# Intensity Gradient Based Registration and Fusion of Multi-modal Images

**E. Haber, J. Modersitzki**
Department of Mathematics and Computer Science, Emory University, Atlanta, Georgia, USA

## Summary

***Objectives:*** A particular problem in image registration arises for multi-modal images taken from different imaging devices and/or modalities. Starting in 1995, mutual information has shown to be a very successful distance measure for multi-modal image registration. Therefore, mutual information is considered to be the state-of-the-art approach to multi-modal image registration. However, mutual information has also a number of well-known drawbacks. Its main disadvantage is that it is known to be highly non-convex and has typically many local maxima.

***Methods:*** This observation motivates us to seek a different image similarity measure which is better suited for optimization but as well capable to handle multi-modal images.

***Results:*** In this work, we investigate an alternative distance measure which is based on normalized gradients.

***Conclusions:*** As we show, the alternative approach is deterministic, much simpler, easier to interpret, fast and straightforward to implement, faster to compute, and also much more suitable to numerical optimization.

## Keywords

Image registration, warping, fusion, distance images, mutual information

**Methods Inf Med 2007; 46: 292–299**

**doi:10.1160/ME9046**

## 1. Objectives

Image registration is one of today's challenging medical image processing problems. The objective is to find a geometrical transformation that aligns points in one view of an object with corresponding points in another view of the same object or a similar one. Particularly in medical imaging, there are many situations that demand image registration. Typical examples include the treatment verification of pre- and post-intervention images, study of temporal series of images, and the monitoring of time evolution of an agent injection subject to a patient-motion. Another important area is the need for combining information from multiple images acquired using different modalities, sometimes called image fusion. Typical examples include the fusion of computer tomography (CT) and magnetic resonance (MRI) images or of CT and positron emission tomography (PET). Image registration is inevitable whenever images acquired from different subjects, at different times, or from different scanners, need to be combined or compared for analysis or visualization. In the past two decades computerized image registration has played an increasingly important role in medical imaging (see, e.g., [2, 7, 12, 14, 15, 27] and references therein).

One of the challenges in image registration arises for multimodal images taken from different imaging devices and/or modalities; see Figure 1 for an example. In many applications, the relation between the gray values of multi-modal images is complex and a functional dependency is generally missing. However, for the images under consideration, the gray value patterns are typically not completely arbitrary or random. This observation motivated the usage of mutual information (MI) as a distance measure between two images; cf. [4, 23]. Starting in 1995, mutual information has shown to be a successful distance measure for multi-modal image registration. Therefore, it is considered to be the state-of-the-art approach to multimodal image registration.

However, mutual information has a number of well-known drawbacks; cf. e.g. [16-18, 22, 25]. Firstly, mutual information is known to be highly non-convex and has typically many local maxima; see for example the discussion in ([15] §6.6), [25], and Section 3. Therefore, the non-convexity and hence non-linearity of the registration problem is enhanced by the usage of mutual information. Secondly, as it has its foundation in information theory, mutual information is typically considered for a discrete number of random variables.

However, fast and efficient registration schemes rely on powerful optimization techniques and thus on smooth functions. Thirdly, mutual information is defined via the joint density of the gray value distribution, and therefore, approximations of the density are required. These approximations are nontrivial to compute and typically involve some very sensitive smoothing parameters (e.g. a binning size or a Parzen window width, see [20]). Fourthly, mutual information completely decouples the gray value from the location information. Therefore, judging the output of the registration process is difficult. Finally, because of the previous difficulties, there is not a unique or common implementation for mutual information and its derivatives.

These difficulties had stimulated a vast amount of research into mutual information registration, introducing many nuisance parameters to help and bypass at least some of the difficulties; see, e.g. [18]. As a result, a practical implementation of mutual information is highly non-trivial.

These observations motivated us to seek a different image similarity measure which is capable to handle multi-modal images but better suited for optimization and interpretation. In this paper, we investigate an alternative distance measure which is based on normalized gradients. As we show, this approach is deterministic, much simpler, easier to interpret, fast and straightforward to implement, faster to compute, and also much more suitable to optimization. The idea of using derivatives to characterize similarity between images is based on the observation that image structure can be defined by intensity changes. The idea is not new. In inverse problems arising in geophysics, previous work on joint inversion [8, 10, 26] discussed the use of gradients in order to solve/fuse inverse problems of different modalities. In image registration, a more general framework similar to the one previously suggested for joint inversion was given in [6]. Further use of gradients in image registration was suggested in [18]. Our approach, the Normalized Gradient Field (NGF) is similar to [18], however, there are some main differences. Unlike the work in [18] we use only gradient information and avoid using mutual information. Furthermore, our computation of the gradient is less sensitive and allows us to deal with noisy images.

The goal of this work is to provide a new approach to multi-modal image registration using properties from differential geometry to characterize similarity between two images. The paper is organized as follows: In Section 2 we shortly lay the mathematical foundation of image registration, present an illustrative example showing some of the drawbacks of mutual information and proposed alternative image distance measure. We then discuss its numerical implementation and lay out a simple algorithm to solve the multi-modal registration problem. Finally, in Section 3 we demonstrate the effectiveness of our method. In particular, we show that our approach is much more effective than mutual information. We also demonstrate that the alternative approach leads to a simple and stable measure for image similarity.

