Three-Dimensional Multimodal Brain Warping Using the Demons Algorithm and Adaptive Intensity Corrections

Alexandre Guimond*, Alexis Roche, Nicholas Ayache, and Jean Meunier

Abstract—**This paper presents an original method for three-dimensional elastic registration of multimodal images. We propose to make use of a scheme that iterates between correcting for intensity differences between images and performing standard monomodal registration. The core of our contribution resides in providing a method that finds the transformation that maps the intensities of one image to those of another. It makes the assumption that there are at most two functional dependencies between the intensities of structures present in the images to register, and relies on robust estimation techniques to evaluate these functions. We provide results showing successful registration between several imaging modalities involving segmentations, T1 magnetic resonance (MR), T2 MR, proton density (PD) MR and computed tomography (CT). We also argue that our intensity modeling may be more appropriate than mutual information (MI) in the context of evaluating high-dimensional deformations, as it puts more constraints on the parameters to be estimated and, thus, permits a better search of the parameter space.**

Index Terms—**Elastic registration, intensity correction, medical imaging, multimodality, robust estimation.**

I. Introduction

OVER the last decade, automatic registration techniques of medical images of the head have been developed following two main trends: 1) registration of multimodal images using low degree transformations (rigid or affine); 2) registration of monomodal images using high-dimensional volumetric maps (elastic or fluid deformations). The first category mainly addresses the fusion of complementary information obtained from different imaging modalities. The second category's predominant purpose is the evaluation of either the anatomical evolution process present in a particular subject or of anatomical variations between different subjects.

Manuscript received November 2, 1999; revised October 20, 2000. The work of A. Guimond was supported in part by Le Fonds FCAR. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was R. Leahy. *Asterisk indicates corresponding author.*

*A. Guimond is with the Projet Epidaure, INRIA, 2004 Route des Lucioles BP 93, 06902 Sophia Antipolis, France. He is also with the Département d'Informatique et recherche opérationnelle, Université de Montréal, CP 6128 succ Centre-Ville, Montréal QC H3C 3J7, Canada. He can currently be reached at Brigham and Women's Hospital, Center for Neurological Imaging, 221 Longwood Avenue, Boston, MA 02115, USA (e-mail: guimond@bwh.harvard.edu).

A. Roche and N. Ayache are with the Projet Epidaure, INRIA, 2004 Route des Lucioles BP 93, 06902 Sophia Antipolis, France.

J. Meunier is with the Département d'Informatique et recherche opérationnelle, Université de Montréal, CP 6128 succ Centre-Ville, Montréal QC H3C 3J7, Canada.

Publisher Item Identifier S 0278-0062(01)00797-2.

These two trends have evolved separately mainly because the combined problem of identifying complex intensity correspondences along with a high-dimensional geometrical transformation defines a search space arduous to traverse. Recently, three groups have imposed different constraints on the search space, enabling them to develop automatic multimodal non-affine registration techniques. All three methods make use of block matching techniques to evaluate local translations. Two of them use mutual information (MI) [1], [2] as a similarity measure and the other employs the correlation ratio [3].

The major difficulty in using MI as a similarity measure for registration is to compute the conditional probabilities of one image's intensities with respect to those of the other. To do so, Maintz *et al.* [4] proposed to use conditional probabilities after rigid matching of the images as an estimate of the real conditional probabilities after local transformations. Hence, the probabilities are evaluated only once before fluid registration. However, Gaens *et al.* [5] argued that the assumption that probabilities computed after affine registration are good approximations of the same probabilities after fluid matching is unsuitable. They also proposed a method in which local displacements are found so that the global MI increases at each iteration, permitting incremental changes of the probabilities during registration. Their method necessitates the computation of conditional probabilities over the whole image for every voxel displacement. To alleviate themselves from such computations owing to the fact that MI requires many samples to estimate probabilities, Lau *et al.* [6] have chosen a different similarity measure. Due to the robustness of the correlation ratio with regards to sparse data [3], they employed it to assess the similarity of neighboring blocks. Hence no global computation is required when moving subregions of the image.

Our method distinguishes itself by looking at the problem from a different angle. In the last years, our group has had some success with monomodal image registration using the demons method [7], [8] an optical flow variant when dealing with monomodal volumetric images. If we were able to model the imaging processes that created the images to register, and assuming these processes are invertible, one could transform one of the images so that they are both represented in the same modality. Then we could use our monomodal registration algorithm to register them. We have, thus, developed a completely automatic method to transform the different structures intensities in one image so that they match the intensities of the corresponding structures in another image, and this without resorting to any segmentation method.

0278–0062/01$10.00 © 2001 IEEE

The rational behind our formulation is that there is a functional relationship between the intensity of a majority of structures when imaged with different modalities. This assumption is partly justified by the fact that the Woods criterion [9] as well as the correlation ratio [3] which evaluate a functional dependence between the images to match, have been used with success in the past, and sometimes lead to better results than MI [10], [11], which assumes a more general constraint.

The idea of estimating an intensity transformation during registration is not new in itself. For example, Feldmar *et al.* [12] as well as Barber [13] have both published methods in which intensity corrections are proposed. These methods restrict themselves to estimating linear intensity transformations in a monomodal registration context. Friston *et al.* [14] also proposed a method to estimate spatial and intensity functions to put positron emission tomography (PET) and magnetic resonance (MR) images into register using a standard least squares solution. Their modeling of the problem is similar to ours but their solution requires segmentation of the images and is not robust to outliers. We propose here a registration model based on one or two high-degree polynomials found using a robust regression technique to enable the registration of raw images from different modalities.

The remaining sections of this paper are organized in the following manner. First, we detail our multimodal elastic registration method. We then describe what kind of images were used to test our method and how they were acquired. Next, results obtained by registering different images obtained from several modalities are presented and discussed, and future tracks are suggested.

II. Method

Our registration algorithm is iterative and each iteration consists of two parts. The first one transforms the intensities of anatomical structures of a source image S so that they match the intensities of the corresponding structures of a target image T. The second part regards the registration of S (after intensity transformation) with T using our elastic registration algorithm.

In the following, we first describe the three-dimensional (3-D) geometrical transformation model and then the intensity transformation model. We believe this ordering is more convenient since it is easier to see what result must provide the intensity transformation once the geometrical transformation procedure is clarified. However, we wish to make clear to the reader that *each iteration* of our registration method proceeds first by estimating the intensity transformation and then the geometrical transformation.

A. Modeling the Geometrical Transformation

Many algorithms have been developed that deform one brain so its shape matches that of another [15]. The procedure used in the following work was influenced by a variety of methods, primarily the demons algorithm [7], [8]. It finds the displacement $v(x)$ for each voxel x of T so it matches the corresponding anatomical location in S. The solution is found using the iterative scheme shown in (1) at the bottom of the page, where G_σ is a Gaussian filter with a variance of σ^2, $\otimes$ denotes the convolution, $\circ$ denotes the composition, ∇ is the gradient operator and the transformation $h(x)$ is related to the displacement by $h(x) = x + v(x)$.

Briefly, (1) finds voxel displacements in the gradient direction $\nabla(S \circ h_n)(x)$. These displacements are proportional to the intensity difference between $S \circ h_n(x)$ and $T(x)$ and are normalized for numerical stability. Convolution with a Gaussian kernel G_σ is performed to model a smoothly varying displacement field. As is common with registration methods, we also make use of multilevel techniques to accelerate convergence. Details about the number of levels and iterations as well as filter implementation issues are addressed in Section IV.

For details on how we obtained (1) and information on the relevance of the resulting transformation, see [16] and [17].

B. Modeling the Intensity Transformation

Prior to each iteration of the geometrical transformation an intensity correction is performed on S so that the intensity of its structures matches those in T, a requirement for the use of (1). The intensity correction process starts by defining the set C of intensity couples from corresponding voxels of T and of the current resampled source image $S \circ h$, which will be designated by S in this section for simplicity. Hence, the set C is defined as

$$C = \{(S(x), T(x)); 1 \leq x \leq N\} \tag{2}$$

where N is the number of voxels in the images. $S(x)$ and $T(x)$ correspond to the intensity value of the xth voxel of S and T, respectively, when adopting the customary convention of considering images as one-dimensional arrays. From there, we show how to perform intensity correction if we can assume that a single intensity value in S has either 1) exactly one corresponding intensity value in T (monofunctional dependence) or 2) at least one and at most two corresponding intensity values in T (bifunctional dependence).

