A new label fusion method using graph cuts: application to hippocampus segmentation

C. Platero , M. C. Tobar , J. Sanguino and O. Velasco

Abstract— **The aim of this paper is to develop a probabilistic modeling framework for the segmentation of structures of interest from a collection of atlases. Given a subset of registered atlases into the target image for a particular Region of Interest (ROI), a statistical model of appearance and shape is computed for fusing the labels. Segmentations are obtained by minimizing an energy function associated with the proposed model, using a graph-cut technique. We test different label fusion methods on publicly available MR images of human brains.**

I. Introduction

Automatic segmentation of subcortical structures in human brain MR images plays a crucial role in clinical practice. Specifically, hippocampus segmentation is an important tool for the study of neurodegenerative diseases. Nowadays, brain MR images have poor quality due to their inherently low spatial resolution, insufficient tissue contrast and ambiguous tissue intensity distributions. To overcome these difficulties, many approaches have been proposed. Atlas-based segmentation has become a standard technique to identify structures from brain MR images. An atlas, in the context of this paper, is an image in one modality with its respective labeling (usually generated by manual segmentation). Segmentations with a single atlas are intrinsically biased towards the shape and the appearance of a subject. Several studies have shown that approaches which incorporate properties of a group of atlases outperform the use of a single atlas [1, 2]. There are two different atlas-based segmentation strategies using multiple atlases: a) Probabilistic atlas and b) Multi-atlas segmentation. In a probabilistic atlas, the information from atlases is combined into a mathematical model in a common coordinate system. The advantage is that, once the probabilistic atlas has been generated, only a single registration is required for obtaining the segmentation. However, this method depends on the success of a single registration. An alternative strategy is to register each atlas to the target image separately. The main benefit of the multi-atlas segmentation approach is that the effect of the errors associated with any single atlas propagation can be reduced in the process of combination. Similar to the probabilistic atlas, the transferred atlases are used to build a model for segmenting the target image. This process is often called label fusion. The main drawback of the multi-atlas segmentation is the computational complexity. However, not all atlases have to be registered into the target image [2]. Aljabar et al [3] showed that an atlas selection framework is required for ranking the atlases and fixing a number of atlases to be fused which depends on the application. These studies also indicate that the similarity between the target image and the atlas is a crucial factor for improving registration and segmentation accuracies. Furthermore, brain images show different structures of interest to be segmented. Therefore, a region-wise approach is more appropriate [4]. This can be achieved by dividing the image into multiples anatomically meaningful regions. Once defined the ROIs, a ranking of atlases is calculated. The transferred labels, which belong to the selected atlases, are fused into the ROI of the target image. The fusion of the propagated segmentations can be achieved in different ways: STAPLE [5], majority voting rule or minimization of an energy function with intensity and prior terms [6]. Recent works have shown that statistical models from the registered atlases can improve the segmentation quality [7, 8]. The aim of this paper is to develop a probabilistic modeling framework for the segmentation of structures of interest from a collection of atlases. The paper is organized as follows. In Section 2, the label fusion method is presented. Experiments for the hippocampus segmentation are described in Section 3. Conclusions are presented in Section 4.

II. Label fusion method

We present a label fusion method based on minimizing an energy function by using graph-cut technique. This energy function incorporates terms of appearance and shape, which are estimated from the training atlases. Other authors have previously used this framework [6, 9]. Our label fusion method has the following differences from previous proposals: a) An appearance generative model based on multiple features extracted from each pixel and its neighborhood, b) A label prior probability is estimated by using a weighted voting method [7] and c) A spatial regularizer that minimizes the surface of separation between two different labels [10].

Consider a set of N training atlases for each ROI $\{A_i\}_{i=1,\dots,N} = \{I_i, S_i\}_{i=1,\dots,N}$ and a target image I, where $I_i : \Omega_i \subset \mathbb{R}^n \to \mathbb{R}$, $n = 3$ and $S_i : \Omega_i \subset \mathbb{R}^n \to \{0,1\}$ are the label maps. We assign to $S(x) = 1$ the foreground pixels and $S(x) = 0$ to the background pixels. We denote $\Phi_i : \Omega \to \Omega_i$ to be the spatial mapping from the target image coordinates to the coordinates of $i-$th training subject. For simplicity, we assume that $\{\Phi_i\}_{i=1,\dots,N}$ have been pre-computed using a pairwise registration procedure. This assumption allow us to shorthand $\mathbb{A} = \{\tilde{S}_i = S_i \circ \Phi_i, \tilde{I}_i = I_i \circ \Phi_i\}_{i=1,\dots,N}$ as the training set into the common coordinates. The segmentation of an image I, based on image intensities and prior knowledge, is computed by the minimization of an energy function.