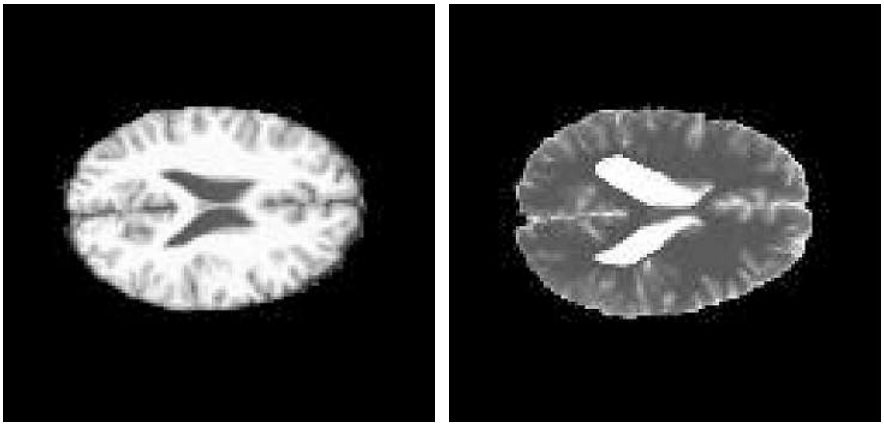


**Fig. 1** Original BrainWeb [3] and T1 (left) and T2 (right) weighted magnetic resonance images

## 2. Methods

### 2.1 The Mathematical Setting

Given a reference image, R, and a template image, T, the goal of image registration is to find a "*reasonable*" transformation such that the "*distance*" between the reference image and a deformed template image is small.

Our overall goal is a fast and efficient optimization of a distance function. We are therefore heading for a continuously differentiable objective function and thus a continuous image model; for a detailed discussion see [11]. Since the images are typically noisy but derivatives are needed we use a smoothing B-spline to approximate the image where the smoothing parameter is chosen using the Generalized Cross Validation method (GCV) [9]; for data interpolation using B-splines see [24]. For the examples in this paper the interpolated images are visually identical to the original images. To minimize notational overhead, the continuous smooth approximations are also denoted by R and T, respectively.

As described in [14], there are basically two registration approaches. One is the so-called parametric and the other one the so-called non-parametric registration technique. Since our interest is the discussion of distance measures, we focus on parametric image registration which is easier to explain.

In parametric registration, the deformation can be parameterized in terms of some basis functions, $\varphi_k$

$$\varphi(\gamma, x) = \sum_{k=1}^{m} \gamma_k \varphi_k(x); \tag{1}$$

see [14] for details. A typical example is the so-called linear registration, where for some appropriate chosen basis functions and the dimension, d = 2,

$$\varphi(\gamma, x) = \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_3 & \gamma_4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \gamma_5 \\ \gamma_6 \end{pmatrix}. \tag{2}$$

Given a distance measure D, the registration problem is to find the minimizer of

$$f(\gamma) := \mathcal{D}\Big(R(\,\cdot\,), T(\varphi(\gamma, \,\cdot\,))\Big). \tag{3}$$

Since the generalized optimization framework is based on minimization rather than optimization, we work with the negative MI throughout this paper whenever we compare it to our NGF approach.

