of model parameters that need to be estimated, and the need to guarantee global consistency.

Under certain conditions a purely geometric transformation is sufficient to model the transformation between a pair of images. Under many real-world conditions, however, the images undergo changes in both geometry and intensity (e.g., brightness and contrast). Many registration techniques attempt to remove these intensity differences with a pre-processing stage, such as histogram matching or homomorphic filtering. The issues involved with modeling intensity differences are similar to those involved in choosing a geometric model. Because the simultaneous estimation of geometric and intensity changes can be difficult, few techniques build explicit models of intensity differences. A few notable exceptions include AIR, in which global intensity differences are modeled with a single multiplicative contrast term, and SPM [1, 2] in which local intensity differences are modeled with a basis function approach.

Having decided upon a transformation model, the task of estimating the model parameters begins. As a first step, an error function in the model parameters must be chosen. This error function should embody some notion of what is meant for a pair of images to be registered. Perhaps the most common choice is a mean square error (MSE), defined as the mean of the square of the differences (in either feature distance or intensity) between the pair of images. This metric is easy to compute and often affords simple minimization techniques. A disadvantage of this error metric is that images that would qualitatively be considered to be in good registration may still have large errors due to, for example, intensity variations, or slight misalignments. Another error metric (included in AIR) is the ratio of image uniformity (RIU) defined as the normalized standard deviation of the ratio of image intensities. Such a metric is invariant to overall intensity scale differences, but typically leads to nonlinear minimization schemes. Mutual information, entropy and the Pearson product moment cross correlation are just a few examples of other possible error functions. Such error metrics are often adopted to deal with the lack of an explicit model of intensity transformations, for example [10].

In the final step of registration, the chosen error function is minimized yielding the desired model parameters. In the most straight-forward case, least-squares estimation is used when the error function is linear in the unknown model parameters. This closed-form solution is attractive as it avoids the pitfalls of iterative minimization schemes such as gradient-descent or simulated annealing. Such nonlinear minimization schemes are, however, necessary due to an often nonlinear error function. A reasonable compromise between these approaches is to begin with a linear error function, solve using least-squares, and use this solution as a starting point for a nonlinear minimization.

In developing our general purpose registration algorithm, we have tried to find a compromise between a flexible and robust technique and computational efficiency. To begin, we have chosen an intensity-based approach so as to avoid the various pitfalls involved in feature selection. Geometrically, we model the transformation with a local affine model and a global smoothness constraint. Intensity variations are explicitly modeled with local changes in brightness and contrast and a global smoothness constraint. These model parameters are simultaneously estimated at each pixel in the image allowing us to capture nonlinear distortions (in both geometry and intensity). We employ a standard MSE error metric on the intensity values. The minimization involves two steps. First an error function linear in the model parameters is minimized using least-squares. This error function is then augmented with a nonlinear smoothness constraint, and the least-squares solution is used to bootstrap an iterative nonlinear minimization. This entire procedure is built upon a differential multiscale framework, allowing us to capture both large- and small-scale transformations, see also [5, 10, 9] for related techniques.

2. Methods

We formulate the problem of image registration between a source and target image within a differential (non feature-based) framework. This formulation borrows from various areas of motion estimation (e.g., [7, 3]). We first outline the basic computational framework, and then discuss several implementation details that are critical for a successful implementation.

2.1. Local affine model

Denote $f(x, y, t)$ and $f(\hat{x}, \hat{y}, t-1)$ as the source and target images, respectively. [1] We begin by assuming that the image intensities between images are conserved (this assumption will be relaxed later), and that the motion between images can be modeled locally by an affine transform:

$$f(x, y, t) \quad = \quad f(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t-1), \tag{1}$$

where m_1, m_2, m_3, m_4 are the linear affine parameters, and m_5, m_6 are the translation parameters. These parameters are estimated locally for each small spatial neighborhood, but for notational convenience their spatial parameters are dropped. In order to estimate these parameters, we define the following quadratic error function to be minimized:

$$E(\vec{m}) \quad = \quad \sum_{x,y \in \Omega} \left[f(x, y, t) - f(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t-1) \right]^2, \tag{2}$$

where $\vec{m} = (\, m_1 \quad \ldots \quad m_6 \,)^T$, and Ω denotes a small spatial neighborhood. Since this error function is nonlinear in its unknowns, it cannot be minimized analytically. To simplify the minimization, we approximate this error function using a first-order truncated Taylor series expansion:

$$\begin{aligned} E(\vec{m}) \quad \approx \quad & \sum_{x,y \in \Omega} \big(f(x, y, t) - [f(x, y, t) + (m_1 x + m_2 y + m_5 - x) f_x(x, y, t) + \\ & (m_3 x + m_4 y + m_6 - y) f_y(x, y, t) - f_t(x, y, t)] \big)^2, \end{aligned} \tag{3}$$

where $f_x(\cdot)$, $f_y(\cdot)$, $f_t(\cdot)$ are the spatial/temporal derivatives of $f(\cdot)$. This error function further reduces to:

$$E(\vec{m}) \quad = \quad \sum_{x,y \in \Omega} \left[f_t(x, y, t) - (m_1 x + m_2 y + m_5 - x) f_x(x, y, t) - (m_3 x + m_4 y + m_6 - y) f_y(x, y, t) \right]^2. \tag{4}$$

[1] We adopt the slightly unconventional notation of denoting the source and target image with a temporal parameter t. This is done for consistency within our differential formulation.

