

Non-rigid Atlas-to-Image Registration by Minimization of Class-Conditional Image Entropy

Emiliano D’Agostino, Frederik Maes*, Dirk Vandermeulen, and Paul Suetens

Katholieke Universiteit Leuven, Faculties of Medicine and Engineering, Medical Image Computing (Radiology - ESAT/PSI), University Hospital Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium
`Emiliano.DAgostino@uz.kuleuven.ac.be`

Abstract. We propose a new similarity measure for atlas-to-image matching in the context of atlas-driven intensity-based tissue classification of MR brain images. The new measure directly matches probabilistic tissue class labels to study image intensities, without need for an atlas MR template. Non-rigid warping of the atlas to the study image is achieved by free-form deformation using a viscous fluid regularizer such that mutual information (MI) between atlas class labels and study image intensities is maximized. The new registration measure is compared with the standard approach of maximization of MI between atlas and study images intensities. Our results show that the proposed registration scheme indeed improves segmentation quality, in the sense that the segmentations obtained using the atlas warped with the proposed non-rigid registration scheme better explain the study image data than the segmentations obtained with the atlas warped using standard intensity-based MI.

1 Introduction

An important problem in medical image analysis is the accurate and reliable extraction of brain tissue voxel maps for white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF) from (possibly multispectral) MR brain images. Typical applications include the visualization of the cortex by volume or surface rendering, quantification of cortical thickness, quantification of intra-subject morphological changes over time and quantification of inter-subject morphological differences in relation to certain neurological or other conditions. While segmentation procedures that require some manual intervention (e.g. for initialization or supervised training) inevitably have to deal with some inter- and intra-observer variability and subjectivity, fully automated procedures for intensity-based tissue classification are more objective and potentially more robust and allow for efficient and consistent off-line processing of large series of data.

* Frederik Maes is Postdoctoral Fellow of the Fund for Scientific Research - Flanders (FWO-Vlaanderen, Belgium).

In previous work, we introduced a model-based approach for automated intensity-based tissue classification of MR brain images [4]. This method uses the Expectation-Maximization (EM) algorithm to iteratively estimate the parameters θ of a Gaussian mixture model (assuming the intensities of each tissue class to be normally distributed with unknown mean and variance, but corrupted by a spatially varying intensity inhomogeneity or bias field) and simultaneously classify each voxel accordingly, such as to maximize the likelihood $p(I|\theta)$ of the intensity data I given the model. The method is initialized by providing initial tissue classification maps for WM, GM, CSF and OTHER derived from a digital brain atlas after appropriate spatial normalization of the atlas to the study images. However, the atlas is not only used to initialize the EM procedure, but also serves as a spatially varying prior that constrains the classification during parameter estimation and in the final classification step. The probabilistic tissue classification L is obtained as the a posteriori probability of tissue label given the observed image intensity and the estimated intensity parameters, which, assuming that all voxels are independent, is computed using Bayes' rule as $p(L_k = j|I_k, \theta) \propto p(I_k|L_k = j, \theta).p(L_k = j)$ with L_k the label assigned to voxel k , j the various tissue classes, I_k the intensity of voxel k , $p(I_k|L_k = j, \theta)$ the probability of the observed intensity given the specified class label as derived from the Gaussian mixture model and $p(L_k = j)$ the prior probability of voxel k to belong to class j , which is simply the atlas registered to the image I . Hence, the quality of the atlas-to-image registration has a direct impact on the segmentation result through the above relation and the impact of the atlas model ($p(L_k = j)$) is as important as that of the intensity data ($p(I_k|L_k = j, \theta)$) itself.

In the method described in [4], affine registration was used to align the brain tissue distribution maps provided with SPM [2] with the study image by maximization of mutual information of corresponding voxel intensities [3] of the study image and the MR template of the SPM atlas. But while affine registration provides an atlas-to-image registration that is globally satisfactory in most cases, it fails to compensate for local morphological differences between atlas and study images, for instance at the cortical surface or at the ventricular boundaries, especially in study images showing enlarged ventricles. Non-rigid image registration, using an appropriate similarity metric and regularization criterion, allows to locally adapt the morphology of the atlas to that of the image under study, such that a more specific prior model $p(L_k = j), \forall j$ is obtained, resulting in a more accurate segmentation that is more consistent with the image data. In this paper we derive a new similarity measure that matches the atlas class labels directly to the study image intensities, such that the likelihood of the data given the spatially deformed prior model is maximized. We show that the proposed scheme for atlas registration results in more optimal tissue segmentations that better explain the data than is the case with non-rigid atlas registration based on matching intensity images instead of class labels.