1) Monofunctional Dependence Assumption: Our goal is to model the transformation that characterizes the mapping from voxel intensities in S to those in T, knowing that some elements of C are erroneous, i.e., that would not be present in C if S and T were perfectly matched. If we can assume a monofunctional dependence of the intensities of T with regards to the those of S as well as additive stationary Gaussian white noise η on the intensity values of T, then we can adopt the model

$$T(x) = f(S(x)) + \eta(x) \tag{3}$$

where f is an unknown function to be estimated. This is exactly the model employed by Roche *et al.* [10], [11] which leads to the

$$v_{n+1}(x) = G_\sigma \otimes \left(v_n + \frac{S \circ h_n(x) - T(x)}{\|\nabla(S \circ h_n)(x)\|^2 + [S \circ h_n(x) - T(x)]^2} \nabla(S_n^\star \circ h_n)(x) \right) \tag{1}$$

correlation ratio as the measure to be maximized for registration. In that approach, for a given transformation, the authors seek the function that best describes T in terms of S. They show that in a maximum likelihood context, the intensity function $\hat{f}$ that best approximates f is a least squares (LS) fit of T in terms of S.

The major difference between our respective problems is that we seek a high-dimensional geometrical transformation. As opposed to affine registration where the transformation is governed by the majority of good matches, the elastic registration model of (1) finds displacements using mainly local information (i.e., gradients, local averages, etc.). Hence, we cannot expect good displacements in one structure to correct for bad ones in another; we have to make certain each voxel is moved properly during each iteration. For this, since the geometrical transformation is found using intensity similarity, the most precise intensity transformation is required. Consequently, instead of performing a standard least squares regression, we have opted for a robust linear regression estimator which will remove outlying elements of C during the estimation of the intensity transformation. To estimate f we use the least trimmed squares (LTS) method followed by a binary reweighted least squares (RLS) estimation [18]. The combination of these two methods provides a very robust regression technique with outlier detection, while ensuring that a maximum of pertinent points are used for the final estimation.

a) LTS Computation: For our particular problem, we will constrain ourselves to the estimation of a polynomial function from the elements in C. We can then relate the intensity correspondences with a set of equations of the form

$$T(x) = \theta_0 + \theta_1 S(x) + \theta_2 S(x)^2 + \cdots + \theta_p S(x)^p \quad (4)$$

where $\boldsymbol{\theta} = [\theta_0, \ldots, \theta_p]$ needs to be estimated and p is the degree of the polynomial function. A regression estimator will provide a $\hat{\boldsymbol{\theta}} = [\hat{\theta}_0, \ldots, \hat{\theta}_p]$ which can be used to predict the value of $T(x)$ from $S(x)$

$$\hat{T}(x) = \hat{\theta}_0 + \hat{\theta}_1 S(x) + \hat{\theta}_2 S(x)^2 + \cdots + \hat{\theta}_p S(x)^p \quad (5)$$

as well as the residual errors

$$r(x) = T(x) - \hat{T}(x). \quad (6)$$

A popular method to obtain $\hat{\boldsymbol{\theta}}$ is to minimize the sum of squared residual errors

$$\min_{\hat{\boldsymbol{\theta}}} \sum_{x=1}^{N} r(x)^2 \quad (7)$$

which leads to the standard LS solution. It is found by solving a linear system using the singular value decomposition (SVD) method. See [11] for a more detailed study. This method is known to be very sensitive to outliers and, thus, is expected to provide a poor estimate of the monofunctional mapping from S to T. The LTS method solves this problem by minimizing the same sum on a subset of all residual errors, thus rejecting large ones corresponding to outliers

$$\min_{\hat{\boldsymbol{\theta}}} \sum_{x=1}^{c} \rho(x) \quad (8)$$

where $\rho(x)$ is the xth smallest value of the set $\{r(1)^2, \ldots, r(N)^2\}$. This corresponds to a standard LS on the c values that best approximates the function we are looking for. Essentially, c/N represents the percentage of "good" points in C and must be at least $(N + p + 2)/2N$, which is the breakdown value of the LTS method [19], i.e., the minimum percentage of points required for a proper estimate. A lesser value would allow to estimate parameters that model a minority of points which could then all be outliers. c will vary according to the modalities registered. The values used for our result and the corresponding modalities are discussed in Section IV.

We have not developed a procedure to find c automatically, but have found reasonable *ad hoc* estimates that provide satisfactory results. A more in depth analysis of the influence of c on our results and its relation with different modalities is beyond the scope of this article, but we have found from visual inspection that a variation of c by 5% does not influence the registration result in a significant way.

Our method for LTS minimization is a simple iterative technique. First, we randomly pick c points from C. We then iterate between calculating $\hat{\boldsymbol{\theta}}$ using the standard LS technique on the selected points and choosing the c closest points from C. Recently, Rousseeuw and Van Driessen [19] proved that this method reduces the error on $\hat{\boldsymbol{\theta}}$ at each iteration. Hence, we carry this process until $\Sigma_{x=1}^{c}\, \rho(x)$ stops decreasing, usually requiring less than five iterations. This provides adequate results for our purposes. Note that this finds a local minimum, although not guarantying the global minimum. Since we assume a good global registration previous to the estimation process, that might not make a large difference. Still, in the same paper, the authors also proposed a new efficient algorithm to find a good approximate solution of the LTS minimization. This strategy will be looked into in future implementations.

b) RLS Computation: Once $\hat{\boldsymbol{\theta}}$ is obtained using LTS, we can compute the standard error σ of our points with respect to our estimate. Of course, this value will also be an estimate corresponding to

$$\hat{\sigma} = \sqrt{\frac{1}{KN} \sum_{x=1}^{c} \rho(x)},$$
$$K = \frac{\displaystyle\int_{-\alpha}^{\alpha} x^2 g(x)\, dx}{\displaystyle\int_{-\alpha}^{\alpha} g(x)\, dx} = \frac{\displaystyle\int_{-\alpha}^{\alpha} x^2 g(x)\, dx}{c/N} \quad (9)$$

where $g(x)$ is the normalized Gaussian distribution ($\mu = 0, \sigma^2 = 1$) and α is the $(0.5 + c/2N)$th quantile of $g(x)$. In (9), K is a normalization factor introduced because $\sqrt{1/N\Sigma_{x=1}^{c}\rho(x)}$ is not a consistent estimator of σ since we only choose the c/N smallest $\rho(x)$.

Using this average deviation of the points from the estimates, we then perform an RLS regression by finding a new $\hat{\boldsymbol{\theta}}$ that minimizes the sum of squared residual for all points within $3\hat{\sigma}$ of the previous estimate,

$$\min_{\hat{\boldsymbol{\theta}}} \sum_{x=1}^{N} w_x \rho(x), \quad \text{where} \quad w_x = \begin{cases} 1, & \text{if} \rho(x) \le 3\hat{\sigma} \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

This optimizes the number of points used to compute $\hat{\boldsymbol{\theta}}$ by considering all the points that relate well to the LTS estimate, not only the best c/N.

2) Bifunctional Dependence Assumption: Functional dependence as expressed in (3) implicitly assumes that two structures having similar intensity ranges in S should also have similar intensity ranges in T. With some combinations of images, this is a crude approximation. For example, ventricles and bones generally give similar response values in a T1 weighted image while they appear with very distinct values in a CT scan. Conversely, white and gray matter are well contrasted in a T1 image while they correspond to similar intensities in a CT.

To circumvent this difficulty, we have developed a strategy that enables the mapping of an intensity value in S to not only one, but two possible intensity values in T. This method is a natural extension of the previous section. Instead of computing a single function that maps the intensities of S to those of T, two functions are estimated and the mapping becomes a weighted sum of these two functions.

We start with the assumption that if a point has an intensity s in S, the corresponding point in T has an intensity t that is normally distributed around two possible values depending on s, $f_1(s)$ and $f_2(s)$. In statistical terms, this means that given s, t is drawn from a mixture of Gaussian distribution

$$P(t|s) = \pi_1(s)N(f_1(s), \sigma^2) + \pi_2(s)N(f_2(s), \sigma^2) \qquad (11)$$

where $\pi_1(s)$ and $\pi_2(s) = 1 - \pi_1(s)$ are mixing proportions that depend on the intensity in the source image, and σ^2 represents the variance of the noise in the target image. Consistently with the functional case, we will restrict ourselves to polynomial intensity functions, i.e., $f_1(s) = \theta_0 + \theta_1 s + \theta_2 s^2 + \cdots + \theta_p s^p$, and $f_2(s) = \psi_0 + \psi_1 s + \psi_2 s^2 + \cdots + \psi_p s^p$.