Note that this quadratic error function is now linear in its unknowns, $\vec{m}$. This error function may be expressed more compactly in vector form as:

$$E(\vec{m}) = \sum_{x,y \in \Omega} \left[k - \vec{c}^T \vec{m} \right]^2, \tag{5}$$

where the scalar k and vector $\vec{c}$ are given as:

$$k = f_t + x f_x + y f_y \qquad \text{and} \qquad \vec{c} = (\, x f_x \quad y f_x \quad x f_y \quad y f_y \quad f_x \quad f_y \,)^T, \tag{6}$$

where again, for notational convenience, the spatial/temporal parameters of $f_x(\cdot)$, $f_y(\cdot)$, and $f_t(\cdot)$ are dropped. This error function can now be minimized analytically by differentiating with respect to the unknowns:

$$\frac{dE(\vec{m})}{d\vec{m}} = \sum_{x,y \in \Omega} -2\vec{c} \left[k - \vec{c}^T \vec{m} \right], \tag{7}$$

setting this result equal to zero, and solving for $\vec{m}$ yields:

$$\vec{m} = \left[\sum_{x,y \in \Omega} \vec{c} \, \vec{c}^T \right]^{-1} \left[\sum_{x,y \in \Omega} \vec{c} \, k \right]. \tag{8}$$

This solution assumes that the first term, a 6×6 matrix, is invertible. This can usually be guaranteed by integrating over a large enough spatial neighborhood Ω with sufficient image content. With this approach a dense locally affine mapping can be found between a source and target image.

2.2. Intensity variations

Inherent to the model outlined in the previous section is the assumption that the image intensities between the source and target are unchanged, i.e., the assumption of brightness constancy. This assumption is likely to fail under a number of circumstances. To account for intensity variations, we incorporate into our model an explicit change of local contrast and brightness. Specifically, our initial model, Equation (1), now takes the form:

$$m_7 f(x, y, t) + m_8 = f(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t - 1), \tag{9}$$

where m_7 and m_8 are two new (also spatially varying) parameters that embody a change in contrast and brightness, respectively [11]. Note that these parameters have been introduced in a linear fashion. As before, this error function is approximated with a first-order truncated Taylor series expansion to yield:

$$E(\vec{m}) = \sum_{x,y \in \Omega} \left[k - \vec{c}^T \vec{m} \right]^2, \tag{10}$$

where the scalar k and vector $\vec{c}$ are now given as:

$$k = f_t - f + x f_x + y f_y \qquad \text{and} \qquad \vec{c} = (x f_x \;\; y f_x \;\; x f_y \;\; y f_y \;\; f_x \;\; f_y \;\; -f \;\; -1)^T. \tag{11}$$

Minimization of this error function is accomplished as before by differentiating $E(\vec{m})$, setting the result equal to zero and solving for $\vec{m}$. The solution takes the same form as in Equation (8), with k and $\vec{c}$ defined as above.

Intensity variations are typically a significant source of error in differential motion estimation. The addition of the contrast and brightness terms allows us to accurately register images in the presence of local intensity variations. It is possible, of course, to fully explain the mapping between images with only a brightness modulation. In the next section we describe how to avoid such a degenerate solution.

2.3. Smoothness

Until now, we have assumed that the local affine and contrast/brightness parameters are constant within a small spatial neighborhood, Equation (10). There is a natural trade-off in choosing the size of this neighborhood. A larger area makes it more likely that the matrix $\sum_{x,y\in\Omega} \vec{c}\,\vec{c}^{\,T}$ in Equation (8) will be invertible. A smaller area, however, makes it more likely that the brightness constancy assumption will hold. We can avoid balancing these two issues by replacing the constancy assumption with a smoothness assumption [7]. That is, we assume that the model parameters $\vec{m}$ vary smoothly across space. A smoothness constraint on the contrast/brightness parameters has the added benefit of avoiding a degenerate solution where a pure brightness modulation is used to describe the mapping between images.

To begin, we augment the error function in Equation (10) as follows:

$$E(\vec{m}) \;=\; E_b(\vec{m}) + E_s(\vec{m}), \tag{12}$$

where $E_b(\vec{m})$ is defined as in Equation (10) without the summation:

$$E_b(\vec{m}) \;=\; \left[k - \vec{c}^{\,T}\vec{m}\right]^2, \tag{13}$$

with k and $\vec{c}$ as in Equation (11). The new quadratic error term $E_s(\vec{m})$ embodies the smoothness constraint:

$$E_s(\vec{m}) \;=\; \sum_{i=1}^{8} \lambda_i \left[\left(\frac{\partial m_i}{\partial x}\right)^2 + \left(\frac{\partial m_i}{\partial y}\right)^2\right], \tag{14}$$

where λ_i is a positive constant that controls the relative weight given to the smoothness constraint on parameter m_i. This error function is again minimized by differentiating with respect to the model parameters, setting the result equal to zero and solving: $dE\left(\vec{m}\right)/d\vec{m} = dE_b\left(\vec{m}\right)/d\vec{m} + dE_s\left(\vec{m}\right)/d\vec{m} = 0$. The derivative of $E_b(\vec{m})$ is:

$$\frac{dE_b\left(\vec{m}\right)}{d\vec{m}} \;=\; -2\vec{c}\left[k - \vec{c}^{\,T}\vec{m}\right]. \tag{15}$$

The derivative of $E_s(\vec{m})$ is computed by first expressing the partial derivatives, $\partial m_i/\partial x$ and $\partial m_i/\partial y$ with discrete approximations [7], and then differentiating, to yield:

$$\frac{dE_s\left(\vec{m}\right)}{d\vec{m}} \;=\; 2L(\overline{\vec{m}} - \vec{m}), \tag{16}$$

where $\overline{\vec{m}}$ is the component-wise average of $\vec{m}$ over a small spatial neighborhood, and L is an 8×8 diagonal matrix with diagonal elements λ_i, and zero off the diagonal. Setting $dE_b\left(\vec{m}\right)/d\vec{m} + dE_s\left(\vec{m}\right)/d\vec{m} = 0$, and solving for $\vec{m}$ at each pixel location yields an enormous linear system which is intractable to solve. Instead $\vec{m}$ is expressed in the following form:

$$\vec{m}^{(j+1)} \;=\; \left(\vec{c}\,\vec{c}^{\,T} + L\right)^{-1}\left(\vec{c}\,k + L\overline{\vec{m}}^{(j)}\right). \tag{17}$$

We employ an iterative scheme to solve for $\vec{m}$ [7]. On each iteration j, $\overline{\vec{m}}^{(j)}$ is estimated from the current $\vec{m}^{(j)}$. The initial estimate $\vec{m}^{(0)}$ is estimated from the closed-form solution of Section 2.2.

