Elastic Registration with Partial Data

Senthil Periaswamy and Hany Farid

Dartmouth College, Hanover, NH, 03755, USA

Abstract. We have developed a general purpose registration algorithm for medical images and volumes. The transformation between images is modelled as locally affine but globally smooth, and explicitly accounts for local and global variations in image intensities. An explicit model of missing data is also incorporated, allowing us to simultaneously segment and register images with partial or missing data. The algorithm is built upon a differential multiscale framework and incorporates the expectation maximization algorithm. We show that this approach is highly effective in registering a range of synthetic and clinical medical images.

1 Introduction

The goal of image registration is to find a transformation that aligns one image to another. Medical image registration has emerged from this broad area of research as a particularly active field (see, for example, [8, 6] 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 (elastic) in nature; 2.) images acquired from different modalities may differ significantly in overall appearance and resolution; 3.) there may not be a one-to-one correspondence between the images (missing/partial data); and 4.) each imaging modality introduces its own unique challenges, making it difficult to develop a single generic registration algorithm.

In our earlier work, we described a general-purpose registration algorithm that contends with both large- and small-scale geometric and intensity distortions [9]. In this paper, we describe an extension to this work that allows us to explicitly contend with missing or partial data. Shown in Fig. 2 are examples of the challenges posed by missing data. In these examples there are large portions of the source image that have no corresponding match in the target image. Without an explicit segmentation or localization of these missing regions, most registration algorithms are unlikely to correctly register these images. Of course,

This work was supported by an Alfred P. Sloan Fellowship, a NSF CAREER Award (IIS-99-83806), and a department NSF infrastructure grant (EIA-98-02068). The authors can be reached at sp@cs.dartmouth.edu and farid@cs.dartmouth.edu.

if the registration between these images were known, then it would be straightforward to perform the segmentation. Similarly, if the segmentation were known, the registration could proceed. Without a known segmentation or registration, however, we are faced with a bit of a chicken and egg problem - which step should be performed first? In order to contend with this problem we have employed the expectation maximization algorithm that simultaneously segments and registers a pair of images or volumes (see also [10]).

For purposes of completeness we will briefly review our previous registration algorithm [9], and then describe the extension that allows us to contend with missing data. We then show the efficacy of this approach on several synthetic and clinical cases.

2 Registration

We formulate the problem of image registration within a differential (non feature-based) framework. This formulation borrows from various areas of motion estimation (e.g., [5, 3]). In order to contend with partial or missing data, the expectation maximization algorithm [2] is incorporated into this framework, allowing for simultaneous segmentation and registration. We first outline the basic computational framework, and then discuss several implementation details that are critical for a successful implementation.

2.1 Local affine

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 geometric transformation between images can be modeled locally by an affine transform:

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

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:

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

where $\boldsymbol{m} = (\, m_1 \quad \ldots \quad m_6 \,)^T$, and Ω denotes a small spatial neighborhood. Since this error function is non-linear in its unknowns, it cannot be minimized

[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.

analytically. To simplify the minimization, we approximate this error function using a first-order truncated Taylor series expansion:

$$E(\boldsymbol{m}) \approx \sum_{x,y\in\Omega} [f_t(x,y,t) - (m_1 x + m_2 y + m_5 - x) f_x(x,y,t) \\ -(m_3 x + m_4 y + m_6 - y) f_y(x,y,t)]^2 , \quad (3)$$

where $f_x(\cdot)$, $f_y(\cdot)$, $f_t(\cdot)$ are the spatial/temporal derivatives of $f(\cdot)$. Note that this quadratic error function is now linear in its unknowns, $\boldsymbol{m}$. This error function may be expressed more compactly in vector form as:

$$E(\boldsymbol{m}) = \sum_{x,y\in\Omega} \left[k - \boldsymbol{c}^T \boldsymbol{m}\right]^2 , \quad (4)$$

where the scalar k and vector $\boldsymbol{c}$ are given as: $k = f_t + x f_x + y f_y$ and $\boldsymbol{c} = (\, x f_x \quad y f_x \quad x f_y \quad y f_y \quad f_x \quad f_y \,)^T$. This error function can now be minimized analytically by differentiating with respect to the unknowns:

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

setting the result equal to zero, and solving for $\boldsymbol{m}$ yielding:

$$\boldsymbol{m} = \left[\sum_{x,y\in\Omega} \boldsymbol{c}\, \boldsymbol{c}^T\right]^{-1} \left[\sum_{x,y\in\Omega} \boldsymbol{c}\, k\right] . \quad (6)$$

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. When an estimate cannot be made, the local parameters are interpolated from nearby regions. 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 (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 [11]. Specifically, our initial model, Equation (1), now takes the form:

$$m_7 f(x,y,t) + m_8 = f(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t-1), \quad (7)$$

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

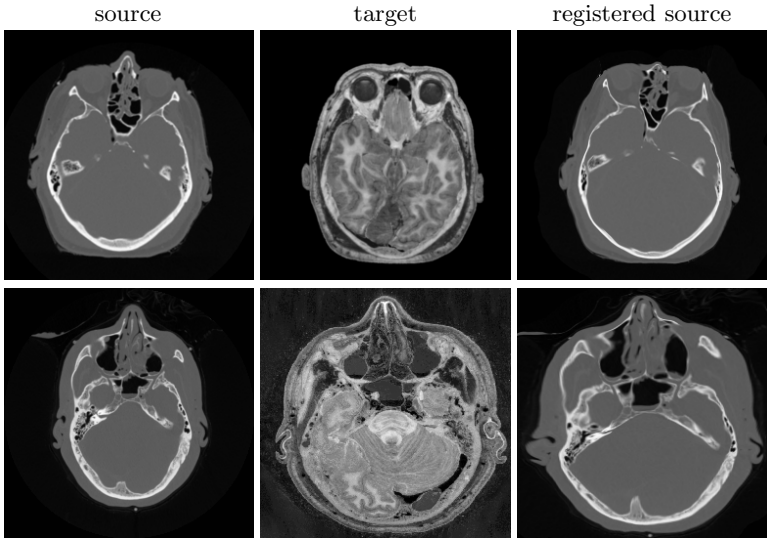

Fig. 1. Shown are examples of registration in the presence of significant geometric and intensity variations. The source are CT images, and the target are photographs from the Visible Human Project.

been introduced in a linear fashion. As before, this error function is approximated with a first-order truncated Taylor series expansion to yield:

$$E(\boldsymbol{m}) = \sum_{x,y \in \Omega} \left[k - \boldsymbol{c}^T \boldsymbol{m} \right]^2, \tag{8}$$

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

$$k = f_t - f + x f_x + y f_y \tag{9}$$

$$\boldsymbol{c} = \begin{pmatrix} x f_x & y f_x & x f_y & y f_y & f_x & f_y & -f & -1 \end{pmatrix}^T, \tag{10}$$

Minimizing this error function is accomplished as before by differentiating $E(\boldsymbol{m})$, setting the result equal to zero and solving for $\boldsymbol{m}$. The solution takes the same form as in Equation (6), with k and $\boldsymbol{c}$ as defined in Equations (9) and (10).

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 these variations, Fig. 1.

2.3 Smoothness

Until now, we have assumed that the local affine and contrast/brightness parameters are constant within a small spatial neighborhood, Equation (8). There is a natural trade-off in choosing the size of this neighborhood. A larger area

makes it more likely that the matrix in Equation (6) will be invertible. A smaller area, however, makes it more likely that the assumption of constant motion will hold. We can avoid balancing these two issues by replacing the assumption of constancy with a smoothness assumption [5]. That is, it is assumed that the model parameters $\boldsymbol{m}$ vary smoothly across space. A smoothness constraint on the contrast/brightness parameters has the added benefit of avoiding a degenerate solution where a pure intensity-based modulation is used to describe the mapping between images.