An intuitive way to interpret this modeling is to state that for any voxel, there is a binary "selector" variable $\epsilon = \{1, 2\}$ that would tell us, if it was observed, which of the two functions f_1 or f_2 actually serves to map s to t. Without knowledge of ϵ, the best intensity correction to apply to S (in the minimum variance sense) is a weighted sum of the two functions

$$f(s,t) = P(\epsilon = 1|s,t)f_1(s) + P(\epsilon = 2|s,t)f_2(s) \qquad (12)$$

in which the weights correspond to the probability that the point be mapped according to either the first or the second function. Applying Bayes' law, we find that

$$P(\epsilon = 1|s,t) = \frac{P(\epsilon = 1|s)P(t|\epsilon = 1, s)}{P(t|s)}$$
$$P(\epsilon = 2|s,t) = \frac{P(\epsilon = 2|s)P(t|\epsilon = 2, s)}{P(t|s)} \qquad (13)$$

and thus, using the fact that $P(\epsilon|s) = \pi_\epsilon(s)$ and $P(t|\epsilon, s) = G_\sigma(t - f_\epsilon(s))$, the weights are determined by

$$P(\epsilon|s,t) = \frac{\pi_\epsilon(s)\, G_\sigma(t - f_\epsilon(s))}{\pi_1(s)\, G_\sigma(t - f_1(s)) + \pi_2(s)\, G_\sigma(t - f_2(s))} \qquad \epsilon = \{1, 2\}. \qquad (14)$$

In order to compute the intensity correction (12), we now need to identify the parameters of our model, i.e., the polynomial coefficients of f_1 and f_2, as well as the mixing proportions $\pi_1(s)$ and $\pi_2(s)$ and the variance σ^2. For this, we employ an *ad hoc* strategy that proceeds as follows.

First, $\boldsymbol{\theta}$ is estimated using the method described in Section II-B. The points not used to compute $\boldsymbol{\theta}$, in a number between zero and $N - c$, are used to estimate $\boldsymbol{\psi} = [\psi_0, \ldots, \psi_p]$ still using the same method. Note that if this number is less than $10 \times p$, p being the polynomial degree, monofunctional dependence is assumed and we fall back to the method described in the previous section.

This provides a natural estimation of the "selector" variable for each voxel: the n_1 points that were used to build f_1 are likely to correspond to $\epsilon = 1$, while the n_2 points used to build f_ψ are likely to correspond to $\epsilon = 2$. Finally, the points that are rejected while estimating $\boldsymbol{\psi}$ are considered as bad intensity matches. A natural estimator for the variance σ^2 is then

$$\hat{\sigma}^2 = \frac{n_1}{n_1 + n_2}\hat{\sigma}_1^2 + \frac{n_2}{n_1 + n_2}\hat{\sigma}_2^2 \qquad (15)$$

where $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are the variances found, respectively, for f_1 and f_2 during the RLS regression (see Section II-B). Similarly, the mixing proportions are computed according to

$$\hat{\pi}_\epsilon(s) = \frac{n_\epsilon(s)}{n_1(s) + n_2(s)}, \qquad \epsilon = \{1, 2\} \qquad (16)$$

in which $n_\epsilon(s)$ is the number of voxels having an intensity s and used to build the function f_ϵ. Notice that in the case where $n_1(s) = n_2(s) = 0$ (i.e., no voxel corresponding to the intensity s has been taken into account in the computation of f_1 or f_2), then we arbitrarily set the mixing proportions to $\hat{\pi}_1(s) = \hat{\pi}_2(s) = 0.5$.

The intensity correction of S can now be performed by reinjecting the estimated parameters in (14) and (12).

C. Combining the Intensity and Geometrical Transformations

As stated previously, the use of (1) in Section II-A to model geometrical transformations makes the assumption that the intensities of structures in S matches those in T. When dealing with images obtained from different modalities, this requirement is not met. To solve this problem, we presented in Section II-B a method to rectify intensity differences between corresponding structures in S and T using an underlying monofunctional or bifunctional model. The result of this intensity correction on S for the nth iteration will be denoted by $S_n^\star$ with $S_0^\star = S$. In the monofunctional case, $S_n^\star(x) = f(S(x))$ [see (3)] and in the bifunctional case $S_n^\star(x) = f(S(x), T(x))$ [see (12)]. Considering this and (1), the geometrical transformation is found using (17), shown at the bottom of the next page.

The reader might wonder why we restricted our technique to estimating at most two functions. In theory, our technique can well accommodate for more than two. The problem encountered regards finding proper values for c, i.e., the number of points used for the LTS computation of each function. As mentioned previously, c needs to be at least $(N+p+2)/2$ for the first polynomial. The number of points available to compute the second function are then at most $(N-p-2)/2$ and may well be less due

to the RLS regression. This means that at best, less than 25% of all points would be available to compute a third function, 12.5% for a fourth function, and so on. We expect that the intensity correspondences reflected by these third and fourth functions would not be of real value, especially during the first iterations of the registration process, where the number of outliers in C is at its highest.

D. Mutual Information

Another approach we tested to compute incremental displacements was inspired by the works of Viola *et al.* [1] and Maes *et al.* [2]. In these approaches, the rigid/affine registration between two images is formulated as a maximization of their MI. This choice is motivated by the fact that MI models the similarity between the images while resorting to assumptions that are much more general than functional dependence. Consequently, and contrary to the approach that was presented in Section II-B, they do not attempt to apply an intensity correction to one image so that it matches the other. Instead, they model the intensity dependence in a purely statistical fashion. Viola *et al.* [1] do this modeling using Parzen windowing, which enables them to differentiate the MI criterion with respect to the geometrical transformation.

As shown by Roche *et al.* [16] and Guimond *et al.* [17], our incremental displacements closely relate with a gradient descent on the sum of squared difference (SSD) criterion. Using this analogy, we implemented an alternative matching strategy where the incremental displacements given by (1) are replaced with the following formula:

$$v_{n+1}(x) = G_\sigma \otimes (v_n + \alpha \nabla MI(v_n)) \tag{18}$$

in which α is a positive constant and $\nabla MI(v_n)$ is the gradient of the MI criterion with respect to the nth displacement vector, whose mathematical expression is found in [1]. This updating corresponds to a first-order gradient descent on the MI criterion, of course up to the Gaussian filtering which is used as a regularization constraint.

Nonrigid registration results using mutual information as presented here are shown in Section IV-C.

III. Data

Most of the data used in the following experiments were obtained from BrainWeb [20], [21]. This tool uses an atlas [22] with a resolution of $1\times 1 \times 1$ mm^3 comprising nine segmented regions (see Table I) from which T1, T2, and PD images can be generated [23]. We use the discrete phantom atlas and three simulated MR images, one of each modality, with the same resolution as the atlas, 5% noise and no intensity nonuniformity. These three images are named $T1_a$, $T2_a$, and PD_a, respectively. Since they are generated from the same atlas, they represent the same underlying anatomy and are all perfectly matched.

TABLE I
Intensity Values Corresponding to the Different Tissue Types Present in the Discrete Phantom Atlas

Tissue type	Intensity	Tissue type	Intensity
Background	0	Muscle / Skin	140
CSF	28	Skin	168
Gray Matter	56	Skull	196
White Matter	84	Glial Matter	224
Fat	112	Connective	252

We also use a T1 MR image and a CT image, both from different subjects and having a resolution of $1\times 1 \times 1$ mm^3. Both these images were affinely registered with the atlas using the correlation ratio method [3]. The two images are called T1 and CT, respectively.

The images all respect the neurological convention, i.e., on axial slices, the patient's left is on the left side of the image.

