Intensity gradient based registration and fusion of multi-modal images

Eldad Haber Jan Modersitzki

Abstract— One of the remaining challenges 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, it 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 minima.

This observation motivate us to seek a different image similarity measure which is better suited for optimization but as well capable to handle multi-modal images. In this work we investigate an alternative distance measure which is based on normalized gradients. 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 optimization.

I. INTRODUCTION

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 instances that demand for 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 fusion. Typical examples include the fusion of computer tomography (CT) and magnetic resonance (MRI) images or of CT and positron emission tomography. 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], [14], [7], [23], [15] and references therein).

One of the remaining challenges in image registration arises for multi-modal 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,

Dept. of Mathematics and Computer Science, Emory University, Atlanta GA 30322
haber@mathcs.emory.edu.
Institute of Mathematics, Wallstraße 40, D-23564 Lübeck, Germany,
modersitzki@math.uni-luebeck.de.

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], [20]. 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 multi-modal image registration.

However, mutual information has a number of well-known drawbacks; cf. e.g., [17], [16], [19]. Firstly, mutual information is known to be highly non-convex and has typically many local minima; see for example the discussion in [15, §6.6] and Section III. 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 has a naturally discrete nature. However, fast and efficient registration schemes rely on powerful optimization techniques and thus on smooth functions. Thirdly, since mutual information is defined via the typically unaccessible joint density of the gray value distribution, approximations of the density are required. These approximations typically involve some very sensitive smoothing parameters (e.g. a binning size or a Parzen window width). 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 even common implementation for mutual information and its derivatives.

These difficulties had stem 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, [17]. As a result, a practical implementation of mutual information is highly non-trivial.

These observations motivate us to seek a different image similarity measure which is capable to handle multi-modal images but better suited for optimization. In this paper we investigate an alternative distance measure which is based on normalized gradients. 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 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 [11], [22], [8], [9] 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].

The goal of this work is to provide a new approach to multi-modal image registration using properties of differential geometry to characterize similarity between two images. The paper is organized as follows: In Section II we shortly lay the mathematical foundation of image registration. Section III presents an illustrative example showing some of the drawbacks of mutual information. In Section IV we discuss the proposed alternative image distance measure. In Section V we discuss its numerical implementation and lie out a simple algorithm to solve the multi-modal registration problem. Finally, in Section VI 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.

II. 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. With $d \in \mathbf{N}$ we denote the spatial dimension of the given images $R, T : \mathbb{R}^d \to \mathbb{R}$ which are assumed to be sufficiently smooth. Thus, $T(\boldsymbol{x})$ gives a gray value at a spatial position $\boldsymbol{x}$. Moreover, we assume that the contents of the images are contained in a bounded domain Ω, where without loss of generality we chose $\Omega := (0,1)^d$. For points $\boldsymbol{x} \notin \Omega$, we thus assume R and T to be constant, i.e., with two constants c_1, c_2, $R(\boldsymbol{x}) = c_1$ and $T(\boldsymbol{x}) = c_2$ for $\boldsymbol{x} \notin \Omega$.

Note that 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 [12]. 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) [10]. For data interpolation using B-splines see [21]. To minimize notational overhead, the continuous smooth approximations are also denoted by R and T, respectively.

As described in [15], 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 $\boldsymbol{\varphi} : \mathbb{R}^d \to \mathbb{R}^d$ can be parameterized in terms of some basis functions $\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_m$ and parameters $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$,

$$\boldsymbol{\varphi}(\boldsymbol{\gamma}, \boldsymbol{x}) = \sum_{k=1}^{m} \gamma_k \boldsymbol{\varphi}_k(\boldsymbol{x}); \qquad (1)$$

see [15] for details. A typical example is the so-called linear registration, where for some appropriate chosen basis functions and $d = 2$,

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

Given a distance measure $\mathcal{D}$, the registration problem is to find a minimizer $\boldsymbol{\gamma}$ of

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

III. 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 [13] for details. Note that the joint density is completely unrelated to the spatial image content. We now slide the template along the horizontal axis. In the language of equation (2), we fix $\gamma_{1,\ldots,5}$ and obtain the transformed image by changing γ_6. Figure 2b) shows the 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 minima. 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 minima is roughly equivalent to the size of the global minima.