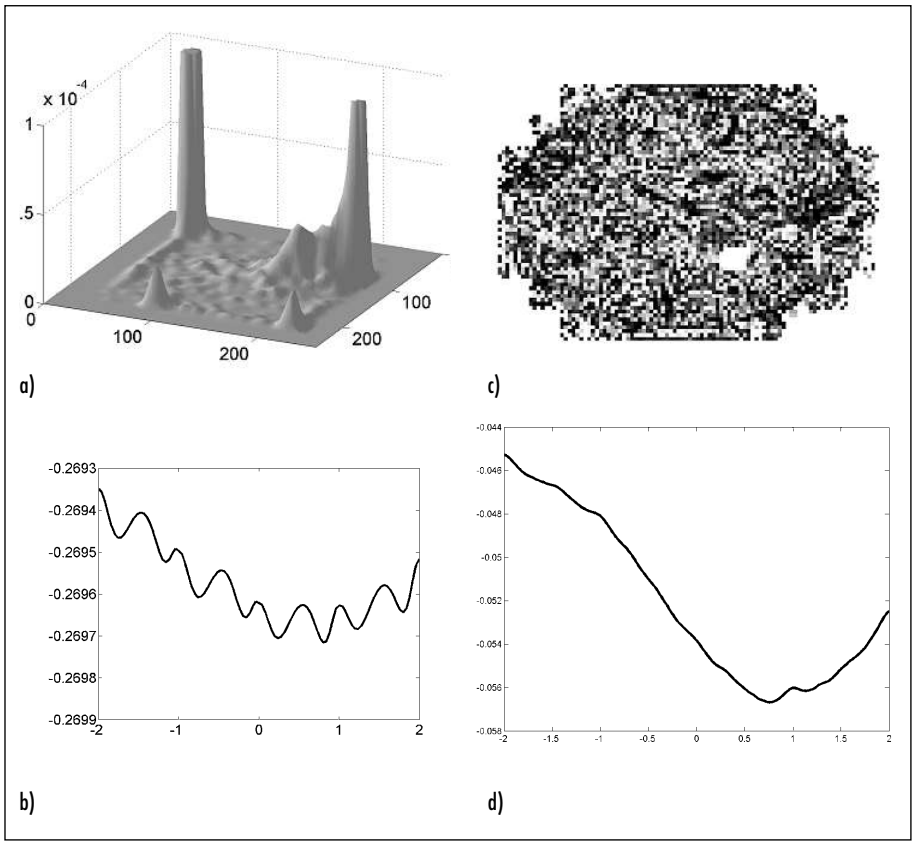


**Fig. 2** Distance measures versus shifts: a) the joint density approximation for *R* and *T*, b) negative mutual information versus shift, c) the normalized gradient field $d^d(T, R)$ (7) for *R* and *T*, d) normalized gradient field versus shift

## 2.2 An Illustrative Example

To emphasize the difficulty explained above, we present an illustrative example. Figure 1 shows a T1 and a T2 weighted magnetic resonance image (MRI) of a brain. Since the image modalities are different, a direct comparison of gray values is not advisable and we hence study a mutual information-based approach.

Figure 2a displays our approximation to the joint density which is based on a kernel estimator, where the kernel is a compactly supported smooth function; see [20] for details. Note that the joint density is completely unrelated to the spatial image content (though there is some interpretation [15]). We now slide the template along the horizontal axis. In the language of Equation 2, we fix $\gamma$ 1,...,5 and obtain the transformed image by changing 6. Figure 2b shows the negative mutual information versus the shift ranging from –2 to 2 pixels. This figure clearly demonstrates that mutual information is a highly non-convex function with respect to the shift parameter. In particular, the curve suggests that there are many pronounced local minima which are closed in value to the global minimum which is at 0.3 for MI and –0.2 for NGF (the true shift should be 0.0). Therefore, any gradient based method can run into difficulties when used to solve this problem. Even statistical techniques such as simulated annealing or genetic algorithms can run into problems when the size of a local minimum is roughly equivalent to the size of the global minimum. Furthermore, note that the dynamical range in Figure 2b is very small and thus even a smoothed version of the curve does not have a strong local minimum. Our particular example is not by any means pathologic. Similar, and even less convex curves of MI appear also in ([25] p 293) who used different interpolation.

Figure 2c displays a typical visualization of our alternative distance between R and T (discussed in the next section). Note that for the alternative distance measure, image differences are related to spatial positions. Figure 2d shows the alternative distance measure versus the shift parameter. For this particular example, it is obvious that the alternative measure is capable for multi-modal registration and it is much better suited for optimization.

## 2.3 Mutual Information

In order to be able to compare our approach and MI we aim for a continuously differentiable Mutual Information measure and therefore we use the algorithm suggested in [17]. Our computation of mutual information is based on a Parzen-Window kernel estimator for the unknown joint density function

$$\rho^h(r,t) = \sum_{\boldsymbol{x}_j} K_\sigma(r - R(\boldsymbol{x}_j)) K_\sigma(t - T(\boldsymbol{x}_j)),$$

where we used the approximation:

$$H^h[\rho^h] \approx H[\rho^h] = -\int_{\mathbb{R}^2} \rho^h(r,t) \log \rho(r,t)\ d(r,t),$$

It is interesting to note that the above implementation is computationally expensive. It requires more than 1000N floating point operations to evaluate the MI function (depending on the width of the Kernel function), where N is the number of pixels in the image. Although cheaper evaluations are possible, they are generally not differentiable.

## 2.4 A Simple and Robust Alternative

The alternative multi-modal distance measure is based on the following simple though general interpretation of similarity: *Two images are considered similar, if intensity changes occur at the same locations.*

An image intensity change can be detected via the image gradient. Since the magnitude of the gradient is dependent

upon the modality of the image, it would be unwise to base an image similarity measure on gradient magnitude. We therefore consider the normalized gradient field, i.e., the local orientation, which is purely geometric information.

$$\mathbf{n}(I,x) := \begin{cases} \frac{\nabla I(x)}{\|\nabla I(x)\|}, \\ 0, \end{cases} \quad \begin{array}{l} \nabla I(x) \neq 0, \\ \text{otherwise.} \end{array} \tag{4}$$

As usual, we set

$$\|x\| = \sqrt{\sum_{\ell=1}^{d} x_\ell^2}$$

and

$$\nabla I := (\partial_1 I, \ldots, \partial_d I)^\top. \tag{5}$$

For two related points we look at the vectors n(R, x) and n(T, x). These two vectors form an angle. Since the gradient fields are normalized, the inner product (dot-product) of the vectors is related to the cosine of this angle, while the norm of the outer product (cross-product) is related to the sine. In order to align the two images, we can either minimize the square of the sine or, equivalently, maximize the square of the cosine. The square is important in order to allow for registration of opposing orientations which is necessary for multimodal registration.

This observation motivates the distance measures shown in Figure 3 (Eqs. 6-9). Note that from an optimization point of view, the two distance measures are equivalent.