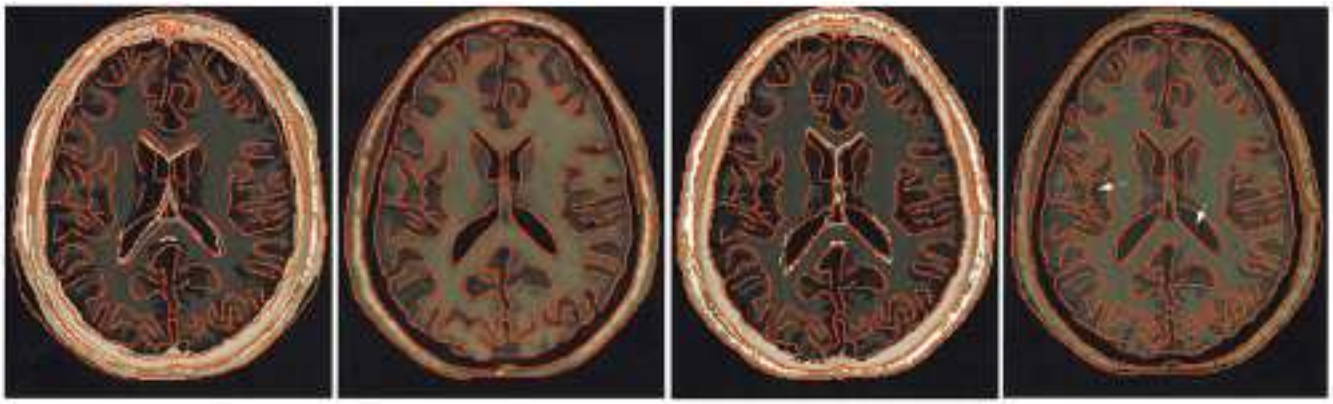
IV. Results and Discussion

In the following section, we present registration results involving images obtained from several different kinds of modalities. First, we show a typical example where monofunctional dependence can be assumed: registration of an atlas with an MR image. Then more practical examples are shown where images from different modalities are registered and where bifunctional dependence may be assumed. We also provide registration results obtained from our implementation of MI maximization as described in Section II-D.

The multilevel process was performed at three resolution levels, namely 4, 2 and 1mm/voxel. Displacement fields at one level are initialized from the result of the previous level. The initial displacement field v_0 is set to zero. 128 iterations are performed at 4 mm/voxel, 32 at 2 mm/voxel and 8 at 1 mm/voxel. These are twice the number of iterations used for registration of monomodal images using the conventional demons algorithm. We believe that making use of a better stopping criterion, such as the difference of the SSD values between iterations, would probably improve the results shown below but this aspect has not been investigated. It should be in future implementations. The Gaussian filter G_σ used to smooth the displacement field has a standard deviation of one voxel regardless of the resolution. This models stronger constraints on the deformation field at the beginning of the registration process to correct for gross displacements, and weaker constraints near the end when fine displacements are sought. The resampling process makes use of trilinear interpolation, except in the case of the atlas where nearest-neighbor interpolation is used.

Computation time to obtain the following results is around 60 min on a 450-MHz PC with 500 MB of RAM (10 min at 4 mm, 20 min at 2 mm and 30 min at 1 mm). Most of this time ($\approx$ 85%) is devoted to the intensity correction part, which has not been optimized in this first version of our program. The other

$$v_{n+1}(x) = G_\sigma \otimes \left(v_n + \frac{S_n^\star \circ h_n(x) - T(x)}{||\nabla (S_n^\star \circ h_n)(x)||^2 + [S_n^\star \circ h_n(x) - T(x)]^2} \nabla (S_n^\star \circ h_n)(x) \right) \tag{17}$$



(a) (b) (c) (d)

Fig. 1. Corresponding axial slices of the atlas to T1 registration result. Contours were obtained using a Canny–Deriche edge detector on T1 and overlaid on the other images to better assess the quality of registration. Arrows point to examples of bad intensity correction. (a) Atlas. (b) T1. (c) Atlas without intensity correction after registration with T1. (d) Atlas with intensity correction after registration with T1.

(a) (b) (c) (d)

Fig. 2. Corresponding axial slices of the atlas to T1′ registration result. Contours were obtained using a Canny–Deriche edge detector on T1′ and overlaid on the other images to better assess the quality of registration. (a) Atlas. (b) T1′. (c) Atlas without intensity correction after registration with T1′. (d) Atlas with intensity correction after registration with T1′.

15% is taken by the standard registration code which is stable and well optimized.

A. *Monofunctional Dependence*

We present here the result of registering the atlas with T1. Since the atlas can be used to generate realistic MR images, it is safe to assume a functional dependence between the intensity of the atlas and those of T1. Also, since T1 and the atlas are well aligned due to the affine registration, we have roughly estimated that the number of points already well matched, i.e., the number of good points in C, are at least $0.80 \times N$, to which we have set the value of c. Since ten classes are present in the atlas, the polynomial degree chosen was set to nine. Having a polynomial of this degree will accommodate any permutation of the structure intensities in the atlas. This is true because it is always possible to fit a polynomial of degree nine to match ten points exactly.

The result of registration is presented in Fig. 1. Fig. 1(a) shows one slice of the atlas. Fig. 1(b) is the corresponding slice of T1. Fig. 1(c) and (d) presents the result of registering the atlas with T1 using our algorithm. Fig. 1(c) shows the result without the intensity transformation; we have simply applied to the atlas the geometrical transformation resulting from the registration procedure. Fig. 1(d) shows the image resulting from the registration process. It has the same shape as the resampled atlas and intensities have been transformed using the intensity correction.

As can be seen, there is one obvious problem with this result: although the shape of the atlas seems well corrected, the CSF intensity is not, as can be seen in the ventricles and around the cortex [see arrows in Fig. 1(d)]. This reflects on the matching, as can be observed around the anterior aspect of the lateral ventricles on Fig. 1(c) and (d).

This problem can also be observed by looking at the intensity transformation function presented in Fig. 3 (Intensity values corresponding to the different tissue types present in the atlas are shown in Table I.) This is due to a spatial overlap of the CSF in the atlas and the gray and white matter in T1, especially around the cortical area which is known to present large variations between subjects. We believe this is due to the strong smoothness constraints imposed by the Gaussian regularization which may prevent the assessment of large and uneven displacements required to match the cortex.

To verify this assumption, we registered $T1_a$ with T1 and deformed $T1_a$ using the resulting displacement field. This provided an image T1′– that closely resembles T1 [Compare Figs. 1(b) and 2(b)] and for which we are assured that the shape

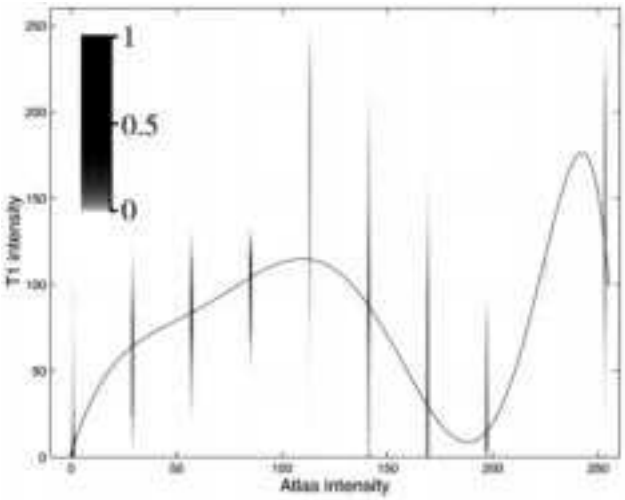


Fig. 3. Intensity transformation found by registering the atlas with T1 and assuming monofunctional dependence. The function f is overlaid on the joint histogram of the two images after registration. The joint histogram values have been compressed logarithmically and normalized as is depicted in the color scale. The tissue types corresponding to the different atlas intensities are presented in Table I.

