Differential Elastic Image Registration

Senthil Periaswamy and Hany Farid
Department of Computer Science
Dartmouth College
Hanover NH 03755

We have applied techniques from differential motion estimation to the problem of automatic elastic registration of medical images. This method models the mapping between images as a locally affine but globally smooth warp. The mapping also explicitly accounts for variations in image intensities. This approach is simple and highly effective across a broad range of medical images. We show the efficacy of this approach on several synthetic and clinical images.

1. Introduction

There are a variety of methods for medical image registration (see [9, 10, 26, 19, 17] for general surveys). Differential registration techniques, however, are often cited as being ineffective, and as such have received little attention (exceptions include [4, 16, 20, 21]). This is unfortunate as differential motion techniques have been quite effective in the Computer Vision community (e.g., [15, 18, 25, 1, 14, 2, 13, 5, 12, 6, 7, 24]).

Here we present an effective technique for elastic image registration built upon a differential framework. This technique models the mapping between images as a locally affine but globally smooth warp, and explicitly accounts for variations in image intensities. The resulting registration is simple and automatic. Results from several synthetic and clinical images are shown.

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. 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, and that the motion between images can be modeled locally by an affine transform:

$$f(x,y,t) = f(m_1x+m_2y+m_5, m_3x+m_4y+m_6, t-1) \quad (1)$$

where m_1, m_2, m_3, m_4 form the 2×2 affine matrix, and m_5, m_6 the translation vector. 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:

$$E(\vec{m}) = \sum_{x,y\in\Omega} [f(x,y,t) - f(m_1x+m_2y+m_5, m_3x+m_4y+m_6, t-1)]^2, \quad (2)$$

where $\vec{m}^T = (\, m_1 \quad \ldots \quad m_6\,)$, and Ω denotes a small spatial neighborhood. Since this error function is non-linear 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:

$$E(\vec{m}) \approx \sum_{x,y\in\Omega} (f(x,y,t) - [f(x,y,t) + (m_1x+m_2y+m_5-x)f_x(x,y,t) + (m_3x+m_4y+m_6-y)f_y(x,y,t) - f_t(x,y,t)])^2, \quad (3)$$

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

$$E(\vec{m}) = \sum_{x,y\in\Omega} [f_t(x,y,t) - (m_1x+m_2y+m_5-x)f_x(x,y,t) - (m_3x+m_4y+m_6-y)f_y(x,y,t)]^2 \quad (4)$$

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:

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

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

$$k = f_t + xf_x + yf_y \quad (6)$$

$$\vec{c}^T = (\, xf_x \quad yf_x \quad xf_y \quad yf_y \quad f_x \quad f_y\,), \quad (7)$$

where again, for notational convenience, the spatial/temporal parameters of $f_x(\cdot)$, $f_y(\cdot)$, and $f_t(\cdot)$

[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.

are dropped. This error function can now be minimized analytically by differentiating with respect to the unknowns:

$$\frac{dE(\vec{m})}{d\vec{m}} = \sum_{x,y\in\Omega} -2\vec{c}\left[k - \vec{c}^T\vec{m}\right], \quad (8)$$

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

$$\vec{m} = \left[\sum_{x,y\in\Omega} \vec{c}\,\vec{c}^T\right]^{-1}\left[\sum_{x,y\in\Omega} \vec{c}k\right]. \quad (9)$$

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. 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 [22]. Specifically, our initial model, Equation (1), now takes the form:

$$m_7 f(x,y,t) + m_8 = f(m_1x + m_2y + m_5, m_3x + m_4y + m_6, t-1), \quad (10)$$

where m_7 and m_8 are two new (spatially varying) parameters that embody a change in contrast and brightness, respectively. 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:

$$E(\vec{m}) = \sum_{x,y\in\Omega}\left[k - \vec{c}^T\vec{m}\right]^2, \quad (11)$$

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

$$k = f_t - f + xf_x + yf_y \quad (12)$$

$$\vec{c}^T = (xf_x \;\; yf_x \;\; xf_y \;\; yf_y \;\; f_x \;\; f_y \;\; -f \;\; -1) \quad (13)$$