The definition of the normalized gradient field (4, 5) is not differentiable in areas where the image is constant and highly sensitive to small values of the gradient field. Suppose that the images have some distinct edges that need to be matched and some other edges, small in their magnitude which may result from noise. The normalized gradient map does not distinguish between the first and the second class. As a result, there is no preference to match wanted structures and to ignore the noisy part of the image. To avoid this problem we define the regularized normalized gradient fields shown in Figure 3 (Eqs. 9a and 9b), where the

$$\begin{aligned} d^c(T,R) &= \|\mathbf{n}(R,x) \times \mathbf{n}(T,x)\|^2, && (6)\\ \mathcal{D}^c(T,R) &= \frac{1}{2}\int_\Omega d^c(T,R)\ dx, && (7)\\ d^d(T,R) &= \langle \mathbf{n}(R,x), \mathbf{n}(T,x)\rangle^2, && (8)\\ \mathcal{D}^d(T,R) &= -\frac{1}{2}\int_\Omega d^d(T,R)\ dx, \\ \mathbf{n}_\varepsilon(I,x) &:= \frac{\nabla I(x)}{\|\nabla I(x)\|_\varepsilon}, && (9a)\\ \|\nabla I(x)\|_\varepsilon &:= \sqrt{\nabla I(x)^\top \nabla I(x) + \varepsilon^2}. && (9b) \end{aligned}$$

**Fig. 3** Equations 6-9

edge parameter, $\varepsilon$, is chosen such that the effect of noise is minimized. With this in mind and similarly to [1], we propose the following *automatic choice:*

$$\varepsilon = \frac{\eta}{V}\int_\Omega |\nabla I(x)|\ dx, \tag{10}$$

where $\eta$ is the estimated noise level in the image and $V$ is the volume of the domain. The measures (6) and (8) are based on local quantities and are easy to compute. Another advantage of these measures is that they are directly related to the resolution of the images. This property enables a straightfor-

**Algorithm 1** Calculation of Normalized gradient fields and their derivatives: $[D, d, d_T] \leftarrow \mathrm{NGF}(R, T, \varepsilon)$;

let $R, T$ be of size $m_1 \times \cdots \times m_d$, $n \leftarrow m_1 \cdots m_d$
compute (using the pointwise operations $\odot$, $./$, $(\cdot)^2$ and $\sqrt{}$)

$$\begin{aligned} \|\nabla^h R\|_\varepsilon &\leftarrow \sqrt{\sum (\partial_i^h R)^2 + \varepsilon^2}, \\ \|\nabla^h T\|_\varepsilon &\leftarrow \sqrt{\sum (\partial_i^h T)^2 + \varepsilon^2}, \\ d &\leftarrow \mathrm{diag}(1./\|\nabla^h R\|_\varepsilon) \cdot \\ &\quad \mathrm{diag}(1./\|\nabla^h T\|_\varepsilon)\left(\sum \mathrm{diag}(\partial_i^h R)\,(\partial_i^h T)\right), \\ D &\leftarrow -\tfrac{1}{2n} d^\top d, \end{aligned}$$

if derivatives are needed compute

$$\begin{aligned} [\|\nabla^h T\|_\varepsilon]_T &\leftarrow \mathrm{diag}(1./\|\nabla^h T\|_\varepsilon)\left(\sum \mathrm{diag}(\partial_i^h T)\partial_i^h\right) \\ d_T &\leftarrow \mathrm{diag}(1./\|\nabla^h R\|_\varepsilon)\cdot \mathrm{diag}(1./\|\nabla^h T\|_\varepsilon) \\ &\quad \cdot\left(\sum \mathrm{diag}(\partial_i^h R)\partial_i^h\right) \\ &\quad - \mathrm{diag}(d)\cdot \mathrm{diag}(1./\|\nabla^h T\|_\varepsilon)\cdot [\|\nabla^h T\|_\varepsilon]_T. \end{aligned}$$

**Fig. 4** Computation of objective function and gradient

**Table 1** Experiments with image 1: chosen vs recovered $\gamma$

| | $\gamma_1$ | $\gamma_2$ | $\gamma_3$ | $\gamma_4$ | $\gamma_5$ | $\gamma_6$ |
|---|---|---|---|---|---|---|
| True | 2 | 0 | 0 | 1 | 0 | 0 |
| Recovered | 1.95 | 0.01 | 0.05 | 1.01 | 0.03 | 0.01 |
| True | 1 | 2 | 0 | 1 | 0 | 0 |
| Recovered | 1.01 | 2.03 | 0.02 | 1.05 | 0.01 | 0.04 |
| True | 1 | 0 | 2 | 1 | 0 | 0 |
| Recovered | 1.01 | 0.01 | 2.02 | 1.02 | 0.02 | 0.00 |
| True | 1 | 0 | 0 | 2 | 0 | 0 |
| Recovered | 1.02 | 0.01 | 0.01 | 1.99 | 0.02 | 0.01 |
| True | 1 | 0 | 0 | 1 | 5 | 0 |
| Recovered | 1.01 | 0.04 | 0.04 | 1.02 | 4.89 | 0.02 |
| True | 1 | 0 | 0 | 1 | 0 | 5 |
| Recovered | 1.01 | 0.03 | 0.02 | 1.02 | 0.01 | 4.78 |
| True | 0.5 | 0.5 | 0.6 | 1.5 | 3 | 3 |
| Recovered | 0.48 | 0.49 | 0.62 | 1.52 | 3.02 | 3.10 |

ward multi-resolution approach. In addition, we can also provide plots of the distance fields $d^c$ and $d^d$, which enables a further analysis of image similarity; (see, e.g., Figure 2c). Note that in particular $d^c = 0$ everywhere if the images match perfectly. Therefore, if in some areas the function $d^c$ takes large values, we know that these areas did not register well.