The use of a smoothness constraint has the benefit that it yields a dense locally affine but globally smooth transformation. The drawback is that the minimization is no longer analytic. We have found, nevertheless, that the iterative minimization is quite stable and converges relatively quickly (on the order of 25 iterations, for the results presented here).

2.4. Implementation details

While the formulation given in the previous sections is relatively straight-forward there are a number of details that are critical for a successful implementation.

First, in order to simplify the minimization, the error function of Equation (13) was derived through a Taylor-series expansion. A more accurate estimate of the actual error function can be determined using a Newton-Raphson style iterative scheme [12]. In particular, on each iteration, the estimated transformation is applied to the source image, and a new transformation is estimated between the newly warped source and target image. As few as five iterations greatly improves the final estimate.

Second, the required spatial/temporal derivatives have finite support thus fundamentally limiting the amount of motion that can be estimated. A coarse-to-fine scheme is adopted in order to contend with larger motions. A Gaussian pyramid is built for both source and target images, and the local affine and contrast/brightness parameters estimated at the coarsest level. These parameters are used to warp the source image in the next level of the pyramid. A new estimate is computed at this level, and the process repeated through each level of the pyramid. The transformations at each level of the pyramid are accumulated yielding a single final transformation. This multiscale approach is critical given the differential nature of our measurements, allowing us to register images with large motions.

Finally, the calculation of the spatial/temporal derivatives is a crucial step. Spatial/temporal derivatives of discretely sampled images are often computed as differences between neighboring sample values. Such differences are typically poor approximations to derivatives and lead to substantial errors. In computing derivatives we employ a set of derivative filters specifically designed for multi-dimensional differentiation [6]. These filters significantly improve the resulting registration.

2.5 Extensions to 3-D

The extension of this algorithm from 2-D images to 3-D volumes is relatively straight forward. Denote $f(x, y, z, t)$ and $f(\hat{x}, \hat{y}, \hat{z}, t-1)$ as the source and target volumes, respectively. Making the same local affine assumption as in Equation (1) yields:

$$\begin{aligned} f(x,y,z,t) &= f(m_1x + m_2y + m_3z + m_{10}, \\ &\quad m_4x + m_5y + m_6z + m_{11}, m_7x + m_8y + m_9z + m_{12}, t-1), \end{aligned} \tag{18}$$

where $\vec{m}$ consists of the linear affine parameters $m_1, ..., m_9$, and the translation parameters m_{10}, m_{11}, m_{12}. As before, we define an error function $E(\vec{m})$, approximate it with a first-order truncated Taylor series expansion, differentiate with respect to the unknowns, set the result equal to zero and solve to obtain:

$$\vec{m} = \left[\sum_{x,y,z \in \Omega} \vec{c}\,\vec{c}^{T} \right]^{-1} \left[\sum_{x,y,z \in \Omega} \vec{c}\,k \right], \tag{19}$$

where $\vec{m}$ is now of size 1×12, and k and $\vec{c}$ are defined as follows:

$$k = f_t + xf_x + yf_y + xf_z \tag{20}$$

$$\vec{c} = \begin{pmatrix} xf_x & yf_x & zf_x & xf_y & yf_y & zf_y & xf_z & yf_z & zf_z & f_x & f_y & f_z \end{pmatrix}^T, \tag{21}$$

where the spatial/temporal derivatives of $f(\cdot)$ are $f_x(\cdot)$, $f_y(\cdot)$, $f_z(\cdot)$, $f_t(\cdot)$. Again for notational convenience, the spatial parameters are dropped.

The contrast and brightness terms (m_{13} and m_{14}) are derived as in the 2-D case, Equations (9) and (10). The resulting solution is once again in the same form as in Equation (19), with $\vec{m}$ now being a vector with fourteen elements, and k and $\vec{c}$ defined as:

$$k = f_t - f + xf_x + yf_y + xf_z \tag{22}$$

$$\vec{c} = \begin{pmatrix} xf_x & yf_x & zf_x & xf_y & yf_y & zf_y & xf_z & yf_z & zf_z & f_x & f_y & f_z & -f & -1 \end{pmatrix}^T. \tag{23}$$

Smoothness is incorporated as before by augmenting the error function of the previous section with a smoothness term $E(\vec{m}) = E_b(\vec{m}) + E_s(\vec{m})$, where $E_b(\vec{m})$ is the error function from above given by: $E_b(\vec{m}) = \left[k - \vec{c}^T\vec{m}\right]^2$, with k and $\vec{c}$ as in Equations (22) and (23), and $E_s(\vec{m})$ is:

$$E_s(\vec{m}) = \sum_{i=1}^{14} \lambda_i \left[\left(\frac{\partial m_i}{\partial x}\right)^2 + \left(\frac{\partial m_i}{\partial y}\right)^2 + \left(\frac{\partial m_i}{\partial z}\right)^2 \right]. \tag{24}$$

An estimate of $\vec{m}$ is obtained once again by differentiating $E(\vec{m})$, setting the result equal to zero, and solving, giving the same iterative solution as in Equation (17):

$$\vec{m}^{(j+1)} = \left(\vec{c}\,\vec{c}^T + L\right)^{-1} \left(\vec{c}\,k + L\overline{\vec{m}}^{(j)}\right), \tag{25}$$

with k and $\vec{c}$ given by Equations (22) and (23). L is now a 14×14 diagonal matrix with diagonal elements λ_i, and zero off the diagonal.