We begin with an error function, $E(\boldsymbol{m}) = E_1(\boldsymbol{m}) + E_2(\boldsymbol{m})$, that combines a smoothness constraint, $E_2(\boldsymbol{m})$, with the previous geometric and intensity transformation constraint, $E_1(\boldsymbol{m})$. The term $E_1(\boldsymbol{m})$ is defined as in Equation (8) without the summation: $E_1(\boldsymbol{m}) = \left[k - \boldsymbol{c}^T\boldsymbol{m}\right]^2$, with k and $\boldsymbol{c}$ given by Equations (9) and (10). The new term $E_2(\boldsymbol{m})$ embodies the smoothness constraint:

$$E_2(\boldsymbol{m}) = \sum_{i=1}^{8} \lambda_i \left[\left(\frac{\partial m_i}{\partial x} \right)^2 + \left(\frac{\partial m_i}{\partial y} \right)^2 \right], \tag{11}$$

where λ_i is a positive constant that controls the relative weight given to the smoothness constraint on parameter m_i. This error term penalizes solutions proportional to the local change in each parameter across a small spatial neighborhood. In so doing, we allow for a locally smooth, but globally elastic transformation. The full error function $E(\boldsymbol{m})$ is minimized, as before, by differentiating, setting the result equal to zero and solving for $\boldsymbol{m}$. The derivative of $E_1(\boldsymbol{m})$ is $dE_1(\boldsymbol{m})/d\boldsymbol{m} = -2\boldsymbol{c}\left[k - \boldsymbol{c}^T\boldsymbol{m}\right]$. The derivative of $E_2(\boldsymbol{m})$ is computed by first expressing the partials, $\partial m_i/\partial x$ and $\partial m_i/\partial y$ with discrete approximations [5], and then differentiating, to yield $dE_2(\boldsymbol{m})/d\boldsymbol{m} = 2L(\overline{\boldsymbol{m}} - \boldsymbol{m})$, where $\overline{\boldsymbol{m}}$ is the component-wise average of $\boldsymbol{m}$ over a small spatial neighborhood, and L is an 8×8 diagonal matrix with diagonal elements λ_i, and zero off the diagonal. Setting $dE_1(\boldsymbol{m})/d\boldsymbol{m} + dE_2(\boldsymbol{m})/d\boldsymbol{m} = 0$, and solving for $\boldsymbol{m}$ at each pixel location yields an enormous linear system which is intractable to solve. Instead $\boldsymbol{m}$ is estimated in the following iterative manner [5]:

$$\boldsymbol{m}^{(j+1)} = \left(\boldsymbol{c}\,\boldsymbol{c}^T + L\right)^{-1} \left(\boldsymbol{c}\,k + L\overline{\boldsymbol{m}}^{(j)}\right). \tag{12}$$

The initial estimate $\boldsymbol{m}^{(0)}$ is determined from the closed-form solution of Section 2.2. On the $j+1^{st}$ iteration $\overline{\boldsymbol{m}}^{(j)}$ is estimated from the previous estimate, $\boldsymbol{m}^{(j)}$.

The use of a smoothness constraint has the benefit that it yields a dense locally affine and 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 (see Section 2.5).

2.4 Partial Data

Inherent to the registration algorithm described above is the assumption that each region in the source image has a corresponding match in the target image.

As illustrated in Fig. 2, this need not always be the case. Under such situations, our registration algorithm typically fails. One way to contend with partial or missing data is to employ a pre-processing segmentation step. We propose, however, a more unified approach in which the registration and segmentation are performed simultaneously.