## 2.5 Numerical Implementation

### 2.5.1 Evaluating Image Distance

While the mathematical framework is clear there are a few obstacles when trying to numerically implement it. Firstly the distance measures $d^c$ and $d^d$ are based on image gradients and therefore their derivatives involve second order image derivatives. This can be a problem since many medical images are noisy. The problems of working with noisy images and calculating their gradients are overcome by using smoothing B-splines approximations of the images. To smooth the image we use Tikhonov regularization where the regularization parameter is computed using the Generalized Cross Validation (GCV) criteria; cf. e.g. [24]. The GCV criteria also help to assess the noise level in the data and therefore for the choice of the edge parameter $\varepsilon$ in (10). For clean images (for images with very low noise), we pick the edge parameter to $h$, where $h$ is the discretization size.

Given the spline-smoothed image and the edge parameter, we are able to approximate the gradient of the image as

$$\nabla I \approx \left(\partial_1^h I, \ldots, \partial_d^h I\right),$$

where $\delta_k^h$ is a difference operator in the k-th direction. Here, similar to other numerical calculations of the absolute value of the gradient [19] we use the forward differences for the approximation. The regularized absolute value of the gradient is defined in a straightforward manner:

$$\|\nabla I\| \approx \sqrt{\sum_k (\partial_k^h I)^2 + \varepsilon^2}.$$

It is interesting to note that a simple calculation of the number of floating point operations required to compute the NGF is roughly 40N where N is the number of pixels/voxels in the image. Observe that this is substantially less than the number of FLOPS needed for the implementation of MI.

### 2.5.2 Numerical Optimization

To find the image deformation we need to minimize the distance function $f$ (cf. (3)) for the distances $Dc$ or $Dd$. Since this function is twice differentiable, we are able to use a Newton type method. The algorithm to compute the NGF and its derivatives is summarized in the pseudo-code of Algorithm 1. Note that the distance measure has a least-squares from

$$\mathcal{D} = -\tfrac{1}{2}\langle d, d\rangle_{L_2} \approx$$

$$-\tfrac{1}{2n}\sum_j d(x_j)^2 =: -\tfrac{1}{2n} d^\top d.$$

Therefore a natural optimization algorithm is the Gauss- Newton method [5]. To use the Gauss-Newton approach, we need to find the Jacobian of $d_c$ or $d_d$ with respect to T. Explicit formulae are given in Algorithm 1.

### 2.5.3 Grid Continuation

Like many other nonlinear problems, substantial computational advantage can be gained by using a multilevel continuation strategy. The idea of multilevel continuation is not new to image registration but most of the work on this topic assumes the sum of square differences as a distance measure. In general, a grid continuation method solves the optimization problem on a sequence of grids starting from the coarse grid. The solution on a finer grid is obtained by interpolating the coarse grid solution and using the interpolated result as a starting guess for the fine grid solution. Care must be taken such that the solution on the coarse grid represents a coarser version of the fine grid.

# 3. Results

In our numerical experience, we have used various examples using T1-T2 MRI images and MRI-CT images with results along the same lines. Since it is impossible to show every result we restrict ourselves to three illustrative and representative examples. For all our examples, we have used the multilevel algorithm. The coarsest grid was chosen to be $8 \times 8$. For parametric registration there is no need to prolong $\gamma$ but for nonparametric registration we use linear interpolation (see [11]). When coarsening, every grid is a factor of two coarser than the previous grid. As common in many multigrid methods, images on coarser grids are

generated by full weight interpolation [21].

In the first example we use the images in Figure 1. We take the T1 image and generate transformed versions of the image by using the affine linear transformation (2). We then use the transformed images to recover the original image. The advantage of this synthetic experiment is that it is controlled, i.e., we know the exact answer to the registration problem and therefore we are able to test our algorithm under a perfectly controlled environment. Running our code we obtain the results summarized in Table I. We see that overall we are able to accurately recover the shift parameters to the level of less than one pixel. On average we had errors of 0.34 of a pixel.

In the second example we use the images from Viola's Ph.D thesis [24]. In the original work a few thousands of iterations of stochastic optimization algorithm where needed to achieve registration using MI as a distance measure. Here, we have used a more efficient implementation of mutual information similar to the one presented in [16] in order to obtain competitive results. We then compared the results of both registrations. The difference between the MI registration and the NGF registration was less than 0.25 of a pixel, thus we conclude that the methods give virtually identical minima. However, to obtain the minima using MI we needed to use a random search technique to probe the space. This technique requires the estimation of many joint-density distributions and therefore it is rather slow.