3 Results

We have tested the efficacy of our registration technique on both synthetic and clinical data in both 2-D and 3-D. In all of the 2-D examples shown here (Figures 1-3), the source and target are 256×256, 8-bit grayscale images with intensity values scaled into the range $[0, 1]$. In order to contend with border effects, each image is padded with zeros to a size of 288×288. A four-level Gaussian pyramid is constructed for both the source and target image (using a 5-tap lowpass filter). At each pyramid level a single global affine transform is first estimated as in Section 2.2, with Ω, the spatial integration window, defined to be the entire image. Then, the local affine and contrast/brightness parameters, $\vec{m}$ are estimated as in Section 2.2, with $\Omega = 5 \times 5$ pixels. This estimate of $\vec{m}$ is used to bootstrap the smoothness iterations, Equation (17). In each iteration, $\lambda_i = 1 \times 10^{11}$, $i = 1, ..., 8$ and $\overline{m_i}$ is computed by convolving with the 3×3 kernel $(1\ \ 4\ \ 1\ ;\ 4\ \ 0\ \ 4\ ;\ 1\ \ 4\ \ 1)/20$. After forty iterations (inner loop), the source is warped according to the final estimate, and this process is repeated five times (outer loop). This entire process is repeated at each level of the pyramid. Although a contrast/brightness map is estimated, it is not applied when warping the source image. In order to minimize artifacts due to the warping, we accumulate successive transforms and apply a single transform to the original source image at each scale. In order to minimize edge artifacts, all convolutions are performed with a mirror-symmetric boundary. The temporal derivatives are computed using a 2-tap filter, and the spatial derivatives using a 3-tap filter. All of these parameters were chosen empirically and are generally consistent with commonly used parameters in the motion estimation literature - they were held fixed in all of the examples shown here. In general we find that the particular choice of these parameters is not crucial. Our current MatLab implementation requires approximately 16 minutes per 256×256 image on a 2 GHz Linux machine.

Shown in Figure 1 are results from four synthetically transformed images. In each case, a different random geometric and contrast/brightness transformation was applied to the source image. In each case, the registered source image is in good agreement with the target image.

Shown in Figure 2 are results from four clinical cases. In each case, the source and target images are either from different subjects or from subjects at different times. Shown across each row are the source and target images, the registered source, and the estimated transform. Shown in Figure 3 are results from four more clinical cases. In each case, the source and target images are from different modalities (from top to bottom: MRI-T1/T2, MRI-T1/MRI-Proton Density, photograph/MRI-T2, and photograph/CT - the bottom two images were taken from the visible human project). Shown across each row are the source and target images, the registered source, and the estimated transform. Even in the presence of significant intensity variations, the registered source is in good agreement with the target image.

The above examples show good qualitative registration results. We also tested quanitatively the overall accuracy of the registration over a broad range of distortions. Source images were rotated by a random amount in the range of -45 to 45 degrees, and scaled randomly between 0.8 and 1.2. In addition, a random local distortion was applied. Averaged over 50 random source images, the mean/median registration error is 1.21/0.15 pixels. The median error is significantly less than the mean due largely to edge artifacts where the estimated warp can vary significantly from the actual warp.

In all of the 3-D examples shown here (Figures 4-5), the source and target are $64 \times 64 \times 64$, 16-bit grayscale volumes with intensity values scaled into the range $[0, 1]$. These experiments closely follow the 2-D experiments, with the following parameters: a two-level Gaussian pyramid, ten inner loop iterations, four outer loop iterations, $\Omega = 5 \times 5 \times 5$ and $\lambda_i = 1 \times 10^{11}$, $i = 1, ..., 14$. The current MatLab implementation requires approximately 80 minutes on a 2 GHz Linux machine with 1 GB of memory (a $128 \times 128 \times 128$ would require approximately 10 hours).

Shown in Figure 4 are results using a synthetically transformed volume with intensity and both local and global geometric distortions. Shown in the first and second columns are source and target volumes, respectively. Shown in the third column is the registered source. Each row corresponds to a specific z-slice in the volume. Shown along the top row are results from a synthetically generated gridded volume, which illustrate the extent of the synthetic transformation and the results of registration. These results demonstrate the ability of our technique to register images with both large and small overall differences.

Shown in Figures 5 is a clinical example of registration between different subjects. Shown in the first and second columns are the source and target volumes, respectively. Shown in the third column is the registered source.

4. Discussion

We have presented a general purpose elastic registration algorithm. Our registration model incorporates both a geometric and intensity transformation. The geometric model assumes a locally affine and globally smooth transformation. The intensity model accounts for local differences in contrast and brightness while imposing a global smoothness on the overall intensity differences. The estimation of these model parameters is built upon a multiscale differential framework allowing us to capture both large- and small-scale transformations. We have tested our algorithm on a number of synthetic images/volumes generated according to the assumptions of our model. We have also shown the efficacy of our algorithm on clinical images/volumes, thus suggesting that these assumptions are reasonable.

Our current implementation suffers from a few shortcomings. First, on a 2 GHz Intel processor with 1 GB memory, the registration of a 256×256 image requires approximately 16 minutes, and the registration of a $64 \times 64 \times 64$ volume requires approximately 80 minutes. We are hopeful that optimization of our algorithm and a C-based implementation will significantly reduce this run-time. Second, we have not implicitly dealt with the problem of missing data, such as structures present in one image but absent in the other. We are currently extending our model to explicitly deal with such situations.