We begin by assuming that each pixel in the source and target are either related through the intensity and geometric model of Equation (7), denoted as model M_1, or cannot be explained by this transformation and therefore belongs to an "outlier" model M_2. Pixels belonging to the outlier model are those that do not have a corresponding match between the source and target images. Assuming that the pixels are spatially independent and identically distributed (iid), the likelihood of observing a pair of images is given by:

$$L(\boldsymbol{m}) = \prod_{x,y \in \Omega} P(\boldsymbol{q}(x,y)), \tag{13}$$

where, $\boldsymbol{q}(x,y)$ denotes the tuple of source, $m_7 f(x,y,t)+m_8$, and target, $f(m_1 x+ m_2 y+m_5, m_3 x+m_4 y+m_6, t-1)$, image intensities, Equation (7). To simplify the optimization of the likelihood function, we consider the log-likelihood function:

$$\begin{aligned} \log[L(\boldsymbol{m})] &= \log\left[\prod_{x,y \in \Omega} P(\boldsymbol{q}(x,y))\right] \\ &= \sum_{x,y \in \Omega} \log\left[P(\boldsymbol{q}(x,y)|M_1)P(M_1) + P(\boldsymbol{q}(x,y)|M_2)P(M_2)\right]. \end{aligned} \tag{14}$$

Assuming that the priors on the models, $P(M_1)$ and $P(M_2)$, are equal, the log-likelihood function simplifies to:

$$\log[L(\boldsymbol{m})] = \sum_{x,y \in \Omega} \log\left[P(\boldsymbol{q}(x,y)|M_1) + P(\boldsymbol{q}(x,y)|M_2)\right], \tag{15}$$

where the factored additive constant is ignored for purposes of maximization. We assume next that the conditional probabilities take the following form:

$$\log[L(\boldsymbol{m})] = \sum_{x,y \in \Omega} \log\left[e^{-r^2(x,y)/\sigma^2} + e^{-c^2}\right]. \tag{16}$$

For model M_1 we assume a Gaussian distribution (with variance σ^2), where $r(x,y)$ is the residual error between the source and target defined as:

$$r(x,y) = [(m_7 f(x,y,t) + m_8) - (f(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6, t-1))]. \tag{17}$$

For model M_2 we assume a uniform distribution (i.e., c is a constant). The log-likelihood function is maximized by differentiating, setting the result equal to zero and solving for $\boldsymbol{m}$:

$$\frac{d\log[L(\boldsymbol{m})]}{d\boldsymbol{m}} = \sum_{x,y \in \Omega} \frac{\frac{dr^2(x,y)}{d\boldsymbol{m}} e^{-r^2(x,y)/\sigma^2}}{e^{-r^2(x,y)/\sigma^2} + e^{-c^2}} = \sum_{x,y \in \Omega} \frac{dr^2(x,y)}{d\boldsymbol{m}} w(x,y) = 0, \tag{18}$$

where $w(\cdot)$ is defined to be the ratio of the exponential distributions. As in the previous sections, the residual $r(\cdot)$ is linearized with respect to the model parameters $\boldsymbol{m}$. The derivative of the residual, $dr^2(x,y)/d\boldsymbol{m}$, is then substituted into the above to yield:

$$\sum_{x,y\in\Omega} -2\boldsymbol{c}[k - \boldsymbol{c}^T\boldsymbol{m}]w = 0, \tag{19}$$

with $\boldsymbol{c}$ and k given by Equations (9) and (10), and, as before, all spatial parameters are dropped for notational convenience. Solving for the model parameters then yields the maximum likelihood estimator:

$$\boldsymbol{m} = \left[\sum_{x,y\in\Omega} (\boldsymbol{c}\boldsymbol{c}^T)w\right]^{-1}\left[\sum_{x,y\in\Omega} (\boldsymbol{c}k)w\right]. \tag{20}$$

Note that this solution is a weighted version of the earlier least-squares solution, Equation (6), where the weighting, w, is proportional to the likelihood that each pixel belongs to model M_1. As before, a smoothness constraint can be imposed to yield the following iterative estimator:

$$\boldsymbol{m}^{(j+1)} = \left((\boldsymbol{c}\ \boldsymbol{c}^T)w + L\right)^{-1}\left((\boldsymbol{c}\ k)w + L\overline{\boldsymbol{m}}^{(j)}\right). \tag{21}$$

This estimator for $\boldsymbol{m}$, however, requires an estimate of the weight w which itself requires an estimate of $\boldsymbol{m}$. The expectation/maximization algorithm (EM) [2] is used to resolve this circular estimator, and proceeds as follows:

1. E-step: compute the weights w (with an initial estimate of $\boldsymbol{m}$ from the solution of Section 2.3).
2. M-step: estimate the model parameters $\boldsymbol{m}$, Equation (21).
3. Repeat steps 1 and 2 until the difference between successive estimates of $\boldsymbol{m}$ is below a specified threshold.