$$S = \arg\min_S E^{\mathbb{A}}(S), \qquad E^{\mathbb{A}}(S) = E_B^{\mathbb{A}}(S) + E_F(S), \tag{1}$$

where the term $E_B^{\mathbb{A}}(S)$ is derived from $\mathbb{A}$ using the framework of Bayesian estimation theory and $E_F(S)$ is associated with an image-based Finsler metric.

A. *Probabilistic model*

To find the MAP estimation is equivalent to minimize the following energy function where the Bayes theorem is applied

$$E_B^{\mathbb{A}}(S) = -\log\left(p(S|I;\mathbb{A})\right) = -\log\left(\frac{p(I|S;\mathbb{A})p(S;\mathbb{A},I)}{p(I;\mathbb{A})}\right).$$

We assume that the observed intensities of I are independent random variables. The image likelihood $p(I|S;\mathbb{A})$ can then be written as a product of the likelihoods of the individual pixels: $p(I|S;\mathbb{A}) = \prod_{x\in\Omega} p(I(x)|S(x);\mathbb{A})$. Usually, the intensity distribution is modeled by a mixture of Gaussians [11]. Alternatively, we use a multivariate Gaussian distribution for each pixel and for each label [4]: $p(I(x)|l;\mathbb{A}) = \frac{1}{(2\pi)^{f/2}|\Sigma_l(x)|^{1/2}} \exp(-\frac{1}{2}(I(x) - \mu_l(x))^T \Sigma_l^{-1}(x)(I(x) - \mu_l(x)))$, where $l \in \{0,1\}$, μ is the mean, Σ is the covariance matrix and f is the dimension of the feature space. The effect of sample size has to be considered on feature selection. The means and covariance matrices are estimated by using a variable number of samples $\#Q_l(x)$, where $Q_l(x) = \{i|\tilde{S}_i(x) = l\}$. The number of observations requires from each of the two class to ensure that the classification error is bounded relative to a infinite number of samples depend on f ($f \leq \frac{1}{5}\min(\#Q_0(x), \#Q_1(x))$ for $f \leq 8$ [12]). To get that the gaussian parameters are the least biased, it is used a neighborhood system around the pixel for obtaining more samples. The gaussian parameters are computed from $\mathbb{A}$:

$$\mu_l(x) = \frac{\sum_{y\in\mathscr{N}(x)} \sum_{i\in Q_l(y)} \tilde{I}_i(y)}{\sum_{y\in\mathscr{N}(x)} \#Q_l(y)} \tag{2}$$

and

$$\Sigma_l(x) = \frac{\sum_{y\in\mathscr{N}(x)} \sum_{i\in Q_l(y)} (\tilde{I}_i(y) - \mu_l(x))(\tilde{I}_i(y) - \mu_l(x))^T}{\sum_{y\in\mathscr{N}(x)} \#Q_l(y) - 1}. \tag{3}$$

Further, f is variable in each pixel. For each pixel is analyzed the correlation matrix. It only selects uncorrelated features in runtime.

The label prior probability $p(S;\mathbb{A},I)$ models the joint probability of all pixels in a particular label configuration. Instead, we assume that the prior probability that pixel x has label l only depends on its position and the similarity between I and $\tilde{I}_i$: $p(S;\mathbb{A},I) = \prod_{x\in\Omega} p(S(x);\mathbb{A},I)$. For each pixel x and each label $l \in \{0,1\}$, we define

$$p(S(x) = l;\mathbb{A},I) = \frac{\sum_{i\in Q_l(x)} m(I(x),\tilde{I}_i(x))^q}{\sum_{l=0}^{1} \sum_{i\in Q_l(x)} m(I(x),\tilde{I}_i(x))^q} \tag{4}$$

where $m(I(x),\tilde{I}_i(x))$ is a global or local similarity measure between the target image and the registered atlas image and q is an associated gain exponent [7].