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 difference is 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.

(For the computation of mutual information we used the $\cos^4$ kernel function with $\sigma = 32$, a midpoint quadrature rule for the integral approximations with a equidistant discretization of the grayvalue range $[-30, 285]$ and 64 discretization points, and a stabilizing tolerance $\mathrm{tol} = 10^{-3}$; see [13] for details. For the normalized gradient field we used a edge parameter $\mathcal{E} = 100$; see below.)

IV. A SIMPLE AND ROBUST ALTERNATIVE

The alternative multi-modal distance measure is based on the following simple though general interpretation of similarity:

> two image are considered to be similar, if intensity changes occur at the same locations.

An image intensity change can be detected via the image gradient. However, since the magnitudes of changes might be related to the imaging devices, and are not related to image differences, it is not advisatory to directly base a

distance measure on gradients. We therefore consider first the normalized gradient field

$$\mathbf{n}(I, \boldsymbol{x}) := \begin{array}{l} \frac{\nabla I(\boldsymbol{x})}{\|\nabla I(\boldsymbol{x})\|}, \quad \text{if} \quad \nabla I(\boldsymbol{x}) \neq \mathbf{0}, \\[2mm] \mathbf{n}(I, \boldsymbol{x}) := 0, \quad \text{otherwise.} \end{array} \tag{4}$$

As usual, for $\boldsymbol{x} \in \mathbb{R}^d$ we set

$$\|\boldsymbol{x}\| = \sqrt{\sum_{\ell=1}^{d} x_\ell^2} \quad \text{and} \quad \nabla I := (\partial_1 I, \dots, \partial_d I)^\top. \tag{5}$$

For two related points $\boldsymbol{x}$ in R and $\boldsymbol{\varphi}(\boldsymbol{x})$ in T or, equivalently, $\boldsymbol{x}$ in $T \circ \boldsymbol{\varphi}$, we look at the vectors $\mathbf{n}(R, \boldsymbol{x})$ and $\mathbf{n}(T \circ \boldsymbol{\varphi}, \boldsymbol{x})$. These two vectors form an angle $\theta(\boldsymbol{x})$. 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.

This observation motivate the following distance measures

$$\begin{aligned} d^c(T, R) &= \|\mathbf{n}(R, \boldsymbol{x}) \times \mathbf{n}(T, \boldsymbol{x})\|^2, \\ \mathcal{D}^c(T, R) &= \frac{1}{2} \int_\Omega d^c(T, R) \; d\boldsymbol{x}, \end{aligned} \tag{6}$$

$$\begin{aligned} d^d(T, R) &= \langle \mathbf{n}(R, \boldsymbol{x}), \mathbf{n}(T, \boldsymbol{x}) \rangle^2, \\ \mathcal{D}^d(T, R) &= -\frac{1}{2} \int_\Omega d^d(T, R) \; d\boldsymbol{x}, \end{aligned} \tag{7}$$

Note that from an optimization point of view, the distances $\mathcal{D}^c$ or $\mathcal{D}^d$ are equivalent.

The definition of the normalized gradient field (4) 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 following regularized normalized gradient fields

$$\mathbf{n}_{\mathcal{E}}(I, \boldsymbol{x}) := \frac{\nabla I(\boldsymbol{x})}{\|\nabla I(\boldsymbol{x})\|_{\mathcal{E}}}, \tag{8}$$

$$\|\nabla I(\boldsymbol{x})\|_{\mathcal{E}} := \sqrt{\nabla I(\boldsymbol{x})^\top \nabla I(\boldsymbol{x}) + \mathcal{E}^2}.$$

In regions where $\mathcal{E}$ is much larger than the gradients the maps $\mathbf{n}_{\mathcal{E}}(I, \boldsymbol{x})$ are almost zero and therefore do not have a significant effect of the measures $\mathcal{D}^c$ or $\mathcal{D}^d$, respectively. However, in regions where $\mathcal{E}$ is much smaller than the gradients, the regularized maps are close to the non-regularized ones and these regions make a substantial difference in the calculation of the measures d^c and d^d. Indeed, the choice of the parameter $\mathcal{E}$ in (8) answers the question, "what can be interpreted as a jump"? Thus, it is the minimal size of a local change in I which is still interpreted as a jump. Such an interpretation is desirable, particularly if true edges are to be respected, but it must be avoided in a band, where the resolution limitation

does not allow a true distinction between what may be zero and what is not.