When probing the space we have found many local minima. Furthermore, the local minima and the global minimum tend to have roughly the same magnitude. Thus, the "landscape" of the MI function for this example is similar to the one plotted in Figure 2. In comparison, our NGF algorithm used 15 iterations on the coarsest grid and 5 iterations on each finer grid. The registration was achieved in a 7.3 seconds and no special space probing was needed to obtain the minimum. In comparison, registration using MI needed roughly 23 seconds. The results of our experiments are also presented in Figure 5.

Another advantage of our method is the ability to quickly evaluate the registration result by looking at the absolute value of the cross-product $|nT \times nR|$. This product has

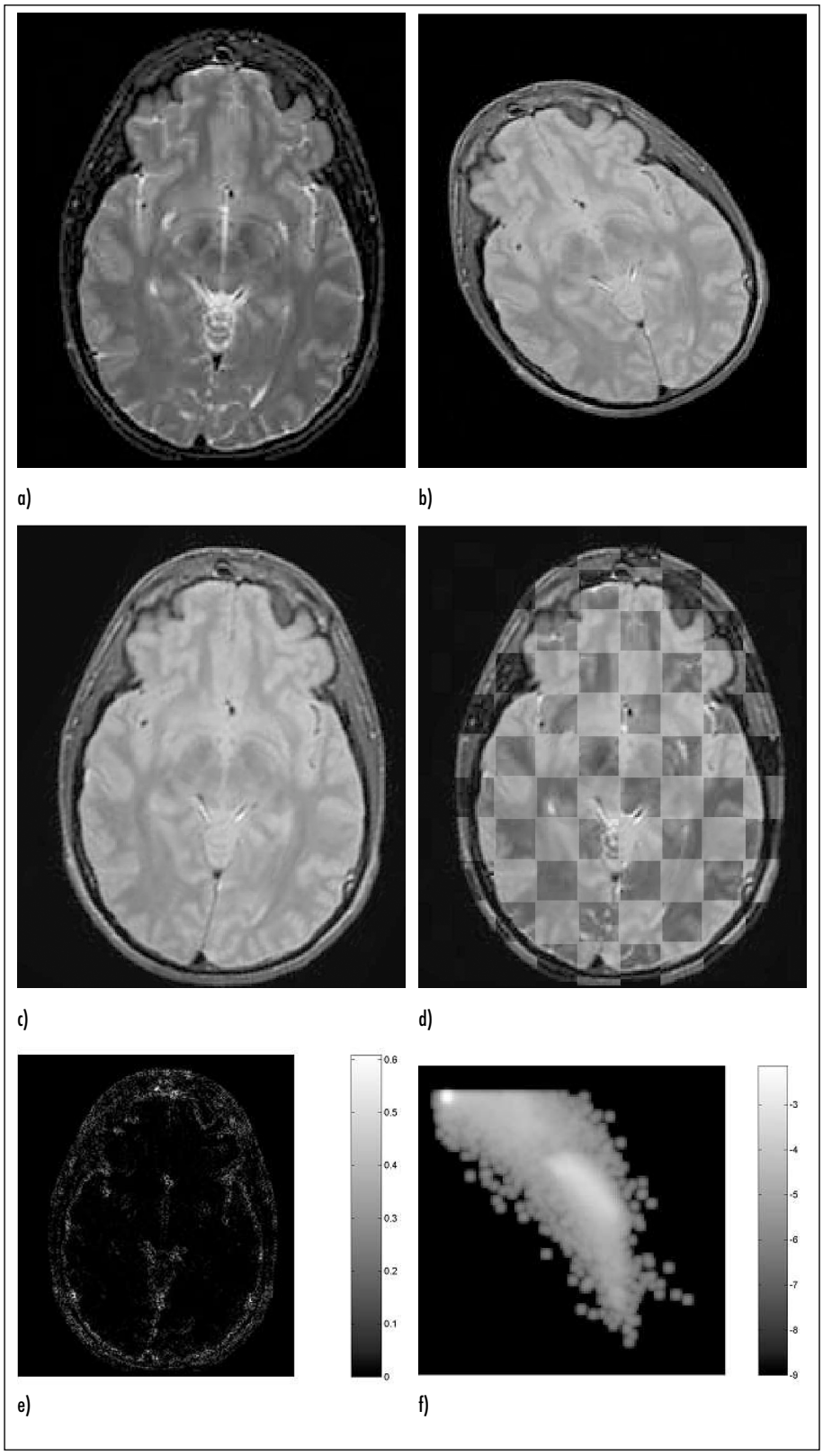


**Fig. 5** Experiments with Viola's example; a) reference $R$, b) template $T$, c) registered $T$, d) overlay of $T$ and $R$ ($20^2$ pixels checkerboard presentation), e) cross product $n_T \times n_R$, f) joint density at the minimum

