Cortical Brain Structures Segmentation Using Constrained Optimization and Intensity Coupling

Alireza Akhondi-Asl[1], Simon K. Warfield[1]

[1]Computational Radiology Laboratory, Children's Hospital Boston, and Harvard Medical School
300 Longwood Ave. Boston MA 02115 USA

ABSTRACT

Brain image segmentation is one of the most important applications in medicine and also is one of the most challenging topics in the field of medical image processing. In general, most automatic segmentation methods consist of an energy function, a shape model, and an optimization strategy. Each plays an important role in the design of an accurate segmentation algorithm. Here we introduce a modified version of a coupled structure segmentation algorithm that is based on earlier paper. Specifically, we have 1) utilized a multiple atlas strategy to estimate a joint probability mass function of the location and tissue type information of the structures; 2) analyzed the relationship among the various structures to achieve more robust probability density function (pdf) estimation; 3) added a constraint to the optimization process to minimize intersection among the different structures; and 4) demonstrated the effectiveness of the method for the segmentation of certain brain structures.

Keywords— *medical image segmentation, Intensity coupling, constrained optimization, Multiple Structure, Level Set, MRI.*

1. Introduction

Magnetic resonance imaging (MRI) is the modality of choice for many medical applications because of its unique ability to achieve highly accurate, non-invasive imaging of human anatomy without the potentially harmful effects of ionizing radiation. However, MRI is not without its limitations in the segmentation of the brain structures (e.g., bias field, intensity inhomogenity, partial volume, and signal-to-noise ratios). There have been numerous attempts to overcome these problems, as evidenced by the many algorithms that have been discussed in the literature [1]. In general, most segmentation methods consist of an energy function, a shape model, and an optimization strategy, and each plays an important role in the design of an accurate segmentation algorithm. In [2-3] we have introduced a coupled structures segmentation algorithm that is based on a principal component analysis (PCA) designed to extract shape relationships among structures. In [4] we have added constraint to the shape parameters to achieve a more robust segmentation algorithm. In [5] we have modified the energy function by considering tissue type and location of the structures as independent information. However, it should be noted that the model of tissue type was for limited number of structures. In [6-7] we modified the method to work with multiple atlases, using it specifically for the volumetry of the hippocampus and lateralization in the epilepsy patients. We have addressed the main limitations in the previous version of our work first, by considering the dependency of tissue type and location of the structures based on the information extracted from the atlases; second, by modifying the intensity pdf estimation to consider intensity relationships among different structures having similar intensity distributions; and third, by adding the constraint to overcome unwanted intersections among different structures.

2. Energy Function

The goal of the segmentation algorithm is to simultaneously segment m different structures in a skull stripped image ($I(x)$) using n skull stripped training intensity datasets ($I_{Tr}^i(x)\, i \in \{1,...,n\}$) and corresponding training label datasets ($L_{Tr}^i(x)\, i \in \{1,...,n\}$) in which all of the m structures are segmented and labeled with $l \in \{1,...,m+1\}$. In these labeled images, label $m+1$ is assigned to the region outside all of these m structures (outside region). In our previous work, for each voxel and for each one of the regions Ω_k, $k \in \{1,...,m+1\}$, we defined a probability density function (pdf) $p(x \in \Omega_k) = p_k(x) \times p_{tk}(x)$ where $p_{tk}(x) = p_{fk}(x) \times p_{sk}(x)$. In this definition, $p_{fk}(x)$ was the probability mass function (pmf) of the tissue type (the intensity ranges); $p_{sk}(x)$ was the pdf of the location; and $p_k(x)$ was the pdf for intensity distribution of the kth structure. To estimate the pmfs of the tissue types of the structures, we applied the clustering twice with 3 and 10