The E-step is the segmentation stage, where pixels that do not have a corresponding match between source and target images have a close to zero weight w. These pixels are therefore given less consideration in the M-step which estimates the registration parameters $\boldsymbol{m}$. The EM algorithm allows for simultaneous segmentation and registration, and hence allows us to contend with missing data.

2.5 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 (8) 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 geometric 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, calculation of the spatial/temporal derivatives in Equations (9) and (10) is a crucial step. These derivatives are often computed using finite differences which typically yield poor approximations. We employ a set of derivative filters, specifically designed for multi-dimensional differentiation [4], that significantly improve the registration results. And third, a coarse-to-fine scheme is adopted in order to contend with larger motions [7, 1]. A Gaussian pyramid is first built for both source and target images, and the full registration is estimated at the coarsest level. This estimate is 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 throughout each level of the pyramid. The transformations at each level of the pyramid are accumulated yielding a single final transformation.

The generalization of the algorithm from 2-D images to 3-D volumes is relatively straight-forward. Briefly, to accommodate a 3-D affine transformation, an additional six affine parameters are added to the geometric and intensity transformation model of Equation (7). Linearization and minimization of this constraint proceeds as in the 2-D case. The smoothness constraint of Equation (11) takes on an additional $(\partial m_i/\partial z)^2$ term, and the iterative estimator of Equation (12) is of the same form, with k and $\boldsymbol{c}$ accommodating a different set of, now 3-D, spatial/temporal derivatives. The solution of Section (2.4) proceeds in a similar manner, with the initial constraint of Equation (13) updated to accommodate the 3-D geometric and intensity transformation model.

In the current MatLab implementation, running on a 2.8 GHz Linux machine, a pair of 256×256 images requires 4 minutes to register. A pair of $64 \times 64 \times 64$ volumes requires 30 minutes.

3 Results

We have tested the efficacy of our registration technique on both synthetic and clinical data in both 2-D and 3-D, Fig. 2-3. In the first two examples, regions were replaced with a uniform black or noise region. In the other examples the skull region was stripped from an axial, sagittal, and complete brain volume. In all cases, the registration is successful even with significant amounts of missing data. This registration would have failed without an explicit model of missing data incorporated directly into the registration algorithm. In all of the results of Fig. 1-3, all system parameters were held fixed.

4 Discussion

We have presented a general purpose registration algorithm. The geometric transformation is modelled as locally affine but globally smooth, and explicitly accounts for local and global variations in image intensities. An explicit model of missing data is also incorporated, allowing us to simultaneously segment and register images with partial or missing data. All of the components