With this in mind and similarly to [1], we propose the following *automatic choice*:

$$\mathcal{E} = \frac{\eta}{V} \int_\Omega |\nabla I(\boldsymbol{x})| \; d\boldsymbol{x}, \tag{9}$$

where η is the estimated noise level in the image and V is the volume of the domain Ω.

The measures (6) and (7) 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 straightforward 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.

V. NUMERICAL IMPLEMENTATION

A. Evaluating image distance

While the mathematical framework is clear there are a few obstacles when trying to numerically implement it. Firstly, 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. [21].

The GCV criteria also help to assess the noise level in the data and therefore for the choice of the edge parameter $\mathcal{E}$ in (9). 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 $\mathcal{E}$ we are able to approximate the gradient of the image as

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

where ∂_k^h is a difference operator in the k-th direction. Here, similar to other numerical calculations of the absolute value of the gradient [18] we use the forward differences for the approximation. The regularized absolute value of the gradient is defined in a straightforward manor:

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

B. Numerical Optimization

To find the image deformation we need to minimize $f(\boldsymbol{\gamma})$ (cf. (3)) for $\mathcal{D}^c$ or $\mathcal{D}^d$. Since this function is twice differentiable with respect to $\boldsymbol{\gamma}$, 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,

$$\mathcal{D} = -\frac{1}{2} \langle d, d \rangle_{L_2} \approx -\frac{1}{2n} \sum_j d(\boldsymbol{x}_j)^2 =: -\frac{1}{2n} \boldsymbol{d}^\top \boldsymbol{d}.$$

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

Algorithm 1 Calculation of Normalized gradient fields and their derivatives:

$[D, \boldsymbol{d}, \boldsymbol{d}_T] \leftarrow \mathrm{NGF}(R, T, \mathcal{E});$

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 $\sqrt{\cdot}$)

$$\|\nabla^h R\|_{\mathcal{E}} \quad \leftarrow \quad \sqrt{\sum (\partial_i^h R)^2 + \mathcal{E}^2},$$

$$\|\nabla^h T\|_{\mathcal{E}} \quad \leftarrow \quad \sqrt{\sum (\partial_i^h T)^2 + \mathcal{E}^2},$$

$$\boldsymbol{d} \quad \leftarrow \quad \mathrm{diag}(1./\|\nabla^h R\|_{\mathcal{E}}) \cdot$$
$$\mathrm{diag}(1./\|\nabla^h T\|_{\mathcal{E}}) \left(\sum \mathrm{diag}(\partial_i^h R)(\partial_i^h T) \right),$$

$$D \quad \leftarrow \quad -\frac{1}{2n} \boldsymbol{d}^\top \boldsymbol{d},$$

if derivatives are needed compute

$$[\|\nabla^h T\|_{\mathcal{E}}]_T \quad \leftarrow \quad \mathrm{diag}(1./\|\nabla^h T\|_{\mathcal{E}}) \left(\sum \mathrm{diag}(\partial_i^h T) \partial_i^h \right)$$

$$\boldsymbol{d}_T \quad \leftarrow \quad \mathrm{diag}(1./\|\nabla^h R\|_{\mathcal{E}}) \cdot \mathrm{diag}(1./\|\nabla^h T\|_{\mathcal{E}}) \cdot$$
$$\left(\sum \mathrm{diag}(\partial_i^h R) \partial_i^h \right) -$$
$$\mathrm{diag}(\boldsymbol{d}) \cdot \mathrm{diag}(1./\|\nabla^h T\|_{\mathcal{E}}) \cdot [\|\nabla^h T\|_{\mathcal{E}}]_T.$$