Minimizing 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 (9), with k and $\vec{c}$ as defined 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 (11). There is a natural tradeoff in choosing the size of this neighborhood. A larger area makes it more likely that the matrix $\sum_{x,y\in\Omega}\vec{c}\,\vec{c}^T$ will be invertible, Equation (9). A smaller area, however, makes it more likely that the constancy assumption will hold. We can avoid balancing these two issues by replacing the constancy assumption with a smoothness assumption [14]. 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 (11) as follows:

$$E(\vec{m}) = E_b(\vec{m}) + E_s(\vec{m}), \quad (14)$$

where $E_b(\vec{m})$ is defined as in Equation (11) without the summation:

$$E_b(\vec{m}) = \left[k - \vec{c}^T\vec{m}\right]^2, \quad (15)$$

with k and $\vec{c}$ as in Equations (12) and (13). The new quadratic error term $E_s(\vec{m})$ embodies the smoothness constraint:

$$E_s(\vec{m}) = \sum_{i=1}^{8}\lambda_i\left[\left(\frac{\partial m_i}{\partial x}\right)^2 + \left(\frac{\partial m_i}{\partial y}\right)^2\right], \quad (16)$$

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, 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:

$$\frac{dE_b\left(\vec{m}\right)}{d\vec{m}} = -2\vec{c}\left[k - \vec{c}^T\vec{m}\right]. \quad (17)$$

The derivative of $E_s(\vec{m})$ is computed by first expressing the partials, $\partial m_i/\partial x$ and $\partial m_i/\partial y$ with discrete approximations [14], and then differentiating, to yield:

$$\frac{dE_s\left(\vec{m}\right)}{d\vec{m}} = 2L(\overline{\vec{m}} - \vec{m}), \quad (18)$$

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

$$dE_b\left(\vec{m}\right)/d\vec{m} + dE_s\left(\vec{m}\right)/d\vec{m} = 0, \quad (19)$$

and solving for $\vec{m}$ at each pixel location yields an enormous linear system which is intractable to solve. As such, we express $\vec{m}$ in the following form:

$$\vec{m} = \left(\vec{c}\,\vec{c}^T + L\right)^{-1}\left(\vec{c}\,k + L\overline{\vec{m}}\right), \quad (20)$$

and employ an iterative scheme to solve for $\vec{m}$ [14]. An initial estimate of $\vec{m}$ is determined using the closed-form solution of Section 2.2. This solution yields an initial estimate of $\overline{\vec{m}}$, from which a new estimate of $\vec{m}$ is obtained, Equation (20). This process is repeated, where on each iteration a new estimate of $\overline{\vec{m}}$ is computed from the previous solution.

The use of a smoothness constraint has the benefit that it yields a dense locally affine but globally smooth mapping. 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.

2.4. Implementation details

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

First, in order to simplify the minimization, the error function of Equation (15) 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 [23]. In particular, on each iteration, the estimated warp is applied to the source image, and a new warp 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 [18, 3]. 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 warps at each level of the pyramid are accumulated yielding a single final warp.

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 [11]. These filters significantly improve the resulting registration.

3. Results

In all of the examples shown here, the source and target are 256×256, 8-bit grayscale images with intensity values scaled into the range $[0, 1]$. A three-level Gaussian pyramid is constructed for both

the source and target image. At each pyramid level a single global affine warp is first estimated according to Equation (11), with Ω, the spatial integration window, defined to be the entire image. Then, the local affine and contrast/brightness parameters, $\vec{m}$ are estimated according to Equation (11), with $\Omega = 5 \times 5$ pixels. This estimate of $\vec{m}$ is used to bootstrap the smoothness iterations, Equation (20). 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 40 iterations, the source is warped according to the final estimate, and this process is repeated 5 times. 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 warps and apply a single warp 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 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 25 minutes per image on a 1.2 GHz Linux machine.