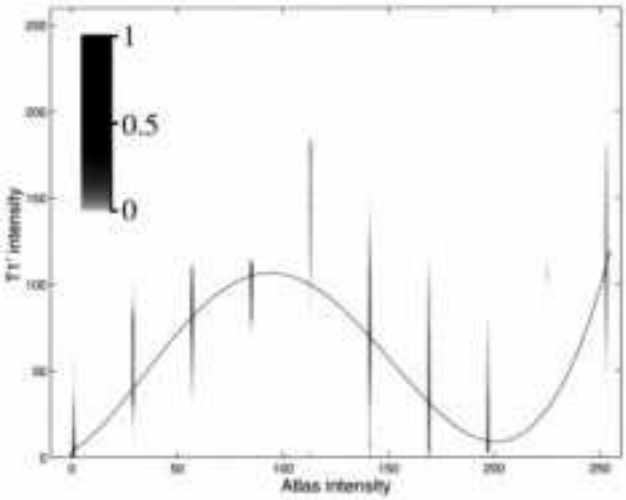


Fig. 4. Intensity transformation found by registering the atlas with T1′ and assuming monofunctional dependence. The function f is overlaid on the joint histogram of the two images after registration. The joint histogram values have been compressed logarithmically and normalized as is depicted in the color scale. The tissue types corresponding to the different atlas intensities are presented in Table I.

variation from the atlas can be assessed by our algorithm. We then registered the atlas with T1′. This result is presented in Fig. 2. As can be observed, the CSF intensity value is now well corrected. By looking at the intensity transformation shown in Fig. 4, we also notice that each atlas structure has corresponding intensity ranges in T1 that are less extended than those of Fig. 3, reflecting a better match between the images. This finding puts forth that the displacement field regularization has to be able to accommodate the large uneven displacement of the cortex. To cope with large displacements, Gaussian filtering may probably be replaced with another regularization strategy such as that based on a fluid model [24] or on a nonquadratic potential energy [25].

Another difference between Figs. 3 and 4 is the difference in the intensity mapping of the glial matter (Intensity 224 in the atlas). This structure forms a small envelope around the ventricles and is represented by so few points that they are considered as outliers. This could be corrected by considering more points during the intensity transformation. In fact, c should be increased at each iteration to reflect that, during registration, S gradually aligns with T and more points in C can be considered as well matched.

B. Bifunctional Dependence

When registering images from different modalities, monofunctional dependence may not necessarily be assumed. Here, we applied the method described in Section II-B where two polynomial functions of degree 12 are estimated. This number was set arbitrarily to a relatively high value to enable important intensity transformations.

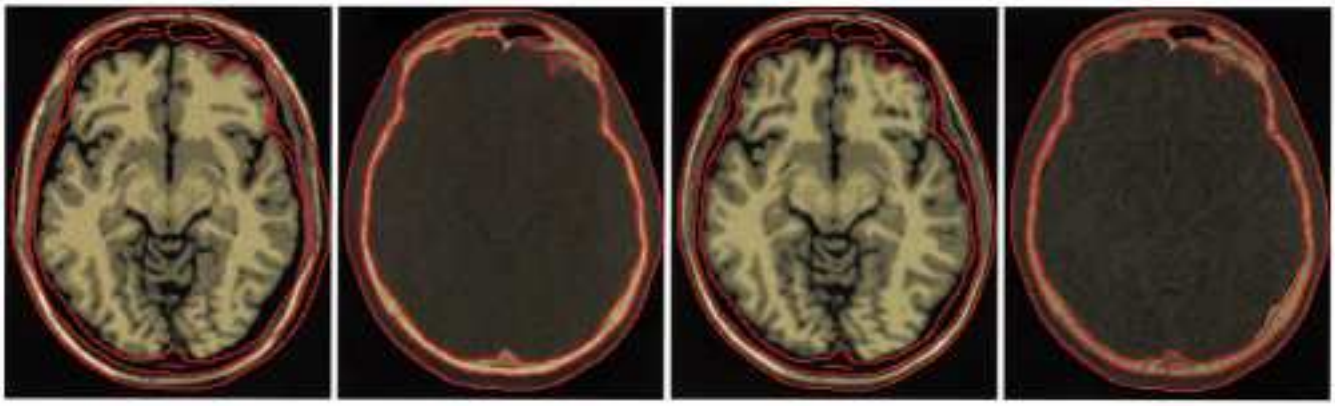
Fig. 5 presents the result of registering $T1_a$ with CT. Using these last two modalities, most intensities of $T1_a$ should be mapped to gray and only the skull, representing a small portion of the image data, should be mapped to white. After affine registration almost all voxels are well matched, i.e., the number of good points in C is large. Hence, in this particular case, we have chosen a high value for c set to $0.90 \times N$.

As we can see in Fig. 5, the skull, shown in black in the MR image and in white in the CT scan, is well registered and the intensity transformation adequate. Fig. 6(a) presents the joint histogram of the two images after registration. This histogram is color-coded and ranges from red representing high point densities to blue depicting low point densities. Fig. 6(b) presents the functions f_1 and f_2 found during the registration process. The red line corresponds to f_1 and the blue one to f_2. The line width for a given intensity s is proportional to the value of the corresponding $\pi_\epsilon(s)$. The gray values represent the joint histogram after registration.

As can be observed in Fig. 6(b), the polynomials found fit well with the high density clusters of the joint histogram. Still, some points need to be addressed.

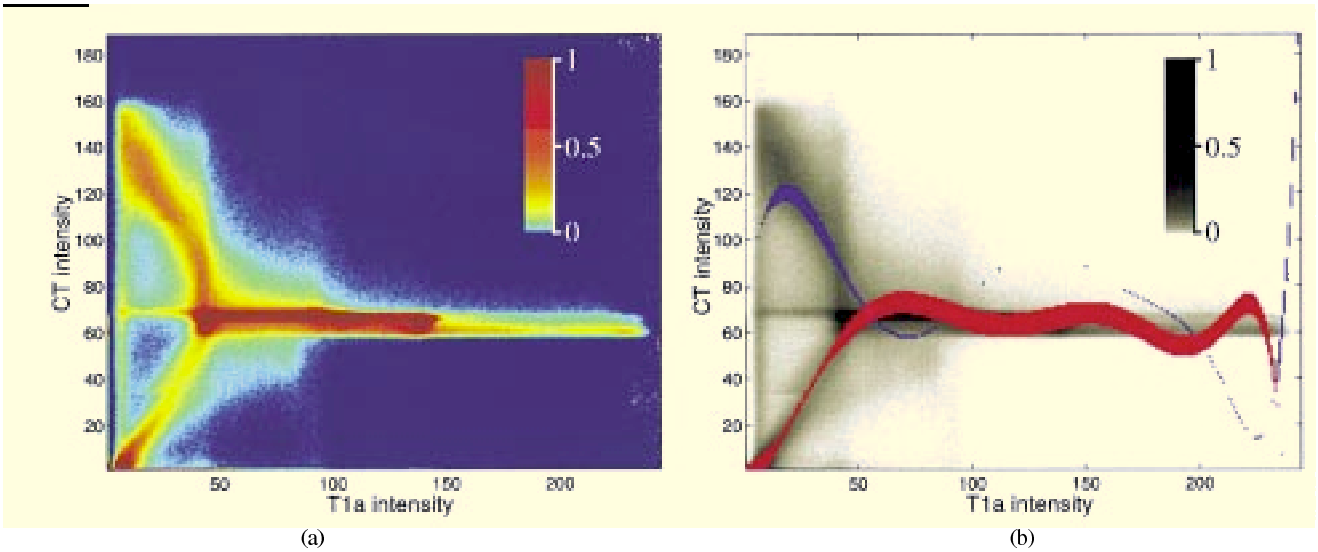
We can observe that due to the restricted polynomial degree, f_1, (shown in red) oscillates around the CT gray value instead of fitting a straight line. This is reflected in the intensity corrected image, shown in Fig. 5(d), where the underlying anatomy can still be observed by small intensity variations inside the skull. This artifact has insubstantial consequences during the registration process since the difference between most of the voxels of T and S is zero, resulting in null displacements. The displacements driving the deformation will be those of the skull and the skin contours, and will be propagated in the rest of the image with the Gaussian filtering of the displacement field.