C. 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. We summarize our multilevel continuation in Algorithm 2.

VI. NUMERICAL EXPERIMENTS

We have done extensive experiments with our method however, to demonstrate the effectiveness of our algorithm we use two different examples. 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 γ and the original image. The advantage of this experiment is that it is controlled and we know the exact answer and therefore

---

Algorithm 2 Multilevel image registration: $\boldsymbol{u} \leftarrow \mathtt{MLIR}$

for level = coarse to fine **do**
  if level=coarse **then**
    choose $\gamma^{(0)}$ on the coarsest grid;
  else
    prolongate γ^* to the finer grid: $\gamma^{(0)} \leftarrow$ prolongate(γ^*);
  end if
  solve the registration problem and obtain γ^* on this grid.
end for

---

we are able to test our algorithm under a perfectly controlled environment. Running our code we obtain the following results summarized in Table I. We see that overall we are able to

TABLE I
EXPERIMENTS WITH IMAGE 1. CHOSEN VS RECOVERED γ

	γ_1	γ_2	γ_3	γ_4	γ_5	γ_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

accurately recover the shift parameters to the level of less than one pixel.

In the second example we use the images from Viola's Ph.D thesis [20]. In the original work a few thousands of iterations of stochastic optimization algorithm where needed to achieve registration using MI as a distance measure. We have used a more efficient implementation of mutual information (see [13]) to obtain the same goal. We then compare the results of both registered images. 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 distribution and therefore it is rather slow. When probing the space we have found many local minima. Furthermore, the local minima and the global minima tend to have roughly the same magnitude. The global minima has the value of about -9.250×10^{-2} while the guess $\gamma = 0$ has the value of about -9.115×10^{-2}. 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 iteration on the coarse grid which is 22×24 and 5 iterations on each finer grid. The registration was achieved in a matter of seconds and no special space probing was needed to obtain the minima. The value of the NGF function at $\gamma = 0$ was -4.63×10^1 while

at the minima its value was -2.16×10^2 thus our minima is much deeper compared with the MI minima. The results of our experiments are also presented in Figure 3.

Another advantage of our method is the ability to quickly evaluate the registration result by looking at the absolute value of the cross-product $|\boldsymbol{n}_T \times \boldsymbol{n}_R|$. This product has the same dimension as the images and therefore can be viewed in the same scale. If the match is prefect then the cross product should vanish and therefore any deviation from zero implies a 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 3e) 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 3f). It is evident that it is not intuitive and hard to interpret.

VII. Conclusion

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

We therefore presented an alternative distance measure which is capable to handle multi-modal images but better suited for optimization. The new measure is based on normalized gradients and therefore naturally links spatial information to image distance. 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.

References

[1] U. Ascher, E. Haber and H. Haung, *On effective methods for implicit piecewise smooth surface recovery*, Submitted, (2004).

[2] Lisa Gottesfeld Brown, *A survey of image registration techniques*, ACM Computing Surveys **24** (1992), no. 4, 325–376.

[3] C.A. Cocosco, V. Kollokian, R.K.-S. Kwan, and A.C. Evans, *Brain-Web MR simulator*, Available at http://www.bic.mni.mcgill.ca/brainweb/.

[4] A. Collignon, A. Vandermeulen, P. Suetens, and G. Marchal, *multi-modality medical image registration based on information theory*, Kluwer Academic Publishers: Computational Imaging and Vision **3** (1995), 263–274.

[5] J. E. Dennis and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, SIAM, Philadelphia, 1996.

[6] M. Droske and M. Rumpf, *A variational approach to non-rigid morphological registration*, SIAM Appl. Math. **64** (2004), no. 2, 668–687.

[7] J. M. Fitzpatrick, D. L. G. Hill, and C. R. Maurer Jr., *Image registration, handbook of medical imaging*, Volume 2: Medical Image Processing and Analysis, SPIE (2000), 447–513.