To test our registration algorithm, we generated synthetic data by applying a locally smooth warp and contrast/brightness field to a target image. These smooth fields were randomly generated by specifying a warp parameter at equally spaced points along a coarse rectilinear grid. Intermediate values were interpolated using Bookstein's thin-plate splines [8]. On average each pixel was warped by ± 8 pixels (not including a possible global affine transform), the multiplicative contrast variation was between 0.8 and 1.0, and the additive brightness variation between 0.0 and 0.2 (with image intensities in $[0, 1]$). Shown in Figure 2(d)-(f) are examples of these warp fields.

Shown in Figure 2(a) and (b) is a synthetically generated source, and target image, respectively. Shown in panel (c) is the source image after registration. Shown in panels (d) and (e) are the applied contrast/brightness maps, and shown in panels (g) and (h) are the estimated maps. Note that while there is a tradeoff between the estimated contrast and brightness maps, errors in this estimate do not impact the estimated warp field. Shown in panel (f) is the synthetic warp as applied to a rectilinear grid. Shown in panel (i) is the result of applying the inverse of the estimated warp to panel (f). If the estimate was perfect, the result should be a rectilinear grid. Notice that in the areas of image content, this is nearly the case.

Shown in Figure 3 are results from four more synthetically warped images. In each case, a different random warp and contrast/brightness field was applied to the source image. In each case, the registered source image is in good agreement with the target image.

Shown in Figure 4 are results from four clinical cases. In each case, the source and target images are either from different subjects, from subjects at different times, or from different modalities. Shown across each row are the source and target images, the registered source, and the estimated warp. Even in the presence of significant intensity variations, the registered source is in good agreement with the target image.

And finally shown in Figure 1 are the results from an extreme and completely unrealistic synthetic warp, which we show to illustrate the robustness and flexibility of our registration technique. Unlike the previous examples, the model used here is one of translation only (i.e., no affine or contrast/brightness terms), and the smoothness parameters on these translation terms was reduced to 1×10^{-2}. These small changes were necessary to accommodate the extreme nature of the synthetic warp.

4. Discussion

We have presented an elastic registration algorithm built upon a differential framework. Our registration model incorporates both a geometric mapping that is locally affine and globally smooth, and contrast/brightness modulations that are globally smooth. The simple differential estimation

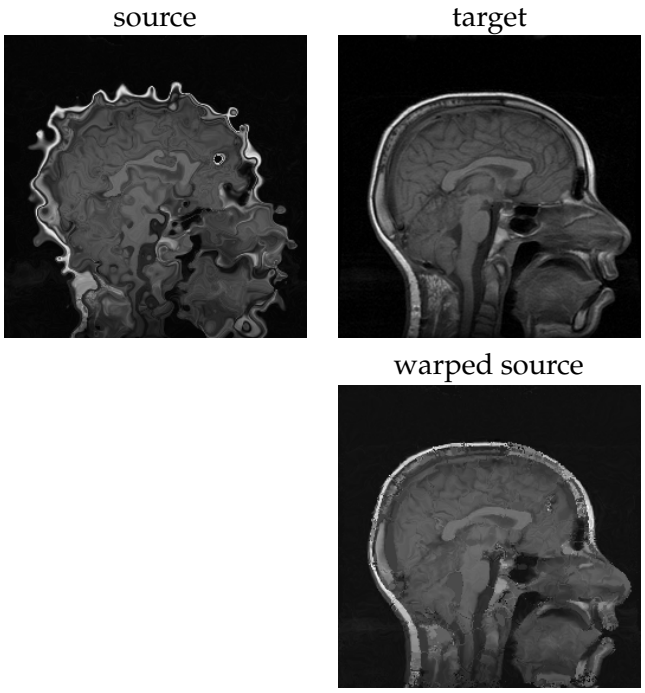


Figure 1: Shown is a target image, a source image exposed to an extreme warp, and the result of our registration.

techniques, and the flexibility of the model has proven to be highly effective across a broad range of medical images. We have tested our algorithm on a number of synthetic images generated according to the assumptions of our model. We have also shown the efficacy of our algorithm on clinical images, thus suggesting that these assumptions are reasonable.