We also notice that f_2 (shown in blue), which is mainly responsible for the mapping of the skull, does not properly model the cluster it represents for intensities smaller than five. The mapping for these intensities is slightly underestimated. This may have two causes. First, as in the previous case, it might be due to the restricted polynomial degree. Second, we can notice that some of the background values in $T1_a$ that have an intensity close to zero are mapped to gray values in the CT which correspond to soft tissues. This means that some of the background in $T1_a$ is matched with the skin in the CT. This has the effect of "pulling" f_2 closer to the small



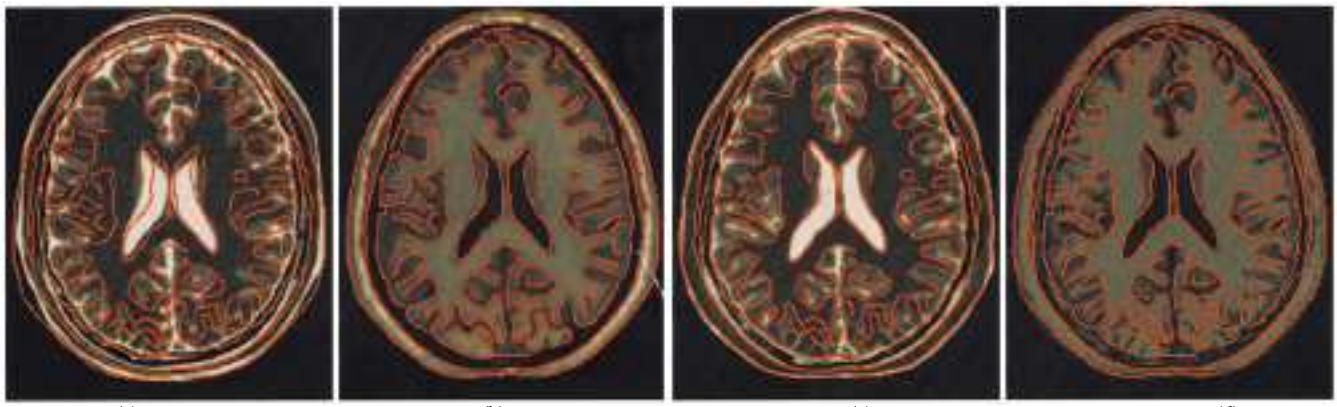
(a) (b) (c) (d)

Fig. 5. Corresponding axial slices of $T1_a$ to CT registration result. Contours were obtained using a Canny–Deriche edge detector on CT (Fig. 1) and overlaid on the other images to better assess the quality of registration. (a) $T1_a$. (b) CT. (c) $T1_a$ without intensity correction after registration with CT. (d) $T1_a$ with intensity correction after registration with CT



(a) (b)

Fig. 6. Intensity transformation found when registering $T1_a$ with CT and assuming bifunctional dependence. (a) The joint histogram values have been compressed logarithmically and normalized as is depicted in the color scale. Values range from red representing high point densities to blue depicting low point densities. (b) The red line corresponds to f_1 and the blue one to f_2. The line width for a given intensity value s in $T1_a$ corresponds to the value of the corresponding $\pi_c(s)$. The gray values represent the joint histogram after registration.



(a) (b) (c) (d)

Fig. 7. $T2_a$ without intensity correction after registration with T1. $T2_a$ with intensity correction after registration with T1. Corresponding axial slices of $T2_a$ to T1 registration result. Contours were obtained using a Canny–Deriche edge detector on T1 (b)and overlaid on the other images to better assess the quality of registration. (a) $T2_a$. (b) T1. (c) $T2_a$ without intensity correction after registration with T1. (d) $T2_a$ with intensity correction after registration with T1.

cluster positioned around (2,65). If the underestimation of f_2 arises because of the second reason, letting the algorithm iterate longer might provide a better result.

In Figs. 7 and 9 we present the result of registering $T2_a$ and PD_a, respectively, with T1. Figs. 8 and 10 show the corresponding intensity transformations. For these ex-

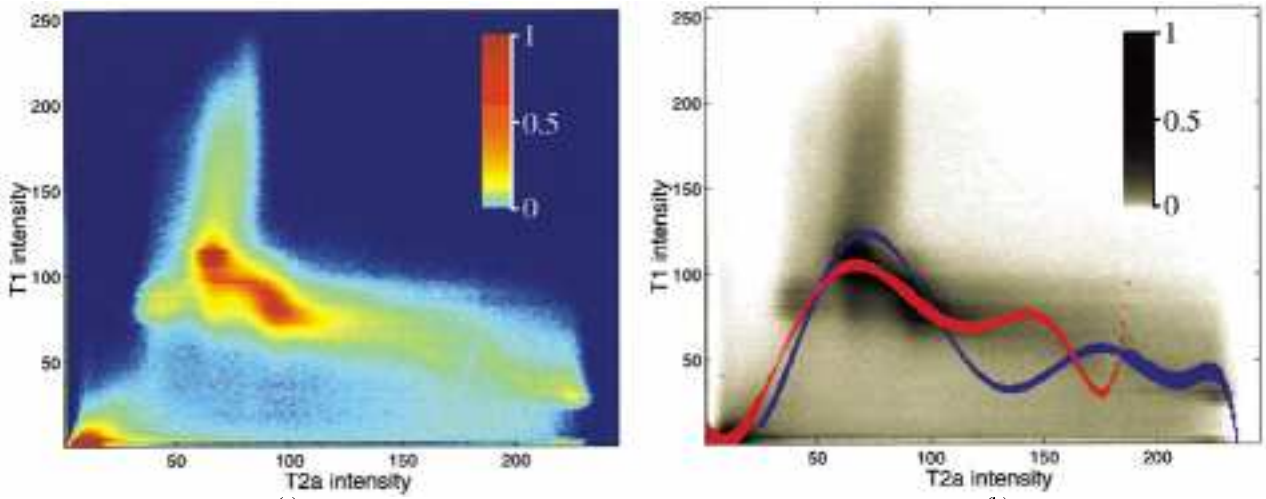


Fig. 8. Intensity transformation found when registering $T2_a$ with T1 and assuming bifunctional dependence. (a) The joint histogram of $T2_a$ and T1 after registration. The joint histogram values have been compressed logarithmically and normalized as is depicted in the color scale. Values range from red representing high point densities to blue depicting low point densities. (b) The red line corresponds to f_1 and the blue one to f_2. The line width for a given intensity value s in $T2_a$ corresponds to the value of the corresponding $\pi_c(s)$. The gray values represent the joint histogram after registration.

Fig. 9. Corresponding axial slices of PD_a to T1 registration result. Contours were obtained using a Canny–Deriche edge detector on T1 (b) and overlaid on the other images to better assess the quality of registration. (a) PD_a. (b) T1. (c) PD_a without intensity correction after registration with T1. (d) PD_a with intensity correction after registration with T1.

periments, c was set to $0.60 \times N$, a value we have found to be effective for these types of modalities after affine registration.

One observation that can be made by looking at the joint histograms of Figs. 8(a) and 10(a) is that there seems to be a functional dependence between the images intensities, i.e., there is a function that can go though the major clusters of the joint histograms. This is also reflected by the closely similar shapes of the corresponding f_1 and f_2.

Thus, we have registered PD_a with T1 using the monofunctional model and $c = 0.8$. This result is presented in Fig. 11 and the corresponding intensity transformation in Fig. 13. As can be seen by comparing the Figs. 9(d) and 11(d), the intensity transformation does not correct as well the CSF intensities and the distinction between the different structures is less contrasted.

This may be explained by a closer look at our bifunctional intensity modeling. Equation (11) reflects the assumption that if an anatomical point has an intensity s in S, the corresponding point has an intensity t in T that is distributed normally around two possible values depending on s. But it makes no assumption about how the intensities in S are distributed. This models the intensities of S without noise, which may not necessarily be well justified, but enables the use of linear regression to estimate the intensity transformation.

The effect of noise in S is reflected in the joint histograms by enlarging clusters along the x axis. This, added to bad matches and partial volume effect, creates many outliers in C and makes the assessment of the true intensity transformation more difficult and more resistant to our robust regression technique. Preprocessing of S using for example anisotropic diffusion may narrow the clusters and provide better results [10].