[8] L.A. Gallardo and M.A. Meju, *Characterization of heterogeneous near-surface materials by joint 2d inversion of dc resistivity and seismic data*, Geophys. Res. Lett. **30** (2003), no. 13, 1658–1664.

[9] ———, *Joint two-dimensional dc resistivity and seismic traveltime inversion with cross-gradients constraints*, J. Geophys. Res. **109B** (2004), 3311–3315.

[10] G. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics **21** (1979), 215–223.

[11] E. Haber and D. Oldenburg, *Joint inversion a structural approach*, Inverse Problems **13** (1997), 63–67.

[12] Eldad Haber and Jan Modersitzki, *A multilevel method for image registration*, Technical Report TR-2004-005-A, Department of Mathematics and Computer Science, Emory University, Atlanta GA 30322, May 2004, Submitted to SIAM J. of Scientific Computing.

[13] Eldad Haber, Jan Modersitzki, and Stefan Heldmann, *Computational methods for mutual information based registration*, Technical Report TR-2004-015-A, Department of Mathematics and Computer Science, Emory University, Atlanta GA 30322, Jun 2004, Submitted to Inverse Problems.

[14] Hava Lester and Simon R. Arridge, *A survey of hierarchical non-linear medical image registration*, Pattern Recognition **32** (1999), 129–149.

[15] J. Modersitzki, *Numerical methods for image registration*, Oxford, 2004.

[16] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever, *Interpolation artefacts in mutual information based image registration*, Proceedings of the SPIE 2004, Medical Imaging, 1999 (K.M. Hanson, ed.), vol. 3661, SPIE, 1999, pp. 56–65.

[17] ———, *Mutual-information-based registration of medical images: a survey*, IEEE Transactions on Medical Imaging **22**, 1999, 986–1004.

[18] L. Rudin, S. Osher and E. Fatemi, *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, 259–268.

[19] M. Unser and P. Thévenaz, *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), May 26-30, 2003, pp. 102–109.

[20] Paul A. Viola, *Alignment by maximization of mutual information*, Ph.D. thesis, Massachusetts Institute of Technology, 1995.

[21] G. Wahba, *Spline models for observational data*, SIAM, Philadelphia, 1990.

[22] J. Zhang and F.D. Morgan, *Joint seismic and electrical tomography*, Proceedings of the EEGS Symposium on Applications of Geophysics to Engineering and Environmental Problems, 1996, pp. 391–396.

[23] Barbara Zitová and Jan Flusser, *Image registration methods: a survey*, Image and Vision Computing **21** (2003), no. 11, 977–1000.

(a)

(b)
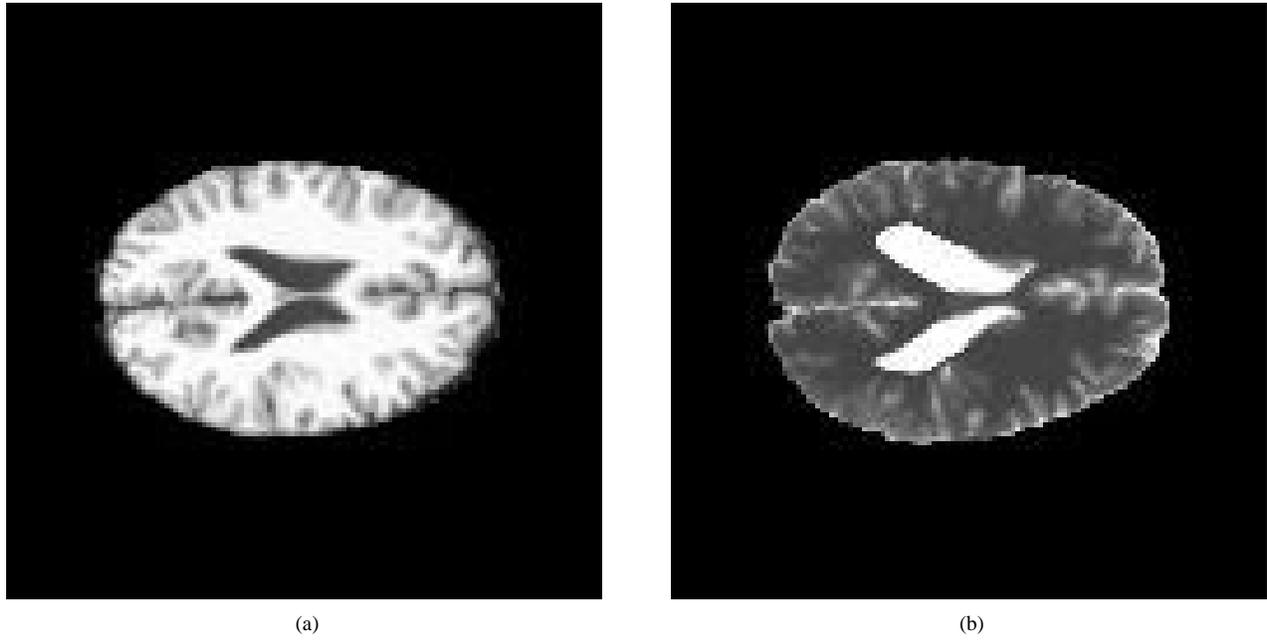
Fig. 1. Original BrainWeb [3] T1 (LEFT) and T2 (RIGHT) images