Our current implementation suffers from a few shortcomings. First, the registration of a 256×256 image requires approximately 45 minutes on a high-end PC. We are hopeful that optimization of our algorithm and a C-based implementation will significantly reduce this run-time. Our implementation is image-based and not yet able to register volume data. The basic framework, however, extends naturally to higher-dimensions and we are currently working on a 3-D implementation.

In general, we believe that differential estimation techniques can be quite elegant and powerful. While we believe that the application to image registration is promising, it is unlikely that this approach will supplant the multitude of existing registration techniques. We expect, rather, that this approach will provide motivation for further investigation into differential methods as well as their incorporation into other registration algorithms.

Acknowledgments

This work has been supported by NSF grants BCS-9978116 (SP), EIA-98-02068 (SP,HF), and NSF CAREER Award IIS-99-83806 (HF).

References

[1] E.H. Adelson and J.R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, 2(2):284–299, 1985.

[2] J. K. Aggarwal and N. Nandhakumar. On the computation of motion from sequences of images - a review. *Proceedings of the IEEE*, 76(8):917–935, 1988.

[3] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, 1989.

[4] D. Barber, W. Tindale, E. Hunt, A. Mayes, and H. Sagar. Automatic registration of spect images as an alternative to immobilization in neuroactivation studies. *Physics in Medicine and Biology*, 40:449–463, 1995.

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

[6] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. volume 588, pages 237–252, 1992.

[7] M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and

piecewise-smooth flow fields. volume 63, pages 75–104, Jan. 1996.

[8] F. L. Bookstein. Principal Warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Patt. Anal. Machine Intell.*, 11:567–585, 1989.

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

[10] P.A. Van den Elsen, E.J.D. Pol, and M.A. Viergever. Medical image matching - a review with classification. *IEEE Engineering in Medicine and Biology*, 12(1):26–39, 1993.

[11] 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, Kiel, Germany, Sept. 1997.

[12] D.J. Fleet and A.D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5:77–104, 1990.

[13] D.J. Heeger. Optical flow using spatiotemporal filters. *International Journal of Computer Vision*, 1(4):279–302, 1988.

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

[15] B.K.P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203, 1981.

[16] K.J. Friston, J. Ashburner, J.B. Poline, C.D. Frith, J.D. Heather, and R.S.J. Frackowiak. Spatial registration and normalization of images. *Human Brain Mapping*, 2:165–189, 1995.

[17] H. Lester and S. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition, Special issue on Image Registration*, 32(1):129–149, 1999.

[18] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, 1981.

[19] J.B.A. Maintz and M.A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

[20] J. Meunier, A. Guimond, C. Janicki, B. Imbert, and J. Soucy. *Computer Assisted Radiology, Vol. 1124 of Excerpta Medica - International Congress Series*, chapter Automatic 3D registration of brain SPECT images, pages 187–192. Elsevier, Amsterdam, 1996.

[21] O. Nestares and D.J. Heeger. Robust multiresolution alignment of mri brain volumes. *Magnetic Resonance in Medicine*, 43:705–715, 2000.