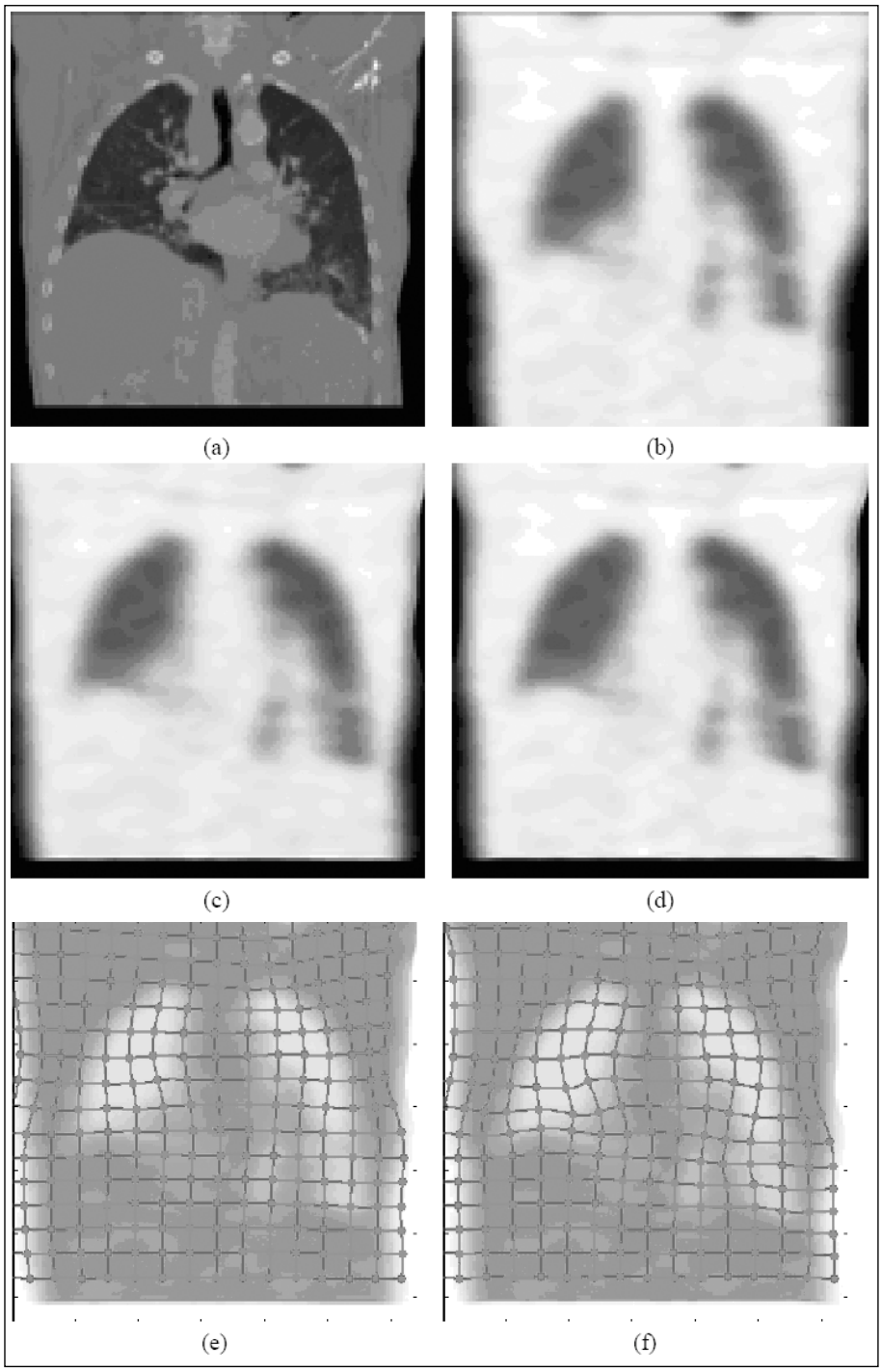


**Fig. 6** PET-CT registration

the same dimensions as the images and therefore can be viewed in the same scale. If the match is perfect then the cross product should vanish and therefore any deviation from zero implies an imperfect fit. Such deviations are expected to be present due to two different imaging processes, the noise in the images and possible additional nonlinear features. Figure 5e shows the cross product for the image matched above. It is evident that the matching is very good besides a small number of locations where we believe the image to be noisy.

In comparison, the final result in MI registration is a joint density map. This map does not have the same dimensions of the image but rather it has the dimensions of the discretization of the gray value spaces. It does not provide direct information on spatial locations. Therefore, it is hard to evaluate the success of the registration based on this map. The log of the joint density map for the same example is shown in Figure 5f. It is evident that it is not intuitive and hard to interpret.

The previous two examples demonstrate the ability of our new algorithm to successfully perform parametric registration. Nevertheless, when the number of parameters is very large, stochastic optimization techniques stall, and differentiable smooth distance measure functions are much more important. We therefore tested our algorithm on a PET/CT registration, see Figure 6. We performed an elastic registration of CT and transmission PET images of a human thorax. The CT image is in Figure 6a and the PET image is presented in Figure 6b. The transformed PET image is presented in Figure 6c and it is visually similar to the CT image. As mentioned earlier, we do not intent to discuss the results from a medical point of view. However, the displayed deformed grid indicates that, for this example, the differences between the MI and NGF approach are very small; cf. Figure 6 (e and f). For the NGF no special effort was needed in order to register the images. For MI we needed a good starting guess to converge to a local minimum.