Adding the estimation of a second function in the bifunctional model helps counter the effect of noise on S. For example, the CSF in the PD_a image has intensity values ranging from about 200 to 240 and gray matter from about 175 to 210. In T1, these ranges are about 30 to 70 and 55 to 80, respectively. As can be seen in Fig. 10, f_1 models well the gray matter cluster but fails to reflect the CSF transformation. This is also well depicted in Fig. 11 in which the CSF and gray matter intensity transformation is modeled using a single polynomial. In this

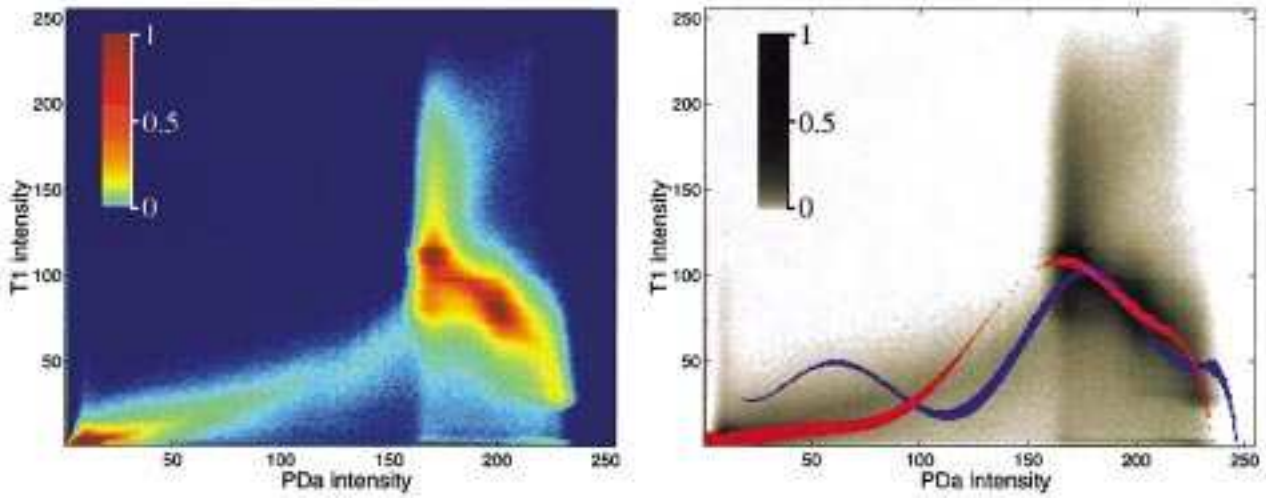


Fig. 10. Intensity transformation found when registering PD_a with T1 and assuming bifunctional dependence. (a) The joint histogram of PD_a and T1 after registration. The joint histogram values have been compressed logarithmically and normalized as is depicted in the color scale. Values range from red representing high point densities to blue depicting low point densities. (b) The red line corresponds to f_1 and the blue one to f_2. The line width for a given intensity value s in PD_a corresponds to the value of the corresponding $\pi_c(s)$. The gray values represent the joint histogram after registration.

Fig. 11. Corresponding axial slices of PD_a to T1 registration result using the monofunctional model. Contours were obtained using a Canny–Deriche edge detector on T1 (b) and overlaid on the other images to better assess the quality of regidtration. (a) PD_a. (b) T1. (c) PD_a without intensity correction after registration with T1. (d) PD_a with intensity correction after registration with T1.

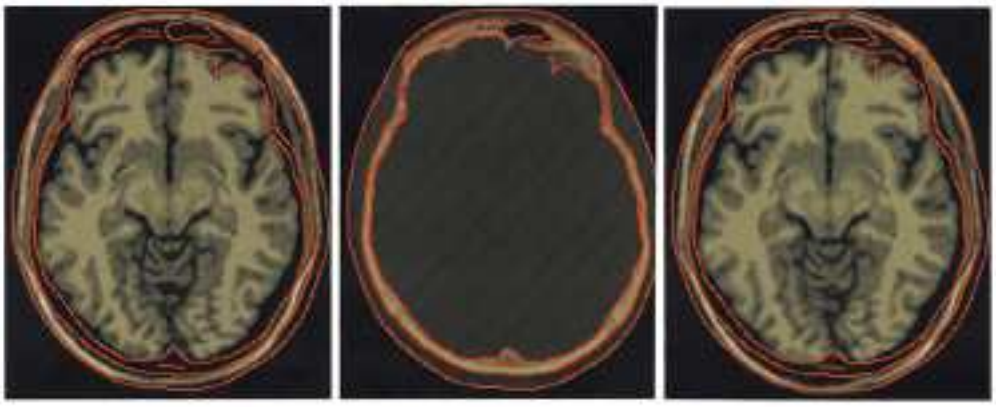


Fig. 12. Corresponding axial slices $T1_a$ to CT registration result using MI. Contours were obtained using a Canny–Deriche edge detector on CT (b) and overlaid on the other images to better assess the quality of registration. (a) $T1_a$. (b) CT. (c) $T1_a$ after registration with CT.

case, the CSF is often mapped as gray matter. Estimating the second polynomial f_2 solves this problem by considering the CSF cluster.

Note that for the registration of the atlas to T1 and of $T1_a$ to CT, we have always deformed the image with the most information, i.e., the one with the higher number of perceivable

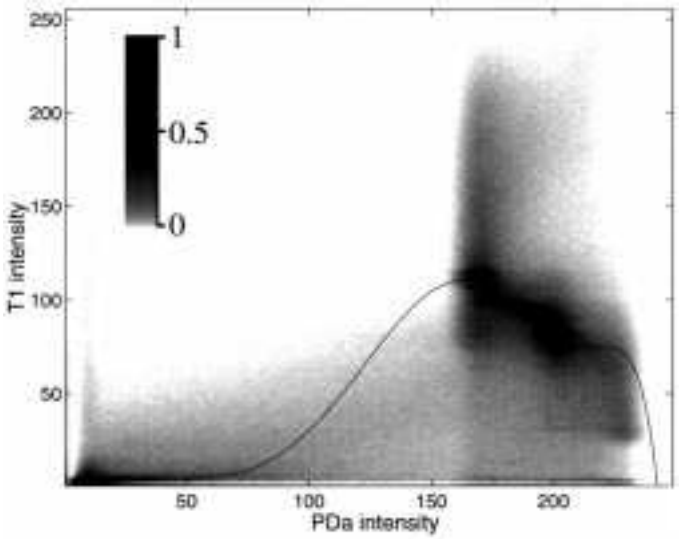


Fig. 13. Intensity transformation found by registering PD_a with T1 and assuming monofunctional dependence. The function f is overlaid on the joint histogram of the two images after registration. The joint histogram values have been compressed logarithmically and normalized as is depicted in the color scale.

structures. This is simply because our algorithm permits many structures of the deformed image to be mapped to a single intensity, as is the case when transforming a T1 image into a CT image (see Fig. 5). But a single intensity in the deformed image can be mapped to at most two intensities in the target image. For example, if we used the CT image as the image to be deformed, the dominant gray intensity value in this image would have to be mapped to gray matter, white matter, CSF, etc. This would require more than two functions to be estimated and complicates the algorithm. Hence, it is always better to use the image with the most number of structures visible as the source image.

C. Mutual Information

We present in Fig. 12 the result of registering $T1_a$ with CT using the MI method described in Section II-D. A typical difference between using our bifunctional method instead of our MI implementation can be appreciated by comparing Figs. 5(c) and 12(c). As can be seen, the contours of corresponding structures do not match after registration using MI. The head contours seem to be attracted by the image borders, which means that the driving forces have misleading directions in this region. This outcome might be due to the fact that Parzen windowing provides too little information on how to match the intensities of the images. As a consequence, the direction of the MI gradient, from which the local driving force is derived, might be unreliable. Many registrations were performed using the MI criterion with varying values for the step length α and several Parzen window sizes. The results we obtained using this strategy were much less convincing than the ones we obtained using our bifunctional method; the source image deformed very little unless the α parameter was set to a high value, in which case the displacements looked random.

The poor results obtained with our implementation of MI might sound surprising as MI performs generally very well in rigid/affine registration. Although it would be irrelevant to conclude that MI is not well-suited for elastic registration, a partial explanation may be given. What MI does is to measure the similarity of the images from the joint probability density function (pdf) of their intensities. In most of the MI-based approaches, the joint pdf is considered as being the image (normalized) joint histogram [2]. There is actually a hidden estimation problem when using MI, as already pointed out by [1]. It can be shown that the joint histogram is the maximum likelihood estimate of the joint pdf among the set of every possible discrete pdf's [10]. This set has a dimension $k_s \times k_t - 1$, where k_s and k_t are the numbers of intensity levels, respectively, in the source and target images.

In the context of rigid/affine registration, where no more than 12 geometrical parameters need to be estimated, such an enormous search space is usually affordable. However, with elastic registration, the geometrical transformation is granted many degrees of freedom and maximizing MI might then become an under-constrained problem. The use of Parzen windowing may be viewed as a regularization strategy as it imposes constraints to the intensity space. While Parzen windowing is a nonparametric approach, our method explicitly restricts the intensity space using polynomial functions. In the case where monofunctional dependence is assumed, only $p + 1$ parameters are estimated to model the intensity dependence, p being the polynomial degree. When assuming a bifunctional relationship, this number becomes $2(p + 1) + 1 + 2k_s$.