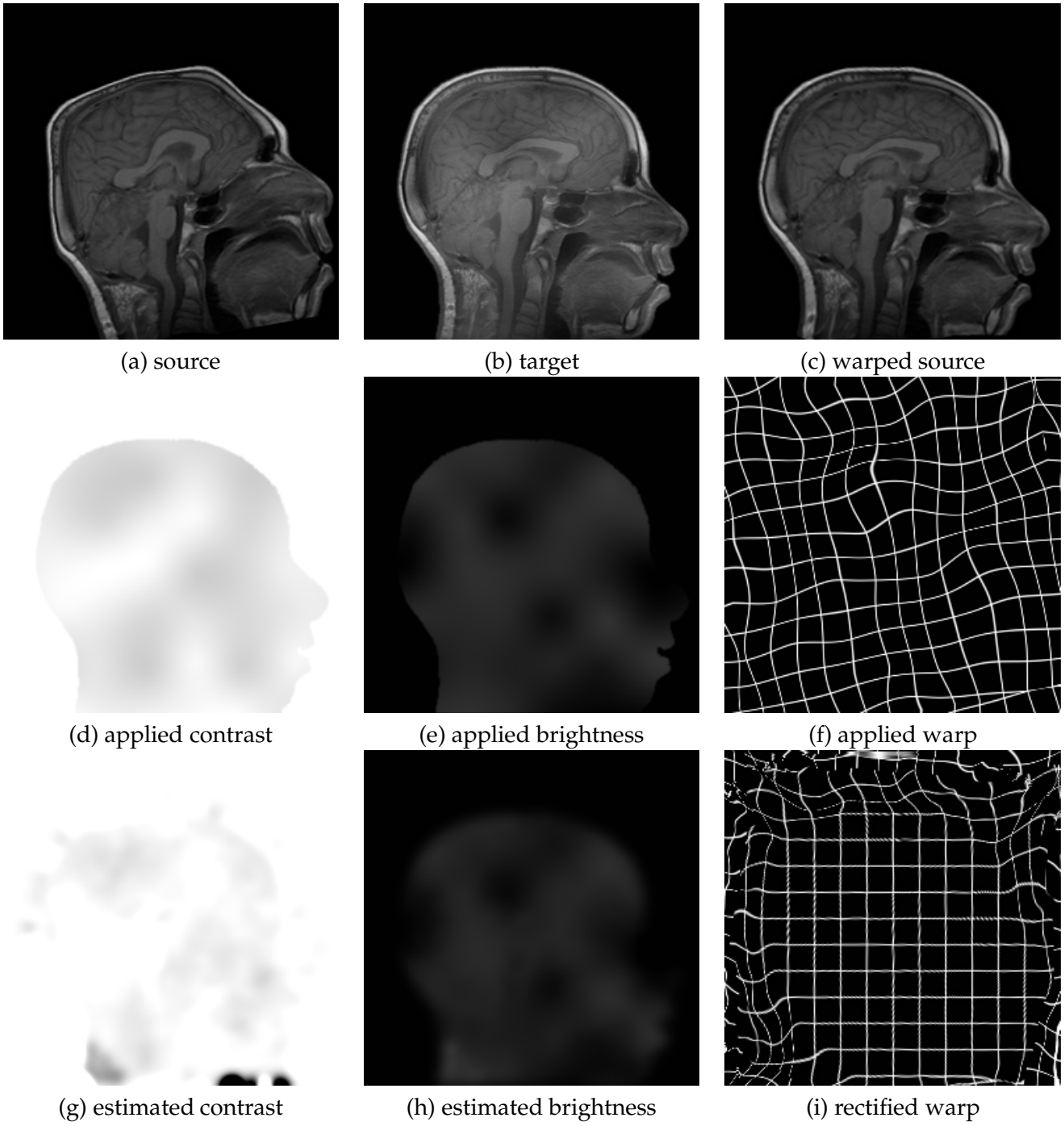
[22] S. Negahdaripour and C-H Yu. A generalized brightness change model for computing optical flow. In *Proceedings of the Fourth ICCV*, volume 2, pages 2–11, Berlin, 1993.

[23] J. Shi and C. Tomasi. Good features to track. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 1994.

[24] E.P. Simoncelli. *Handbook of Computer Vision and Applications*, chapter Bayesian Multiscale Differential Optical Flow, pages 397–420. 1999.

[25] A.B. Watson and A.J. Ahumada. *Motion: Perception and representation*, chapter A look at motion in the frequency domain, pages 1–10. 1983.

[26] 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 *Image Processing*, volume 2710, pages 332–347, Newport Beach, Apr. 1996.



(a) source (b) target (c) warped source

(d) applied contrast (e) applied brightness (f) applied warp

(g) estimated contrast (h) estimated brightness (i) rectified warp

Figure 2: Complete results from a synthetic warp. Shown along the top are the source, target and estimated warped source. Shown in panels (d), (e) and (g), (h) are the applied and estimated contrast/brightness maps. Shown in panel (f) is the applied warp. Shown in panel (i) is the inverse of the estimated warp applied to panel (f) - if the estimate was perfect, this result should be a rectilinear grid, as is nearly the case.