Acknowledgments

We would like to thank Souheil Inati and John Weaver for many helpful suggestions and providing us with data. This work is supported by NSF grants BCS-9978116 (SP), EIA-98-02068 (SP,HF), an Alfred P. Sloan Fellowship (HF) and a NSF CAREER Award IIS-99-83806 (HF).

References

[1] J. Ashburner and K.J. Friston. Multimodal image coregistration and partitioning - a unified framework. *NeuroImage*, 6(3):209–217, 1997.

[2] J. Ashburner, P. Neelin, D.L. Collins, A.C. Evans, and K.J. Friston. Incorporating prior knowledge into image registration. *NeuroImage*, 6:344–352, 1997.

[3] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, Feb. 1994.

[4] L.G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, 1992.

[5] D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans. Automatic 3d inter-subject registration of mr volumetric data in standardized talairach space. *Journal of Computer Assisted Tomography*, 18(2):192–204, 1994.

[6] H. Farid and E.P. Simoncelli. Optimally rotation-equivariant directional derivative kernels. In *International Conference on Computer Analysis of Images and Patterns*, pages 207–214, Berlin, Germany, Sept. 1997.

[7] B.K.P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.

[8] H. Lester and S.R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32(1):129–149, Jan 1999.

[9] O. Nestares and D.J. Heeger. Robust Multiresolution Alignment of MRI Brain Volumes. *Magnetic Resonance in Medicine*, 43:705–715, 2000.

[10] A. Roche, G. Malandain, N. Ayache, and S. Prima. Towards a better comprehension of similarity measures used in medical image registration. In *MICCAI*, volume 1679, pages 555–566, 1999.

[11] S. Negahdaripour and C.-H. Yu. A generalized brightness change model for computing optical flow. In *International Conference of Computer Vision*, pages 2–11, Berlin, Germany, May 1993.

[12] J. Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition*, pages 593–600, Seatle, WA, USA, June 1994.

[13] R.P. Woods, S.T. Grafton, C.J. Holmes, S.R. Cherry, and J.C. Mazziotta. Automated image registration: I. general methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography*, 22:141–154, 1998.

[14] R.P. Woods, S.T. Grafton, C.J. Holmes, S.R. Cherry, and J.C. Mazziotta. Automated image registration: II. intersubject validation of linear and nonlinear models. *Journal of Computer Assisted Tomography*, 22:155–165, 1998.