B. *Spatial regularization*

Following the work of Boykov and Kolmogorov [10], the smoothness term E_F of the energy function is defined from a Finsler metric. These authors decomposed the energy into E_R and E_f with weights $\lambda_1, \lambda_2 \in \mathbb{R}, \lambda_1 \geq 0,$ that is,

$$E_F(S) = \lambda_1 E_R(S) + \lambda_2 E_f(S).$$

The first part minimizes the segmentation surface by a Riemannian metric and the second one takes into account the orientation of the segmentation surface in the metric. We consider that the isotropic Riemannian metric from the image is defined by $D(x) = g(\|\nabla I(x)\|)\mathbb{I}$, where $\mathbb{I}$ is an identity matrix, $g(x) = (\exp(-x/\gamma))^{1/3}$ and γ is estimated by the average of $\|\nabla I(x)\|$. The energies are defined by

$$E_R(S) = \sum_x \sum_{y,\,\{x,y\}\in\mathscr{N}} \omega_x^R(y)(1-S(x))S(y),$$

$$E_f(S) = \sum_x \sum_{y,\,\{x,y\}\in\mathscr{N}} \omega_x^f(y)(S(x)(1-S(y)) - S(y)(1-S(x))),$$

where $\mathscr{N}$ is a neighborhood system, $\omega_x^R(y) = \frac{g(\|\nabla I(x)\|)}{\|x-y\|}$ and $\omega_x^f(y)$ is the component of the vector $\nabla I(x)$ along the vector defined by x and y.

C. Optimization

For the min-cut/max-flow algorithms, the energy to be minimized is represented by a weighted graph, $\mathscr{G} = \langle \mathscr{V}, \mathscr{E} \rangle$, with two special nodes, namely the source s and the sink t. The rest of nodes represents the pixels of the image. The set of edges is denoted by $\mathscr{E} = \mathscr{E}_{\mathscr{N}} \bigcup \mathscr{E}_{\mathscr{T}}$, where $\mathscr{E}_{\mathscr{N}}$ denotes the set of pixel-to-pixel edges in the defined neighborhood system (n-links) and $\mathscr{E}_{\mathscr{T}}$ denotes the set of pixel-to-terminal s or t edges (t-links). We assign a nonnegative cost $c(x,y)$ to each edge $(x,y) \in \mathscr{E}$. In the considered energy, the cost of a $t-$link is defined by the prior probabilities and the coefficients ω_x^f, while the cost of a $n-$link is given by the coefficients ω_x^R. It easily follows that for all x and y with $(x,y) \in \mathscr{N}$

$$c(s,x) = -\log\left(p(S(x) = 1|I;\mathbb{A})\right) + \lambda_2 \sum_{y,\,\{x,y\} \in \mathscr{N}} \omega_x^f(y),$$

$$c(x,t) = -\log\left(p(S(x) = 0|I;\mathbb{A})\right) + \lambda_2 \sum_{y,\,\{x,y\} \in \mathscr{N}} \omega_y^f(x),$$

$$c(x,y) = \lambda_1(\omega_x^R(y) + \omega_y^R(x)). \tag{5}$$

III. Experiments with Brain MR data

To evaluate the performance of the different label fusion methods, we employ an available database of T1-weighted MR images of epileptic and nonepileptic subjects [13]. Images were acquired by using two MR imaging systems with field strengths (1.5 T and 3.0T) and thus have different resolutions (0.78 x 2 x 0.78 mm^3 and 0.39 x 2 x 0.39 mm^3). All atlases are skull-stripped using BET [14]. An atlas (HFH_021) is selected as a reference to which all atlases are then co-registered with an affine transformation using FLIRT [15]. After spatial normalization, a region is defined for each structure studied (left and right hippocampus) as the minimum bounding box containing the structure for all training atlases expanded by three pixels along each dimension. The size of these ROIs is in the range of 105 x 38 x 115 pixels. The target image is also normalized and parceled by using BET, FLIRT and predefined ROIs. For each ROI, the atlases are ranked based on their similarities with the target image. Then, the selected atlases are co-registered non-rigidly to the ROI of the query image using a B-spline registration with an isotropic grid spacing of 3.0 mm. All nonrigid registration are computed using *Elastix* [16]. The negative Mutual Information (MI) is used as the cost function. Finally, the transferred labels are fused and an inverse affine transformation is applied to return the segmentation into the native space.

A. Setting parameters

The proposed label fusion method has several input parameters: a) Similarity measures, b) Scalar features for the appearance term, c) The Lagrange multipliers λ_1 and λ_2 of the energy function and d) The optimal number of atlases to be fused. To set these parameters, twenty five leave-one-out segmentations on the training atlases are performed to determinate the tunable parameters. These parameters are varied in certain ranges and their effects are measured by the overlap between the resulting segmentation and the ground truth. Dice coefficient is chosen as a measure of the segmentation overlaps. The parameters are adjusted to give the highest values of Dice coefficient.