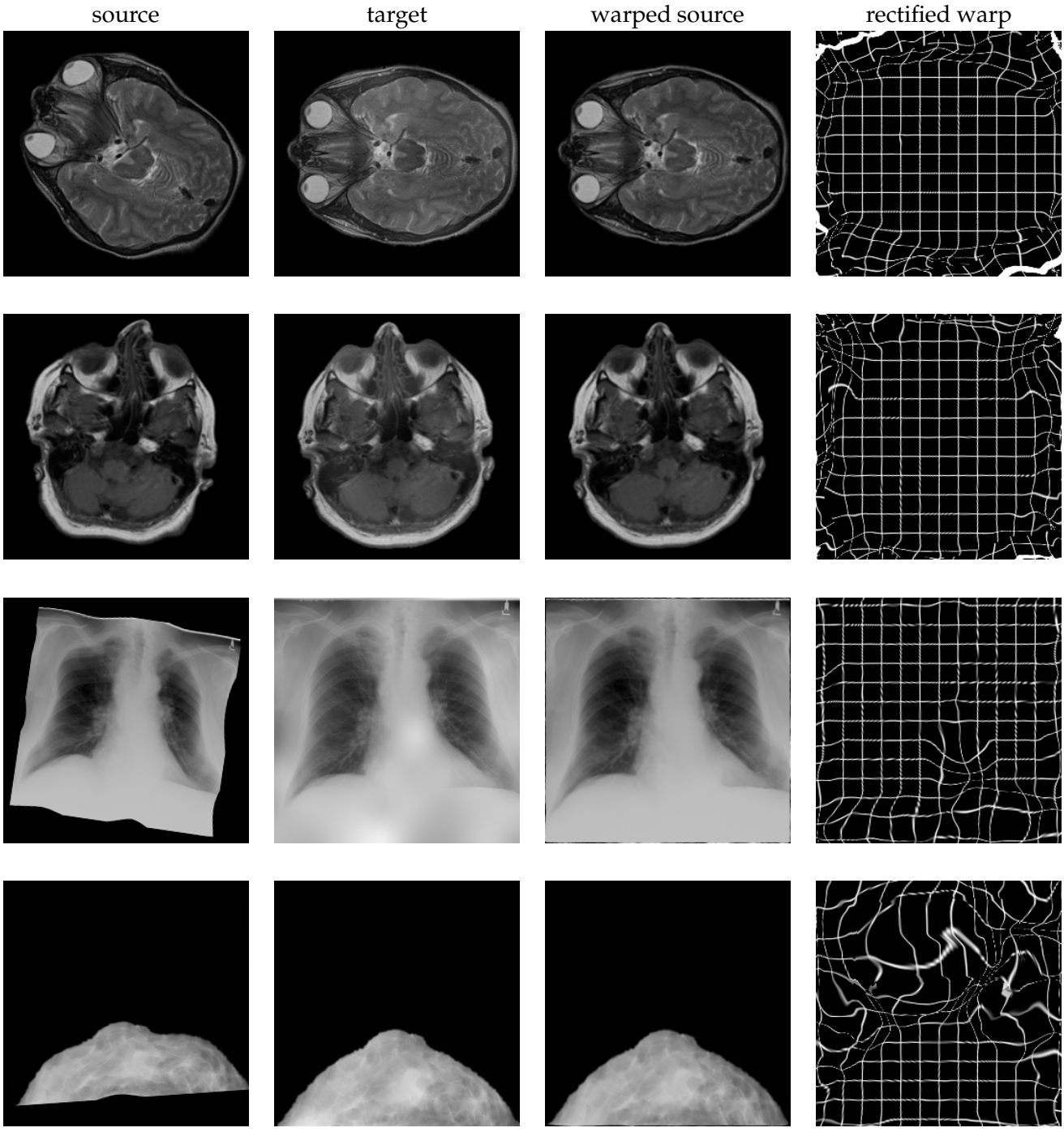


Figure 3: Results from synthetic warps. Shown in each row is the source, target and estimated warped source. If the estimated warp was perfect, the last column should appear as a rectilinear grid, see also Figure 2.