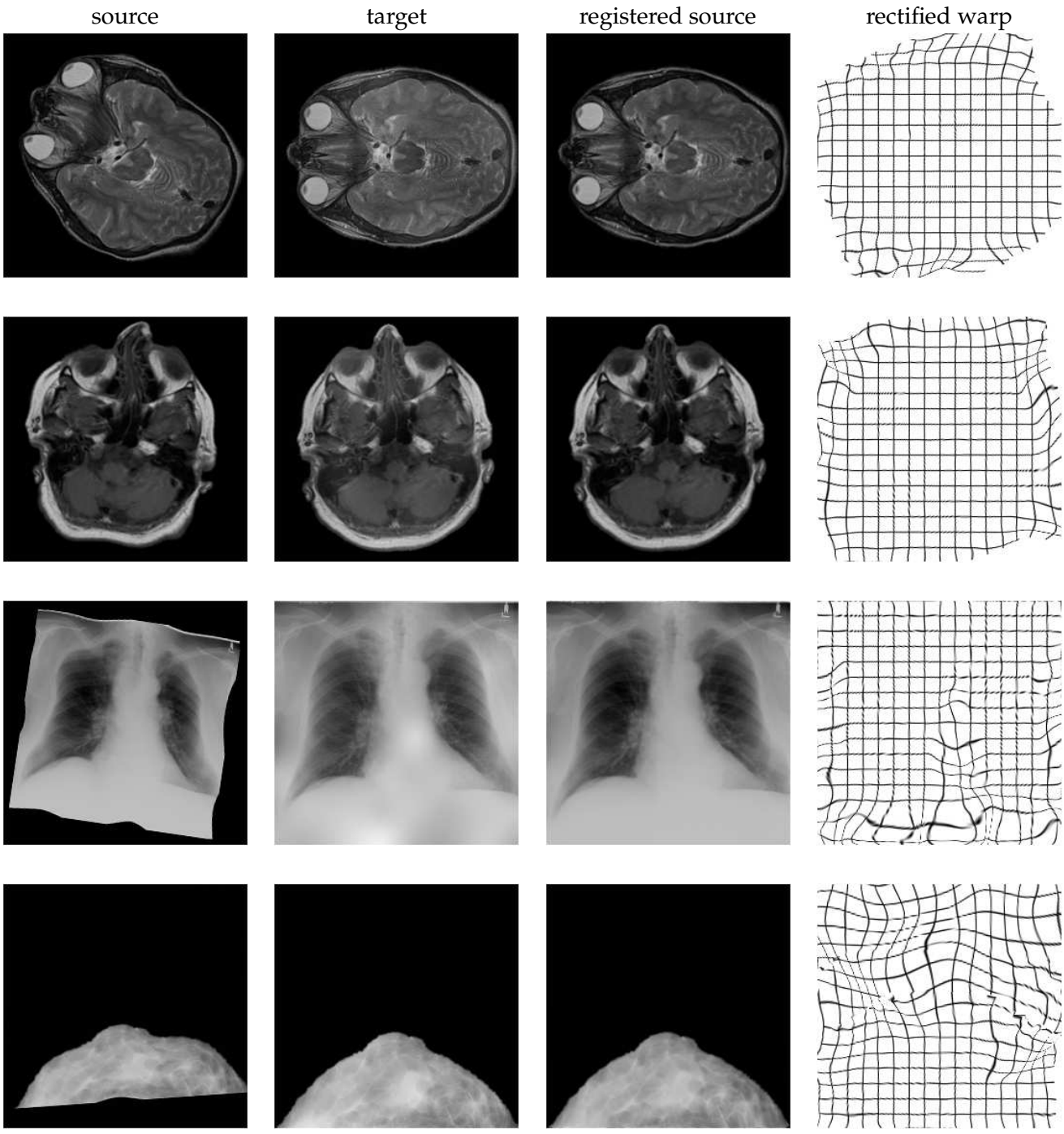

Figure 1: Results from synthetic transformations. Shown in each row is the source, target, registered source, and the rectified warp. If the estimated transform was perfect, the last column should appear as a rectilinear grid.

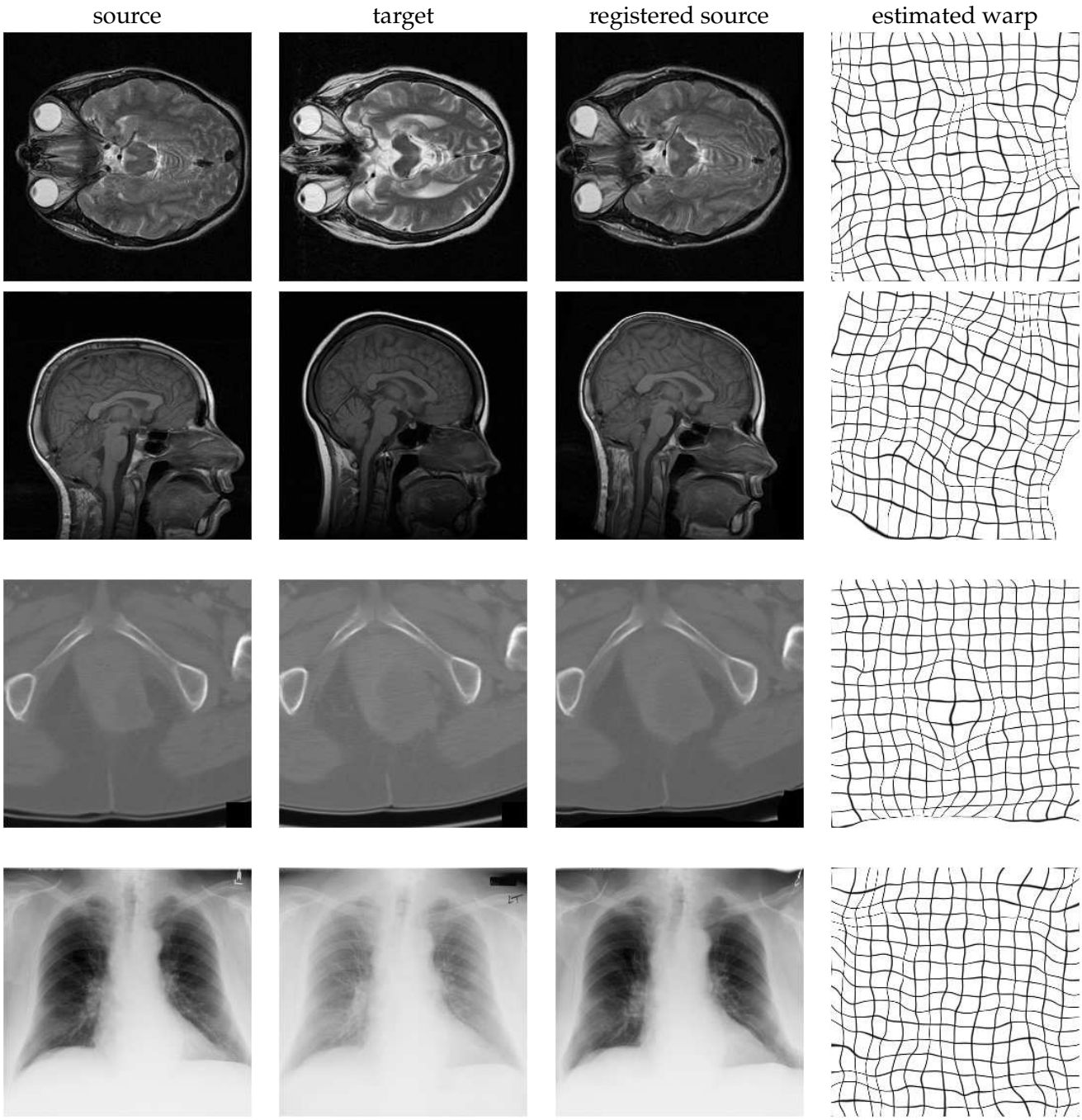

Figure 2: Results from clinical images with unknown registration. Shown in each row is the source, target, registered source, and estimated warp.