2 Method

2.1 Similarity Measure

Our aim is to apply non-rigid registration using an appropriate similarity measure to optimally align the prior tissue distribution maps of the atlas to the image under study, such that the a priori classification provided by the atlas best fits the image intensities. This is not necessarily guaranteed by matching the atlas MR template to the study image using an intensity-based similarity metric such as mutual information of corresponding voxels intensities, as is typically done [4]. Instead, what is needed is a measure that directly evaluates the correspondence between atlas tissue probabilities and study image intensities without having to rely on atlas intensity images. In this context, we propose the new information-theoretic similarity measure $I(Y, L)$ for voxel-based non-rigid image matching that measures the mutual information between study image intensities Y and atlas class label probabilities L . $I(Y, L)$ is defined as

$$I(Y, L) = \sum_k \sum_y p(k, y) \cdot \log \frac{p(k, y)}{p(k) \cdot p(y)} \quad (1)$$

with k indexing the different atlas class labels (WM, GM, CSF and OTHER), y the image intensity in the study image, $p(k, y)$ the joint probability distribution of class k and intensity y and $p(k)$ and $p(y)$ the corresponding marginal distributions for class label k and intensity y respectively. This measure is analogous to the traditional mutual information of voxel intensities similarity measure [3], but differs in the features used to estimate similarity and in the way the joint probability distribution $p(k, y)$ is computed. Samples i with intensity y_i at positions p_i in the image Y are transformed into the corresponding positions q_i in the atlas space using the transformation T_α with parameters α . The joint probability distribution $p(k, y)$ is constructed using partial volume distribution (PV) interpolation [3]:

$$p(k, y) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^8 w_{i,j} \delta(y - y_i) \cdot c_{i,j}(k) \quad (2)$$

with N the number of samples in the image, j indexing each of the 8 nearest neighbours on the grid of the atlas images of the transformed location q_i of sample i , $w_{i,j}$ the trilinear interpolation weight associated with neighbour j and $c_{i,j}(k)$ the probability for tissue class k at this grid point as given by the atlas. The marginal distributions $p(y)$ and $p(k)$ are derived by integration of $p(k, y)$ over k and over y respectively. $p(k, y)$ is a continuous and a.e. differentiable function of the registration parameters α through the weights $w_{i,j}$. Further on, we derive an expression for the derivative of $I(Y, L)$ with respect to local displacements of individual voxels to define a force field that drives the registration in order to maximize $I(Y, L)$.

$I(Y, L)$ can be interpreted as $I(Y, L) = H(Y) - H(Y|L)$ with $H(Y)$ the entropy of the image intensity data and $H(Y|L)$ the residual entropy (or uncertainty) of Y given the atlas classification L . Because $H(Y)$ is constant during registration (all samples of Y contribute always to the joint probability distribution $p(k, y)$), maximization of $I(Y, L)$ is equivalent to minimizing the class-conditional image intensity entropy $H(Y|L)$. The proposed method thus aims at aligning the atlas labels such that the atlas best explains the data, i.e. such that the intensities corresponding to each atlas class are optimally clustered.

2.2 Force Field Computation

To assess the effect of a displacement u_i of a particular voxel i on the similarity measure $I(Y, L)$, we differentiate $I(Y, L)$ with respect to u_i , using a similar approach as in [5]:

$$\begin{aligned} \frac{\partial I(\mathbf{u} + \epsilon \mathbf{h})}{\partial u_i} &= \sum_k \sum_y \frac{\partial}{\partial u_i} \left[p(k, y) \log \frac{p(k, y)}{p(y) \cdot p(k)} \right] \\ &= \sum_k \sum_y \left[\log \frac{p(k, y)}{p(k)} \cdot \frac{\partial p(k, y)}{\partial u_i} \right] \end{aligned} \quad (3)$$

using the fact that $\sum_k \sum_y \frac{\partial p(k, y)}{\partial u_i} = \sum_k \frac{\partial p(k)}{\partial u_i} = 0$. The derivative of $p(k, y)$ with respect to the displacement u_i of sample i is given by

$$\frac{\partial p(k, i)}{\partial u_i} = \frac{1}{N} \sum_{j=1}^8 \frac{\partial w_{i,j}}{\partial u_i} \delta(y - y_i) \cdot c_{i,j}(k)$$