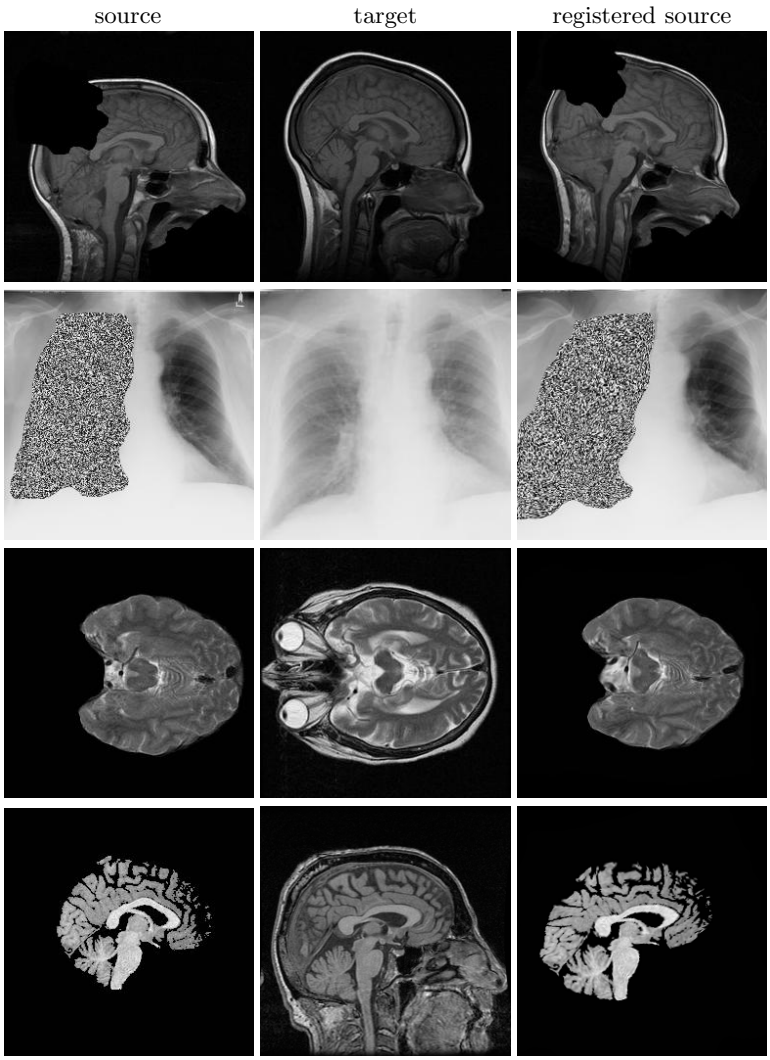

Fig. 2. Shown are synthetic and clinical examples of registration with significant portions of missing data.

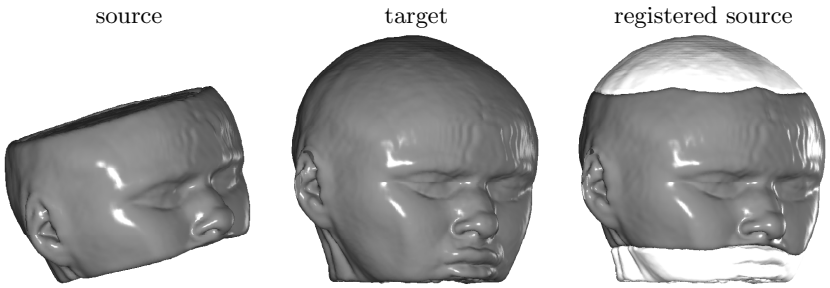

Fig. 3. Shown is an example of 3-D registration with partial data. The brighter regions shown with the registered source are the portions of the target that are missing in the source - these regions are superimposed to show the accuracy of the registration.

are combined within an integrated framework yielding a robust and effective registration algorithm within and across different imaging modalities.

References

1. P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, 1989.
2. D.B. Rubin A.P. Dempster, N.M. Laird. Maximum lilelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 99(1):1–38, 1977.
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, 1994.
4. 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, 1997.
5. B.K.P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.
6. H. Lester and S.R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32(1):129–149, 1999.
7. B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, 1981.
8. J.B.A. Maintz and M.A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.
9. S. Periaswamy and H. Farid. Elastic registration in the presence of intensity variations. *IEEE Transactions on Medical Imaging*, 2003 (in press).
10. Z. Chen A. Rangarajan J. Knisely R.Nath R. Bansal, L. Staib and J.S. Duncan. A novel approach for the registration of 2D and 3D CT images for treatment setup verification in radiotherapy. In *Medical image computing and computer-assisted intervention (MICCAI)*, pages 1075–1086, 1998.
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, 1993.
12. J. Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition*, pages 593–600, Seatle, WA, USA, 1994.