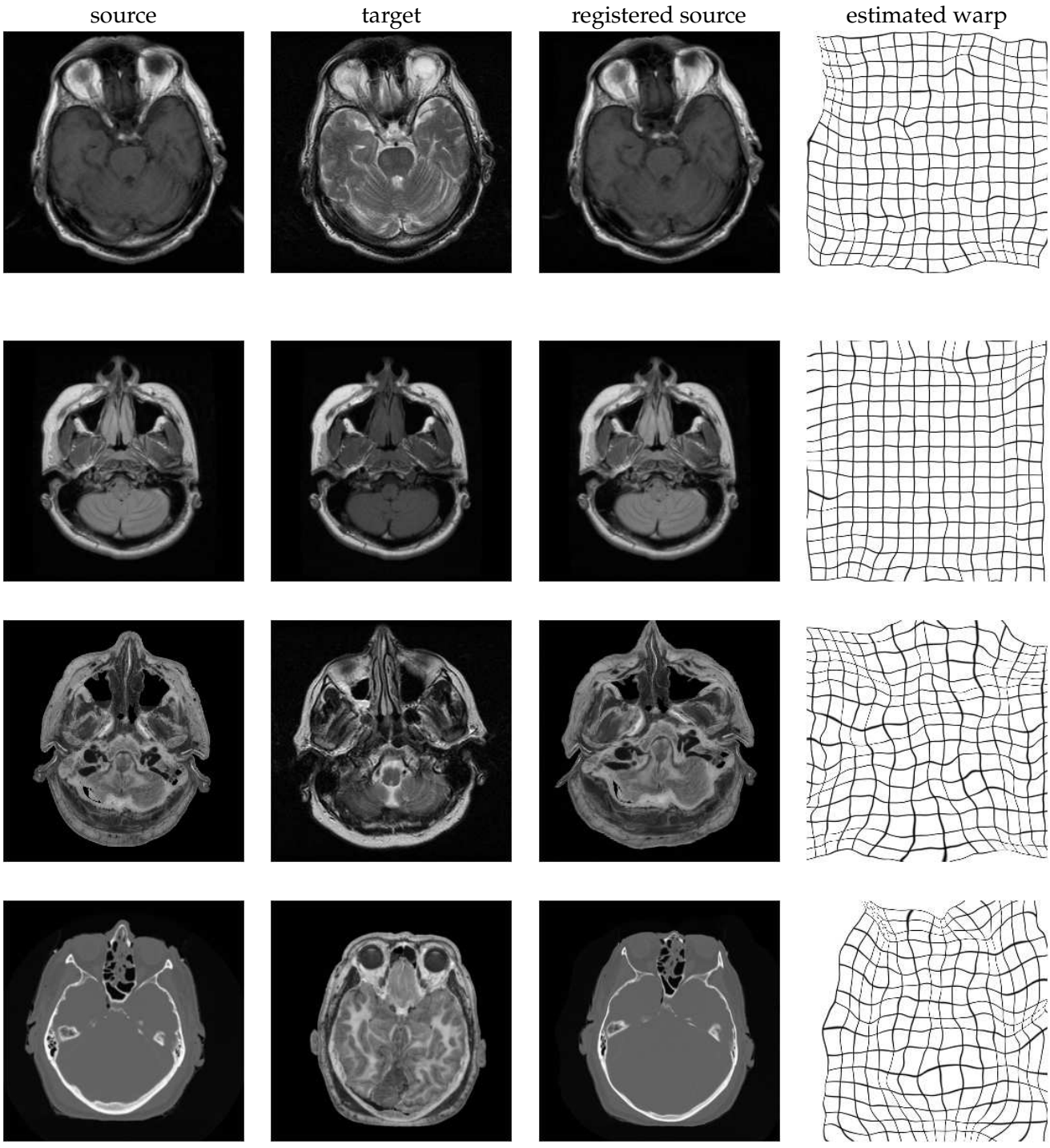

Figure 3: Results from clinical multi-modal images with unknown registration. Shown in each row is the source, target, registered source, and estimated warp.