The derivative of the joint probability distribution is itself a joint probability constructed using the PV interpolation scheme, with weights that are the spatial derivatives of the weights of the original joint histogram. We can thus define the driving forces F_i in each voxel i as:

$$F_i = \frac{\partial I(Y, L)}{\partial \epsilon} = \frac{1}{N} \sum_k \log \frac{p(k, y_i)}{p(k)} \cdot \left(\sum_{j=1}^8 \frac{\partial w_{i,j}}{\partial u_i} c_{i,j}(k) \right)$$

2.3 Viscous Fluid Regularization

We adopt the free-form registration approach of [5] and use the force field $\mathbf{F}(\mathbf{x}, \mathbf{u}) = F_i$ as defined above to drive a viscous fluid regularizer by iteratively solving its Navier-Stokes governing equation:

$$\nabla^2 \mathbf{v} + \nabla (\nabla \cdot \mathbf{v}) + \mathbf{F}(\mathbf{x}, \mathbf{u}) = 0 \quad (4)$$

with $\mathbf{v}(\mathbf{x}, t)$ the deformation velocity experienced by a particle at position \mathbf{x} . An approximate solution of (4) is obtained by convolution with a Gaussian kernel

ψ and the deformation field $\mathbf{u}^{(m+1)}$ at iteration $(m+1)$ is found by integration over time:

$$\mathbf{v} = \psi \star \mathbf{F} \quad (5)$$

$$\mathbf{R}^{(m)} = \mathbf{v}^{(m)} - \sum_{i=1}^3 v_i^{(m)} \left[\frac{\partial \mathbf{u}^{(m)}}{\partial x_i} \right] \quad (6)$$

$$\mathbf{u}^{(m+1)} = \mathbf{u}^{(m)} + \mathbf{R}^{(m)} \cdot \Delta t \quad (7)$$

The time step Δt is constrained by $\Delta t \leq \max(\|\mathbf{R}\|) \cdot \Delta u$, with Δu the maximal voxel displacement that is allowed in one iteration. Regridding and template propagation are used as in [5] to preserve topology.

3 Results

The method presented above was implemented in Matlab. Image resampling, joint histogram construction and force field computation were coded in C. The maximal voxel displacement Δu at each iteration was set to 0.25 voxels and regridding was performed when the Jacobian of the deformation field became smaller than 0.5. Iterations were continued as long as negative $I(Y, L)$ decreased, with a maximal number of iterations of 180. The number of classes used for the template image was 4 (WM, GM, CSF and OTHER). After linear rescaling of the study image intensities to the range $[0, 127]$, the number of bins for the reference image was 128, such that the joint histogram size was 4×128 .

Several experiments were conducted to evaluate the impact of the proposed registration measure on atlas-driven intensity-based tissue segmentation quality. In a first experiment the Brainweb MR brain template [1] was warped to 5 different high resolution MR images of normal brains. For each of these images, tissue maps for WM, GM and CSF were obtained independently using the method described in [4]. The atlas was first affinely aligned with the subject images using MI. Subsequently, three different non-rigid atlas-to-image registration schemes were compared: matching of atlas to subject image intensities using the MI measure as described in [5], matching of atlas to subject tissue class labels using the divergence measure as described in [6], and matching of atlas class labels to subject image intensities using the method proposed here. We refer to these methods as MI, D and HMI (hybrid mutual information) respectively. The performance of the different methods is compared by the overlap coefficient for WM, GM and CSF computed between the warped atlas tissue maps and the segmented tissue maps. All maps were hard segmented (after warping) to compute the overlap coefficients.

Table 1 shows the results. Non-rigid atlas warping using the HMI criterion proposed here generates, in all cases and for all tissue classes, tissue maps that are more similar to the intensity-based tissue segmentations itself than is the

case with standard MI. The results obtained with measures HMI and D are comparable. However, the use of measure D, which matches atlas labels to subject class labels directly, requires availability of the subject segmentation maps, while the HMI measure proposed here does not. Moreover, matching atlas class labels to subject class labels using measure D does not consistently improve the registration quality compared to matching to subject intensities using the HMI measure.

In a second experiment an in-house constructed brain atlas, build from a database of 64 normal brain images by non-rigid warping, was warped to the same five subject images considered above. The atlas was first aligned with the subject images by affine registration and was subsequently non-rigidly warped to the subject images using the standard MI measure and the HMI measure proposed here (figure 1). The subject images were segmented with the method of [4] using a priori tissue probability maps derived from the affinely registered atlas and from the non-rigidly warped atlas using either of both methods.