(a)

(b)

(c)

(d)
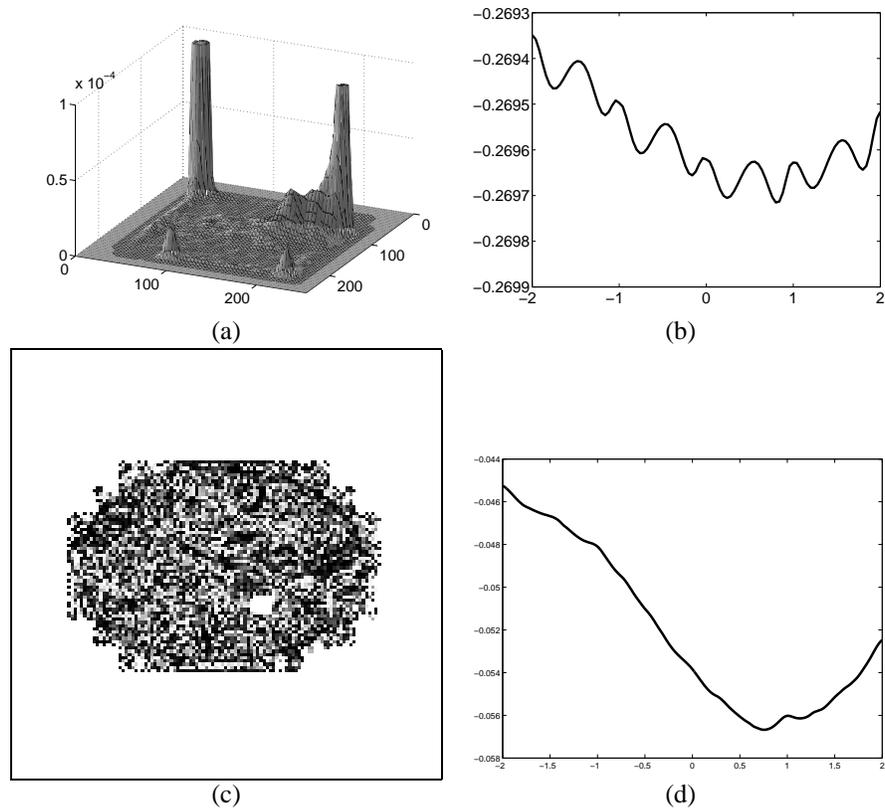
Fig. 2. Distance measures versus shifts; (a) the joint density approximation for R and T, (b) mutual information versus shift, (c) the normalized gradient field for R and T, (d) normalized gradient field versus shift.