MI is used as the similarity measure. For the weighted voting rule, a semi-local strategy is used to calculate the similarity for each registered atlas. A mask image is built by joining all transferred label images. This mask image is used to define the domain for measuring the similarity between the image target and the registered atlases. This strategy is specially suitable when contrast between neighboring structures is low [7], as in the hippocampus.

For each T1-weighted MR image, the following features are calculated: intensity, gradients, laplacians, curvatures and local entropies in different scales. Some of these features are not invariants in gray level so an intensity normalization is applied to the registered atlas images by histogram matching. Spatial derivatives are implemented by Gaussian-derivative filters. In our experiments, the optimal scale is $\sigma = 2$ for the Guassian masks. Bhattacharyya distances and Dice coefficients are used to identify those features which are more important in discrimination among labels. These scalar features are the intensity, the gradient module and the local entropy. To estimate the statistical parameters of the appearance term and since the number of samples for any label is low, a 26-neighborhood system in the sagittal plane is tuned and applied to the equations (2) and (3). In runtime, a matrix of correlation coefficients is calculated for building the quadratic classifiers. The features, whose correlation coefficients are below 0.6 in absolute value, are considered independents and are used in the classifier. The dimension of the feature space is variable for each pixel and can be 3, 2 or 1.

The Lagrange multipliers λ_1 and λ_2 of the energy function are tuned by Dice evaluation. We have observed that the Riemannian metric is more influential than the surface orientation term in the optimization process. Considering 3D grid-graphs with 6 neighborhood system in (5), the edge weights are calculated with $\lambda_1 = 4$ and $\lambda_2 = 1$. The computational burden is reduced by calculating the edge weights of the n-link only for pixels whose labels have uncertainty.

Table 1: Correlation for manual and automatic volumes and mean and standard deviation values of the five quality measures: correlation (r), Dice coefficient (m_1), relative absolute volume difference (m_2), average symmetric surface distance (m_3), root mean square symmetric surface distance (m_4), and maximum symmetric surface distance (m_5).

Type		r	m_1	m_2	$m_3\,[mm]$	$m_4\,[mm]$	$m_5\,[mm]$
STAPLE	LH	0.48	0.722 ± 0.121	0.22 ± 0.19	0.79 ± 0.37	1.03 ± 0.44	5.26 ± 1.73
	RH	0.52	0.737 ± 0.063	0.24 ± 0.15	0.74 ± 0.12	0.97 ± 0.22	5.17 ± 1.53
Majority	LH	0.36	0.722 ± 0.127	−0.15 ± 0.14	0.77 ± 0.37	0.99 ± 0.46	4.62 ± 1.68
Voting	RH	0.51	0.744 ± 0.073	0.11 ± 0.09	0.70 ± 0.14	0.91 ± 0.21	4.41 ± 1.21
Weighted	LH	0.57	0.732 ± 0.069	−0.16 ± 0.12	0.72 ± 0.17	0.96 ± 0.28	4.74 ± 1.70
Voting	RH	0.63	0.757 ± 0.045	−0.12 ± 0.09	0.66 ± 0.06	0.84 ± 0.10	4.06 ± 0.85
Our approach	LH	0.66	0.753 ± 0.073	0.22 ± 0.15	0.74 ± 0.13	0.98 ± 0.20	6.38 ± 1.66
1F	RH	0.77	0.773 ± 0.055	0.20 ± 0.11	0.70 ± 0.10	0.91 ± 0.21	5.41 ± 1.56
Our approach	LH	0.70	0.754 ± 0.061	0.13 ± 0.11	0.74 ± 0.11	0.99 ± 0.21	6.18 ± 1.59
3F	RH	0.79	0.778 ± 0.050	0.11 ± 0.09	0.69 ± 0.10	0.92 ± 0.22	5.06 ± 1.92

In the atlas selection framework, once the atlases are ranked by MI in the whole ROI, for all segmentation methods we employ a leave-one-out validation strategy, where an optimal number of atlases is tuned. According to the ROI and the segmentation method, the number of the fused atlases varies from 6 to 15.