Figure 2 shows the conditional probability $p(y|k)$ of intensities y in one of the subject images for the GM class before (i.e. only affine registration) and after non-rigid matching and, for both cases, before and after subsequent segmentation, i.e. using the warped a priori atlas GM classification or the computed a posteriori GM classification respectively. These graphs show the impact of the prior model and of the image data on the GM intensity model estimated during segmentation. It can be seen that non-rigid atlas registration (blue curve) clusters the GM intensities more than merely affine registration (black curve). HMI based warping (left) clusters the GM intensities more efficiently than standard MI (right). Subsequent intensity-based classification of the images using either the affinely or the non-rigidly warped atlas priors results in a class-conditional intensity distribution that is nearly Gaussian and that is fairly similar for segmentations obtained with both affinely and non-rigidly warped priors, especially with standard MI. Nevertheless, even if the intensity distribution within each tissue class might be more or less identical, the classification map itself can be quite different due to the impact of the prior model on the classification itself.

The effect of atlas warping on the segmentation quality can also be appreciated from the negative log-likelihood curves $-\log P(Y|\theta)$ of the data Y given the model parameters θ during iterative classification and parameter estimation [4]. As illustrated in figure 2 for one of the five subject images considered here, the negative log-likelihood associated with the segmentation based on priors warped using the HMI method presented here (red curve) is smaller than with the segmentation based on priors warped using standard MI (black curve). This confirms that the HMI warped prior better explains the image data than the standard MI warped prior.

Considering the tissue maps obtained independently using the affinely registered SPM atlas as ground truth (as in the first experiment), we can compute the overlap coefficients between this ground truth and the segmentations obtained with our atlas prior, after affine and after non-rigid warping using HMI and standard MI respectively. These results are summarized in table 2.

Table 1. Overlap coefficients (in %) for different tissue classes in 5 different subject images between the segmented tissue maps and atlas tissue maps warped to the study image using affine and subsequent non-rigid matching with three different registration schemes.

Case	Affine			MI			D			HMI		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
1	67.28	69.12	42.70	67.78	69.99	52.73	75.52	78.08	69.90	81.89	79.57	63.11
2	66.13	68.95	41.73	70.59	72.78	58.73	77.39	79.94	70.68	81.78	79.39	65.48
3	61.21	66.79	40.48	67.54	68.53	52.38	76.11	79.81	70.22	80.18	78.94	62.90
4	66.77	68.54	41.80	68.26	70.85	58.33	75.78	78.71	70.42	77.22	74.78	60.09
5	65.37	67.39	44.97	69.05	70.85	60.28	71.57	73.83	70.43	80.94	77.62	66.37

Table 2. Overlap coefficients (in %) for different tissue classes of tissue maps obtained with affinely registered and with non-rigidly warped atlas priors using HMI and standard MI, using an independent segmentation of the subject image as ground truth.

Case	Affine registration			HMI			MI		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
1	91.54	90.59	69.00	99.06	98.49	95.41	94.20	93.67	76.01
2	92.41	90.51	67.66	97.35	96.32	79.74	95.23	94.45	76.41
3	91.10	90.58	72.99	97.92	96.43	80.34	95.13	94.50	82.95
4	92.64	91.40	69.70	96.74	96.14	81.35	93.22	91.99	76.06
5	91.79	89.77	72.21	96.88	95.96	80.25	94.88	94.29	81.04

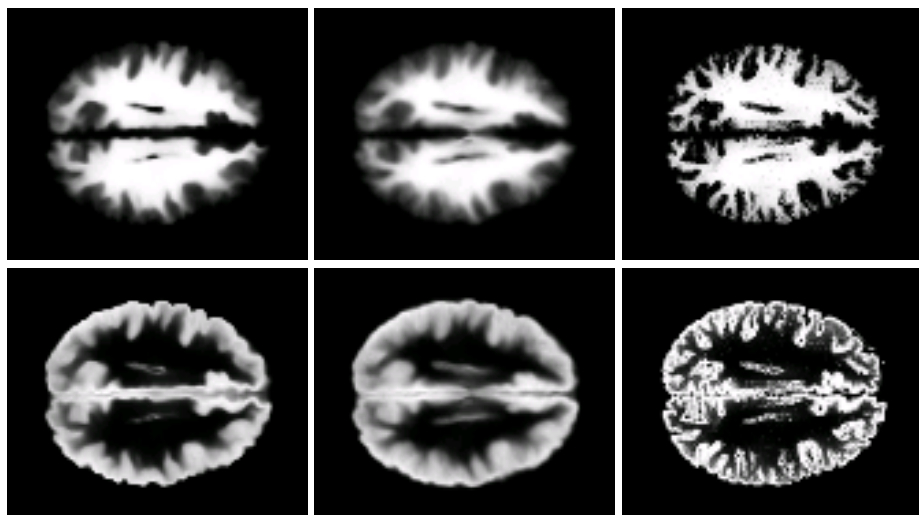


Fig. 1. WM (top) and GM (bottom) atlas priors warped to a particular subject brain using non-rigid registration with the proposed HMI measure (left column) and with standard MI (middle column). Right column: reference segmentation maps (ground truth).