TABLE II
STATISTICS REGARDING THE DIFFERENCES BETWEEN DISPLACEMENTS PROVIDED BY EACH TYPE OF REGISTRATION. EACH CELL PRESENTS THE MEDIAN LENGTH, THE AVERAGE LENGTH WITH THE CORRESPONDING STANDARD DEVIATION AND THE MAXIMUM LENGTH. ALL MEASURES ARE IN MILLIMETERS

	Atlas-T1 Atlas-T1'	Atlas-T1 $T1_a$-T1	Atlas-T1 $T2_a$-T1	Atlas-T1 PD_a-T1	$T2_a$-T1 PD_a-T1
median	0.85	1.46	1.13	1.67	1.32
average	0.97	1.58	1.23	1.76	1.40
std. dev.	0.60	0.84	0.63	0.79	0.68
maximum	4.57	6.99	5.14	7.10	6.86
	Atlas-T1' $T1_a$-T1	Atlas-T1' $T2_a$-T1	Atlas-T1' PD_a-T1	$T1_a$-T1 $T2_a$-T1	$T1_a$-T1 PD_a-T1
median	0.89	0.95	1.15	1.00	1.01
average	1.00	1.07	1.25	1.18	1.16
std. dev.	0.55	0.60	0.64	0.78	0.71
maximum	4.48	5.46	7.43	7.17	8.08

D. Displacement Field Comparison

Since the atlas, $T1_a$, $T2_a$, and PD_a have all been registered with T1, it is relevant to compare some statistics of the resulting displacement fields to assess if our algorithm provides consistent results across modalities.

We computed statistics regarding the difference between any two of these displacement fields. The length of the vectors of the resulting difference fields were calculated. Each cell of Table II presents, for each combination of displacement fields, the median length, the average length with the corresponding standard deviation and the maximum length of the difference field.

The two largest average differences are 1.76 and 1.58 mm and were found when registering the atlas with $T1_a$ and PD_a, respectively. This may be explained by the intensity correction

bias for the CSF that would tend to attenuate displacements and produce larger errors, a problem invoked in Section IV-A. Aside from these, the average error length varies between 0.97 and 1.40 mm and the median error is between 0.85 and 1.32 mm, with the largest errors occurring in the cortex area. These are values in the range of the image resolution of 1.0 mm. Note also that all the standard deviations are below this value.

We point out that these are global measures that are presented to provide an idea of the differences between the displacement fields. They do not strictly provide a validation of the method, but do show a certain coherence between the different results we obtained.

V. Conclusion

In this paper, we introduced an original method to perform nonrigid registration of multimodal images. This iterative algorithm is composed of two steps: the geometrical transformation and the intensity transformation. Two intensity transformation models were described which assume either monofunctional or bifunctional dependence between the images to match. Both of these models are built using robust estimators to enable precise and accurate transformation solutions. Results of registration were presented and showed that the algorithm performs very well for several kinds of modalities including T1, T2, PD, CT, and segmentations, and provides consistent results across modalities. Our algorithm was compared with the maximization of the MI criterion and seems to be more apt at evaluating high-dimensional deformations. Our explanation is that it puts more constraints on the intensity transformation that relates the images and, thus, permits a better search of the parameter space.

References

[1] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.

[2] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, pp. 187–198, Apr. 1997.

[3] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," in *Lecture Notes in Computer Science*, vol. 1496, Proc. 1st Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI'98), W. M. Wells, A. Colchester, and S. Delp, Eds., pp. 1115–1124.

[4] J. B. A. Maintz, E. H. W. Meijering, and M. A. Viergever, "General multimodal elastic registration based on mutual information," in *SPIE Proc.*, vol. 3338, Medical Imaging 1998: Image Processing (MI'98), K. M. Hanson, Ed., San Diego, CA, pp. 144–154.

[5] T. Gaens, F. Maes, D. Vandermeulen, and P. Suetens, "Non-rigid multimodal image registration using mutual information," in *Lecture Notes in Computer Science*, vol. 1496, Proc. 1st Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI'98), W. M. Wells, A. Colchester, and S. Delp, Eds., pp. 1099–1106.

[6] Y. H. Lau, M. Braun, and B. F. Hutton, "Non-rigid 3D image registration using regionally constrained matching and the correlation ratio," in *Proc. Int. Workshop Biomedical Image Registration (WBIR'99)*, F. Pernus, S. Kovačič, H. S. Stiehl, and M. A. Viergever, Eds., Bled, Slovenia, pp. 137–148.

[7] J.-P. Thirion. (1995) Fast nonrigid matching of 3D medical images. Institut National de Recherche en Informatique et en Automatique, Sophia-Antipolis, France. [Online]. Available: http://www.inria.fr/RRRT/RR-2547.html

[8] ——, "Image matching as a diffusion process: An analogy with Maxwell's demons," *Med. Image Anal.*, vol. 2, no. 3, pp. 243–260, 1998.

[9] R. P. Woods, J. C. Mazziotta, and S. R. Cherry, "MRI-PET registration with automated algorithm," *J. Comput. Assist. Tomogr.*, no. 4, pp. 536–546, 1993.

[10] A. Roche, G. Malandain, N. Ayache, and S. Prima, "Toward a better comprehension of similarity measures used in medical image registration," in *Proc. 2nd Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI'99)*, vol. 1679, Oct. 1999, pp. 555–566.

[11] A. Roche, G. Malandain, and N. Ayache, "Unifying maximum likelihood approaches in medical image registration," *Int. J. Imag. Syst. Technol.*, vol. 11, pp. 71–80, 2000.

[12] J. Feldmar, J. Declerck, G. Malandain, and N. Ayache, "Extension of the ICP algorithm to nonrigid intensity-based registration of 3D volumes," *Comput. Vis. Image Understanding*, vol. 66, no. 2, pp. 193–206, May 1997.

[13] D. C. Barber, "Registration of low resolution medical images," *Phys. Med. Biol.*, vol. 37, no. 7, pp. 1485–1498, 1992.

[14] 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," in *NeuroImage, Proc. 2nd Int. Conf. Functional Mapping of the Human Brain (HBM'95)*, pp. 165–189.

[15] A. W. Toga, *Brain Warping*. New York: Academic, 1999.

[16] A. Roche, A. Guimond, N. Ayache, and J. Meunier, "Multimodal elastic matching of brain images," presented at the 6th Eur. Conf. Computer Vision (ECCV'00), Dublin, Ireland, June 26–July 1, 2000.

[17] A. Guimond, A. Roche, A. Ayache, and J. Meunier. (1999, Nov.) Multimodal brain warping using the demons algorithm and adaptative intensity corrections. Institut National de Recherche en Informatique et en Automatique, Sophia Antipolis, France. [Online]. Available: http://www.inria.fr/RRRT/RR-3796.html

[18] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, ser. Probability and mathematical statistics. New York: Wiley, 1987.

[19] P. J. Rousseeuw and K. Van Driessen. (1999) Computing LTS regression for large data sets. Tech. Rep. Statistics Group, Univ. Antwerp, Antwerp, Belgium. [Online]. Available: http://win-www.uia.ac.be/u/statis/publicat/fastlts_readme.html

[20] Simulated brain database [Online]. Available: http://www.bic.mni.mcgill.ca/brainweb/

[21] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, and A. C. Evans, "Brainweb: Online interface to a 3D MRI simulated brain database," in *NeuroImage, Proc. 3rd Int. Conf. Functional Mapping of the Human Brain (HBM'97)*, vol. 5, Copenhagen, , Denmark, p. S425.

[22] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Med. Imag.*, vol. 17, pp. 463–468, June 1998.

[23] R. K.-S. Kwan, A. C. Evans, and G. B. Pike, "An extensible MRI simulator for post-processing evaluation," in *Lecture Notes in Computer Science*, vol. 1131, Proc. 4th Int. Conf. Visualization in Biomedical Computing (VBC'96), K. H. Höhne and R. Kikinis, Eds., Hamburg, , Germany, pp. 135–140.

[24] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "Deformable templates using large deformation kinematics," *IEEE Trans. Med. Imag.*, vol. 5, pp. 1435–1447, Oct. 1996.

[25] P. Hellier, C. Barillot, E. Mémin, and P. Pérez, "Medical image registration with robust multigrid techniques," in *Lecture Notes in Computer Science*, vol. 1679, Proc. 2nd Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI'99), Cambridge, U.K., pp. 680–687.