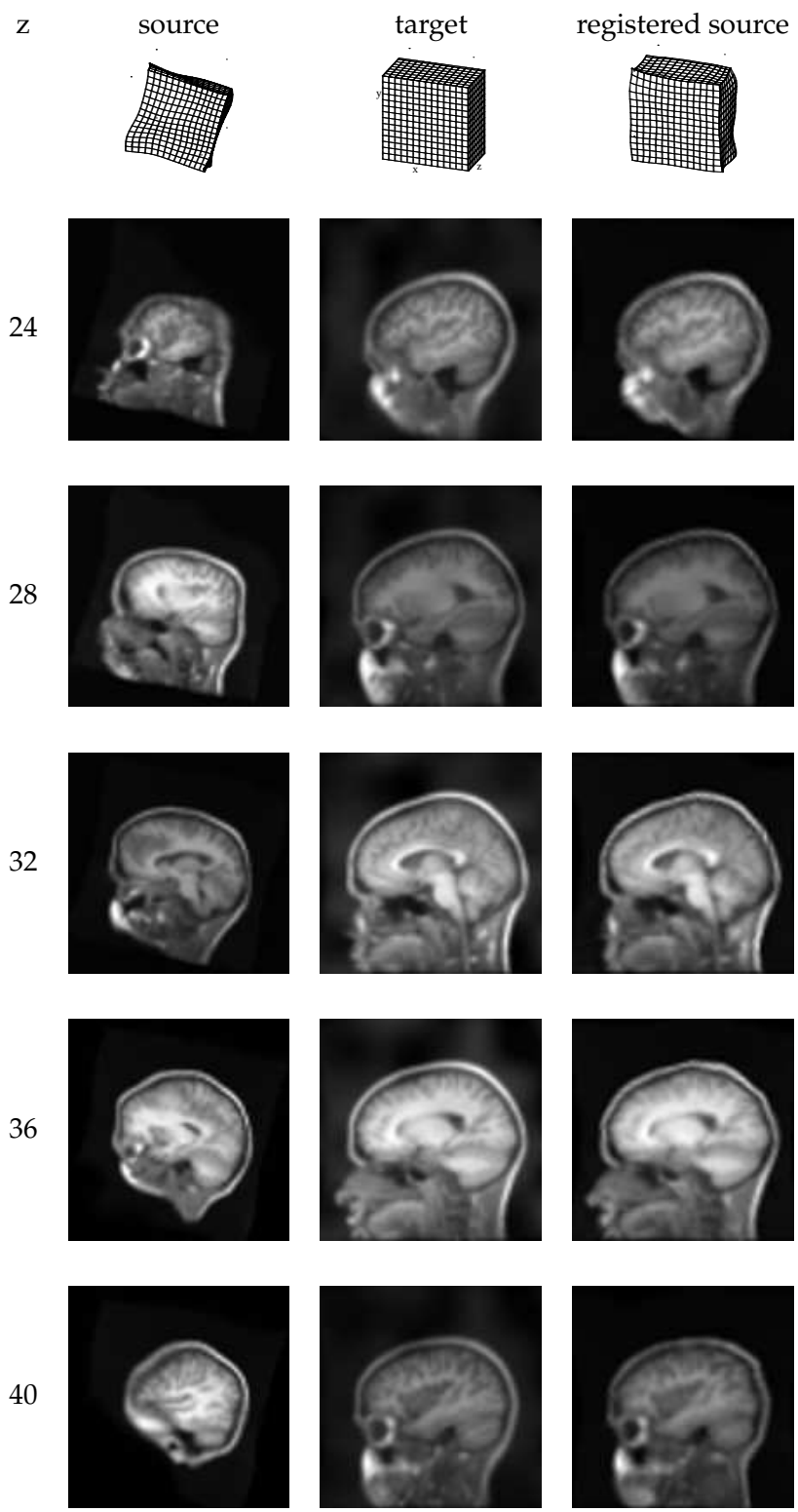

Figure 4: Results from a synthetic 3-D transform. Shown in each row is the source, target, and registered source for several z-slices.

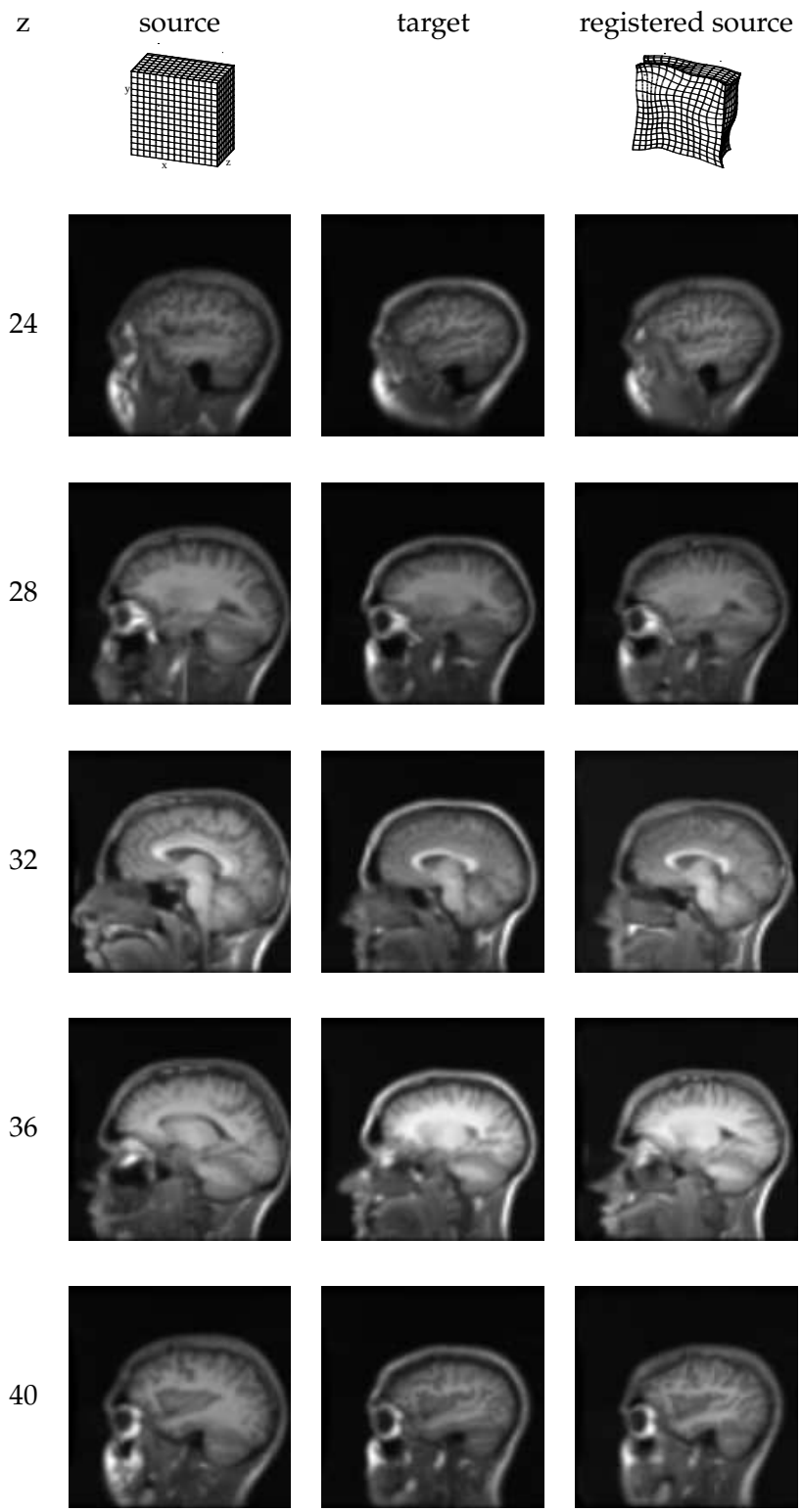

Figure 5: Results from a clinical (different subjects) 3-D case. Shown in each row is the source, target, and registered source for several z-slices.