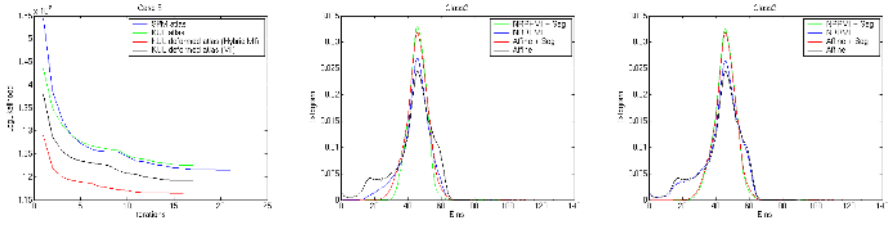


Fig. 2. Left: Negative log-likelihood $-P(Y|\theta)$ curves during iterative model-based tissue classification and parameter estimation for different segmentation strategies. Middle and right: conditional probabilities $p(y=k)$ given one class (gray matter) before and after segmentation and/or non-rigid registration for one of the five cases. Middle: hybrid mutual information based non-rigid registration; Right: mutual information based non-rigid registration.

4 Discussion

In this paper we present a hybrid mutual information registration criterion for atlas-to-image matching that allows matching of atlas class probabilities to study image intensities. The new registration criterion measures how well the probabilistic atlas tissue distributions explain the intensities observed in the study image. A similar approach has been presented in [7] with the difference that in this paper the author uses mutual information between reference image intensities and binary labels in the floating image. Compared to the standard MI measure of corresponding voxel intensities, the proposed approach has the advantage that spatially varying tissue information is explicitly introduced in the registration criterion, which makes it more robust. On the other hand, in contrast with the divergence criterion introduced in [6] for matching atlas and study image class labels, no segmentation of the study image needs to be available for the method presented here.

The joint probability distribution $p(k, y)$ between atlas class labels k and study image intensities y from which the HMI measure $I(Y, L)$ is computed, is estimated during registration using PV interpolation [3], such that it varies smoothly with individual voxel displacements and can be analytically differentiated. A force field is thus obtained that acts to displace individual voxels such that the mutual information between atlas class labels and study image intensities is maximized. The approach presented here is completely discrete due to the PV interpolation scheme and, in contrast with the method of [5], the force field does not depend on spatial image intensity gradients.

Several experiments were performed to evaluate the impact of atlas warping using various registration schemes on atlas-driven intensity-based tissue segmentation, showing that the hybrid registration measure proposed here indeed performs better for this particular task than standard MI. Our further work will focus on merging atlas-driven labelling and label-based atlas matching in a single information-theoretic framework, whereby each process benefits from the output of the other.

References

- [1] C.A. Cocosco and V. Kollokian and R.K.-S. Kwan and A.C. Evans. BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. *Proceedings of 3-rd International Conference on Functional Mapping of the Human Brain*, 5(4), part 2/4, S425, 1997.
- [2] The SPM package is available online at <http://www.fil.ion.ucl.ac.uk/spm/>
- [3] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Medical Imaging*, 16(4):187–198, 1997.
- [4] K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Medical Imaging*, 18(10):897–908, 1999.
- [5] E. D’Agostino, F. Maes, D. Vandermeulen, and P. Suetens. A viscous fluid model for multimodal non-rigid image registration using mutual information *Medical Image Analysis*, 7(4):565–575, 2003.
- [6] E. D’Agostino, F. Maes, D. Vandermeulen, and P. Suetens. An information theoretic approach for non-rigid image registration using voxel class probabilities *MICCAI 2003*, Lectures Notes in Computer Science, vol 2878:812–820
- [7] J. Kim, J. W. Fisher III, A. Jr. Yezzi, M. Cetin and A. S. Willsky. Nonparametric methods for image segmentation using information theory and curve evolution *IEEE ICIP 2002*, IEEE International Conference on Image Processing, vol. 3, pp. 797–800, Rochester, New York, September 2002.