## 4. Conclusion

Mutual information is to be considered as state-of-the-art distance measure for multimodal image registration. The measure has proven to be successful although it has a number of well-known disadvantages: it is highly non-convex, with typically many pronounced local maxima, it is naturally discrete, based on a nontrivial density, requires some critical smoothing parameters, and a common implementation does not exist.

We therefore presented an alternative distance measure which is better suited for

optimization. The new measure is based on normalized gradients and therefore naturally links spatial information to image distance. For the examples we have in this work, we show that the alternative approach is deterministic, much simpler, easier to interpret, fast and straightforward to implement, faster to compute, and also much more suitable to numerical optimization. The performance of the new approach is demonstrated by a few numerical examples.

Since our approach is based on matching edge information it can be sensitive to problems where edges in one image do not match any edges in the other. For example, it is difficult to register images obtained from emission tomography, where objects in the image are based on activity to transmission tomography, MRI or CT where edges are based on anatomy. In future work we intend to further test and develop our approach for other imaging modalities and to further improve our abilities to deal with local minima and noise.

## References

1. Ascher U, Haber E, Haung H. On effective methods for implicit piecewise smooth surface recovery. SIAM J of Scientific Computing 2006; 28 (1); 339-358.
2. Gottesfeld Brown L. A survey of image registration techniques. ACM Computing Surveys 1992; 24 (4); 325-376.
3. Cocosco CA, Kollokian V, Kwan RK, Evans AC. Brain-Web MR simulator. Available at http://www.bic.mni.mcgill.ca/brainweb/.
4. Collignon A, Vandermeulen A, Suetens P, Marchal G. Multimodality medical image registration based on information theory. Computational Imaging and Vision 1995; 3; 263-274.
5. Dennis JE, Schnabel RB. Numerical methods for unconstrained optimization and nonlinear equations. Philadelphia: SIAM; 1996.
6. Droske M, Rumpf M. A variational approach to non-rigid morphological registration. SIAM Appl Math 2004; 64 (2); 668-687.
7. Fitzpatrick JM, Hill D LG, Maurer CR. Image registration. Handbook of medical imaging, Volume 2: Medical Image Processing and Analysis, 2000; SPIE, pp 447-513.
8. Gallardo LA, Meju MA. Characterization of heterogeneous nearsurface materials by joint 2d inversion of dc resistivity and seismic data, Geophys Res Lett 2003; 30 (13); 1658-1664.
9. Golub G, Heath M, Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 1979; 21; 215-223.
10. Haber E, Oldenburg D. Joint inversion a structural approach. Inverse Problems 1997; 13; 63-67.
11. Haber E, Modersitzki J. A multilevel method for image registration. SIAM J of Scientific Computing 2006; 27 (5); 1594-1607.
12. Lehmann TM, Meinzer HP, Tolxdorff T. Advances in biomedical image analysis – past, present and future challenges. Methods Inf Med 2004; 43 (4); 308-314.
13. Lester H, Arridge SR. A survey of hierarchical non-linear medical image registration. Pattern Recognition 1999; 32; 129-149.
14. Modersitzki J. Numerical methods for image registration. Oxford, 2004.
15. Park H, Bland PH, Brock KK, Meyer CR. Adaptive registration using local information measures. Medical Image Analysis 2004; 8; 465-473.
16. Josien PW, Pluim JPW, Maintz JBA, Viergever MA. Image registration by maximization of combined mutual information and gradient information. IEEE TMI 2000; 19 (8); 809-814.
17. Pluim JPW, Maintz JBA, Viergever MA. Interpolation artifacts in mutual information based image registration. CVIV 2000; 77 (2); 211-232.
18. Pluim JPW, Maintz JBA, Viergever MA. Mutual-information-based registration of medical images: a survey. IEEE Transactions on Medical Imaging 1999; 22 (9): 986-1004.
19. Rudin L, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Proceedings of the eleventh annual international conference of the Center for Nonlinear Studies on Experimental mathematics: computational issues in nonlinear science, 1992. pp 259-268.
20. Silverman R. Density estimation for statistics and data analysis. Chapman & Hall, 1992.
21. Trottenberg U, Oosterlee C, Schuller A. Multigrid. Academic Press, 2001.
22. Unser M, Thévenaz P. Stochastic sampling for computing the mutual information of two images. Proceedings of the Fifth International Workshop on Sampling Theory and Applications (SampTA'03) (Strobl, Austria), 2003; pp 102-109.
23. Viola PA. Alignment by maximization of mutual information. Ph.D. thesis. Massachusetts Institute of Technology, 1995.
24. Wahba G. Spline models for observational data. SIAM, Philadelphia, 1990.
25. Yoo TS. Insight into images: Principles and practice for segmentation, registration, and image analysis. AK Peters Ltd, 2004.
26. Zhang J, Morgan FD. Joint seismic and electrical tomography, Proceedings of the EEGS Symposium on Applications of Geophysics to Engineering and Environmental Problems, 1996; pp 391-396.
27. Zitová B, Flusser J. Image registration methods: a survey. Image and Vision Computing 2003; 21 (11); 977-1000.

**Correspondence to:**
Eldad Haber
Department of Mathematics and Computer Science
Emory University
Atlanta, GA 30233
USA
E-mail: haber@mathcs.emory.edu