B. Results

We compare five label fusion methods: STAPLE, Majority Voting (MV), Weighted Voting (WM) and two methods that we derive from our proposal. In the appearance term, we consider either a singular feature using the intensity (1F) or the proposal with multi-features (3F). The same parameters are applied to WM with respect to our proposal in label prior (semi-local, MI and $q = 4$). The performances of these approaches are evaluated by comparing six measures for the cases of left (LH) and right (RH) hippocampus segmentations. Table 1 gives the correlation for manual and automatic volumes and the mean and standard deviation values of the five quality measures for each method and each ROI. The minus sign in m_2 indicates a result of under-segmentation. Our scores on MV are slightly worse than those given in [13]. This could be because it has not been applied any manual correction in BET. A paired t-test is applied in Dice distributions between MV and the other methods with the following results: STAPLE (LH $p = 0.888$, RH $p = 0.105$), WV (LH $p = 0.515$, RH $p = 0.124$), 1F (LH $p = 0.056$, RH $p = 0.035$) and 3F (LH $p = 0.028$, RH $p = 0.008$).

IV. Conclusions

We introduce a new label fusion method. It combines an appearance generative model based on multiple features with a label prior using a weighted voting method and a spatial regularizer that minimizes the surface of separation between two different labels. The proposed combination provides high accuracy in segmentation, it shows significant improvements in relation to the conventional framework and also the best correlation between manual and automatic volumes. The proposed method is generic and could be incorporated to other applications.

References

1. Rohlfing T., Brandt R., Menzel R., Russakoff D., Maurer C.. Quo vadis, atlas-based segmentation? *Handbook of Biomedical Image Analysis.* 2005:435–486.
2. Heckemann R.A., Hajnal J.V., Aljabar P., Rueckert D., Hammers A.. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion *NeuroImage.* 2006;33:115–126.
3. Aljabar P., Heckemann RA, Hammers A., Hajnal JV, Rueckert D.. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy *NeuroImage.* 2009;46:726–738.
4. Han X., Fischl B.. Atlas renormalization for improved brain MR image segmentation across scanner platforms *Medical Imaging, IEEE Transactions on.* 2007;26:479–486.
5. Warfield S.K., Zou K.H., Wells W.M.. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation *Medical Imaging, IEEE Transactions on.* 2004;23:903–921.
6. Lijn F., Heijer T., Breteler M., Niessen W.J.. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts *NeuroImage.* 2008;43:708–720.
7. Artaechevarria X., Muñoz-Barrutia A., Solorzano C.. Combination strategies in multi-atlas image segmentation: Application to brain MR data *Medical Imaging, IEEE Transactions on.* 2009;28:1266–1277.
8. Sabuncu M.R., Yeo B.T.T., Van Leemput K., Fischl B., Golland P.. A generative model for image segmentation based on label fusion *Medical Imaging, IEEE Transactions on.* 2010;29:1714–1729.
9. Wolz Robin, Heckemann Rolf A, Aljabar Paul, et al. Measurement of hippocampal atrophy using 4D graph-cut segmentation: application to ADNI *NeuroImage.* 2010;52:109.
10. Kolmogorov V., Boykov Y.. What metrics can be approximated by geo-cuts, or global optimization of length/area and flux in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*;1:564–571 2005.
11. Pohl K.M., Fisher J., Grimson W.E.L., Kikinis R., Wells W.M.. A Bayesian model for joint segmentation and registration *NeuroImage.* 2006;31:228–239.
12. Raudys S.J., Jain A.K.. Small sample size effects in statistical pattern recognition: Recommendations for practitioners *IEEE Transactions on pattern analysis and machine intelligence.* 1991;13:252–264.
13. Jafari-Khouzani Kourosh, Elisevich Kost V, Patel Suresh, Soltanian-Zadeh Hamid. Dataset of magnetic resonance images of nonepileptic subjects and temporal lobe epilepsy patients for validation of hippocampal segmentation techniques *Neuroinformatics.* 2011;9:335–346.
14. Smith Stephen M. Fast robust automated brain extraction *Human brain mapping.* 2002;17:143–155.
15. Jenkinson Mark, Bannister Peter, Brady Michael, Smith Stephen, others . Improved optimization for the robust and accurate linear registration and motion correction of brain images *Neuroimage.* 2002;17:825–841.
16. Klein S., Staring M., Murphy K., Viergever M.A., Pluim J.P.W.. elastix: a toolbox for intensity-based medical image registration *Medical imaging, IEEE transactions on.* 2010;29.