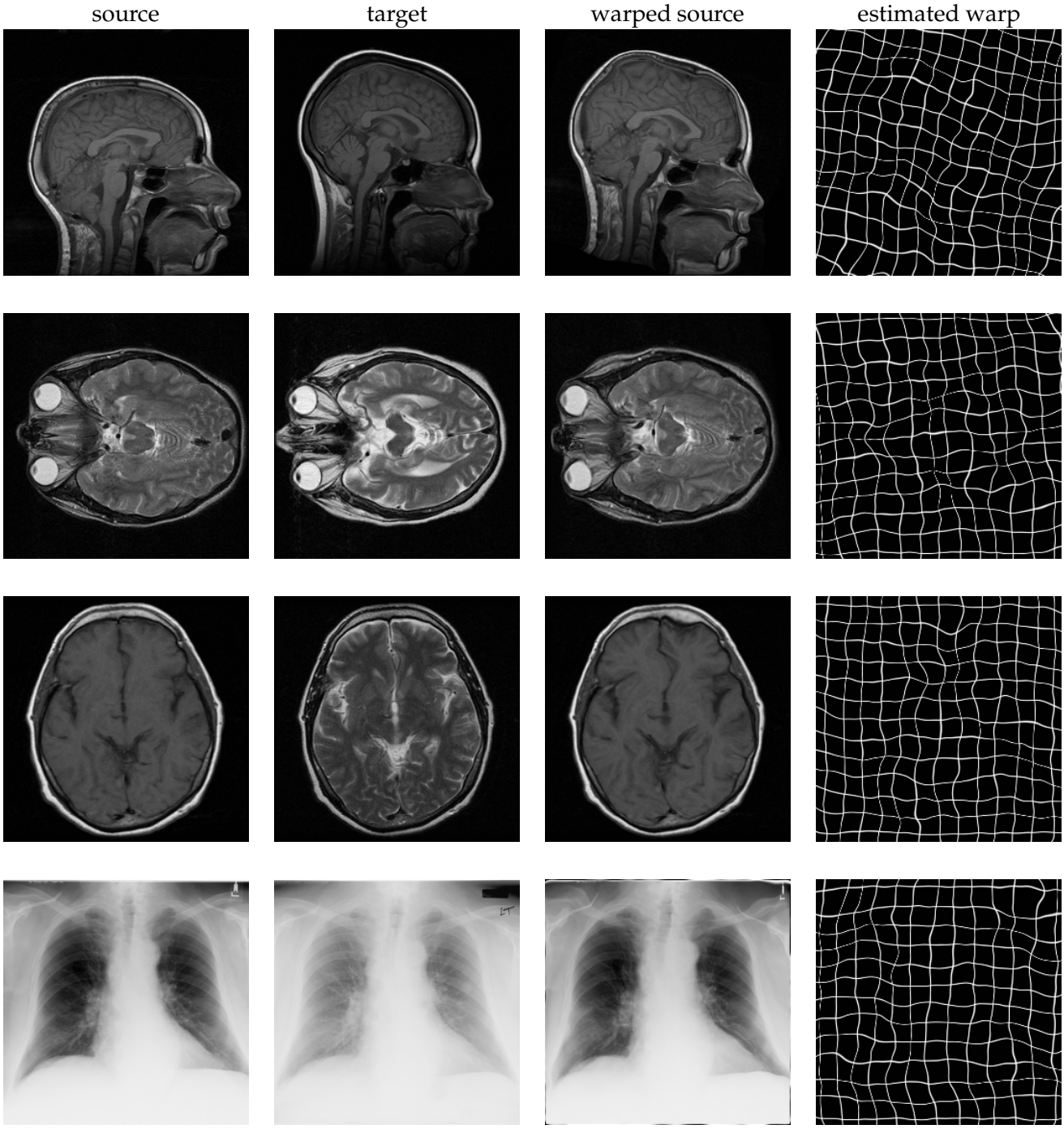


Figure 4: Results from clinical images with unknown registration. Shown in each row is the source, target and estimated warped source. Shown in the last column is the estimated warp field.