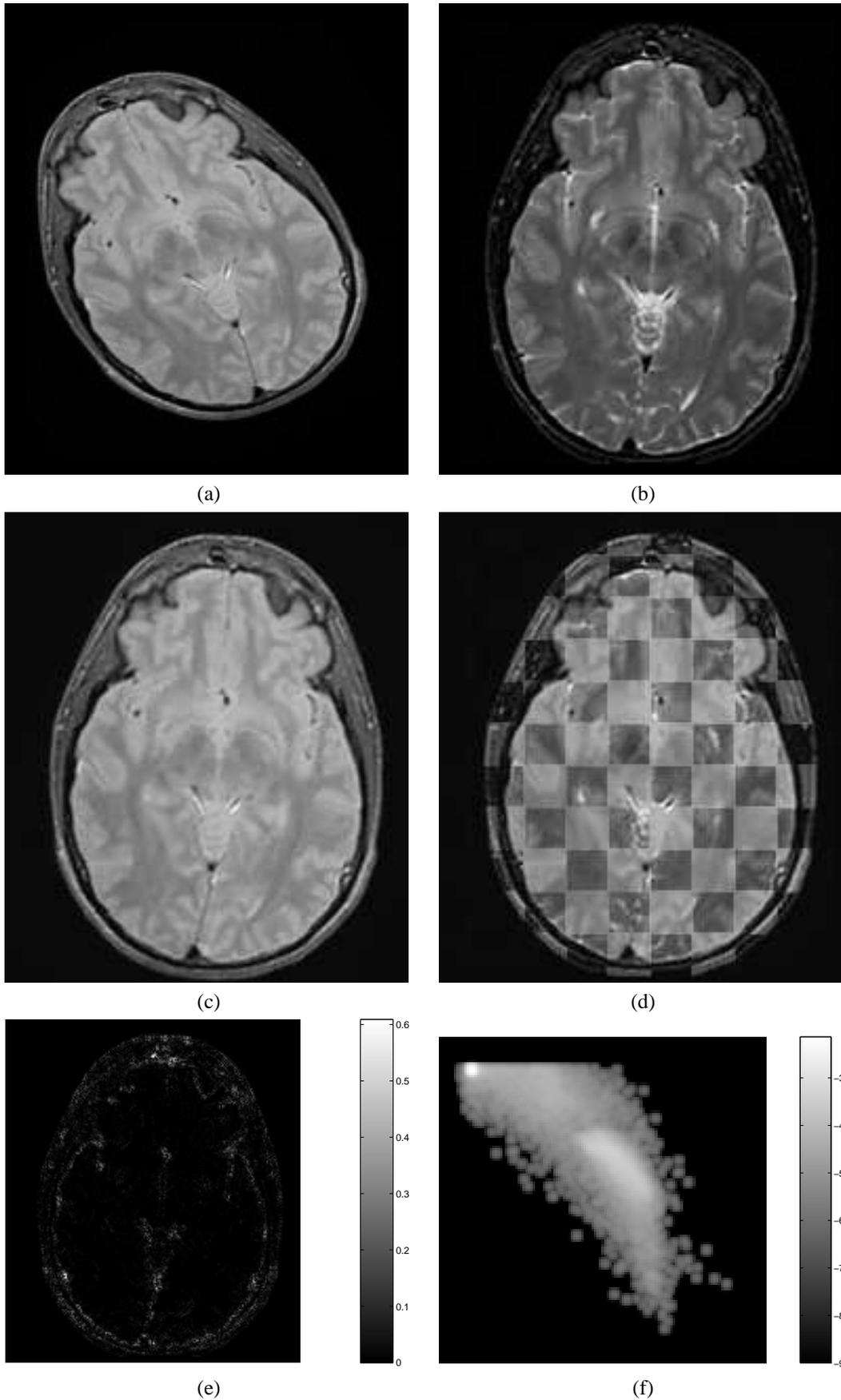
Fig. 3. Experiments with Viola's example; (a) template T, (b) reference R, (c) registered T, (d) overlay of T and R (20^2 pixels checkerboard presentation), (e) cross product $\boldsymbol{n}_T \times \boldsymbol{n}_R$, (f) joint density at the minimum.