clusters respectively. The 3-class segmentation results captured the global intensity information of the normal tissues in the brain; while the 10-class segmentation captured their local or fine intensity information. Equipped with this information, we then modeled the pmf of the tissue type of each structure (the probabilities of each one of the 10-class clusters being present in the structure) [5]. To compute $p_{sk}(x)$, we used the multiple atlas strategy to register $I_{Tr}^i(x)\, i \in \{1,...,n\}$ to $I(x)$ with a non-rigid registration method. To this end, for each of the datasets, we found the transform and named it as T^i . Next, we applied these transforms to the label map images ($L_{Tr}^i(x)\, i \in \{1,...,n\}$) to find ${}^{T^i}L_{Tr}^i(x)$ (the transformation of $L_{Tr}^i(x)$ with T^i). We then applied these values to estimate the pdf of location for each one of the structures [6]. Thus, we considered $p_{fk}(x)$ and $p_{sk}(x)$ as independent pdf and pmf estimates, respectively, and we defined each separately. However, there is a relationship between them that can be extracted based on the training data. In considering these relationships, we defined a pmf that takes into account both location and tissue type information. First, we clustered the intensity image $I(x)$ based on the method in [5] to 10 classes to extract the image $S(x)$. In addition, we clustered, ${}^{T_i}I_{Tr}^i(x)\, i \in \{1,...,n\}$ using the same strategy to extract ${}^{T_i}S_{Tr}^i(x)\, i \in \{1,...,n\}$. Then we defined:

$$W_i(x) = \exp\left(-\left({}^{T_i}S_{Tr}^i(x) - S(x)\right)^2 / 2\right) \qquad (1)$$

For a voxel x in the Image I, $W_i(x)$ will be equal to one, if both of the $S(x)$ and ${}^{T_i}S_{Tr}^i(x)$ are in the same class; further, it will be smaller than one based on the difference between the two classes. We suppose the resulting clusters are ordered such that the center of the first cluster has the lowest value (representing darkest regions of the image), and the last cluster center has the highest value (representing the brightest region of the image). This approach is more effective than the algorithm in [8] where the differences among the intensities are considered directly. In the next step, for each cluster, $k \in \{1,...,m+1\}$, we define:

$$p_{tk}(x) = \frac{\sum_{j=1}^{n} W_j(x)\delta_k\left({}^{T^j}L_{Tr}^j(x)\right)}{\sum_{j=1}^{n} W_j(x)} \qquad (2)$$

where $\delta_k\left(u(x)\right) = \begin{cases} 1 & if \quad u(x) = k \\ 0 & if \quad u(x) \neq k \end{cases}$. This is different from

the $p_{tk}(x) = p_{fk}(x) \times p_{sk}(x)$ in [5], which models them as independent pdf and pmf. The new probability mass

function is more accurate, because it considers the relationship between the intensity and location of the structures. Next, we estimated the intensity pdf for each one of the structures. Accordingly, to estimate $p_k(x)$, we use the Parzen window estimator that is defined in the form of

$$p_k(x) = \frac{1}{|\Omega|}\int_\Omega K\left(I(x) - I(\hat{x})\right)d\hat{x} \quad [9].$$ In this equation,

$K(t)$ is a Gaussian kernel with the variance σ, which is estimated by the method described in [10]. In addition, $|\;|$ is used to calculate the cardinality which forms the number of members in a given set or region. In general, for any structure, points identified inside a given region of the structure are used to estimate the intensity of the pdf [2]. However, structures of the same type are known to have very similar pdfs and, as a result, can be used to develop a more robust estimation of pdfs. Thus, in considering intensity information and the relationships among structures; we have defined the following set:

$$S_i = \left\{\forall k \,\middle|\, \Omega_k \in family(\Omega_i)\right\} \qquad (3)$$

where $family(\Omega_i) = \left\{\forall\Omega_k \,\middle|\, pdf(\Omega_k) \equiv pdf(\Omega_i)\right\}$ and $\equiv$ is used to show the equivalency. This set shows the indices of the regions that have the same intensity distribution as region i. We can therefore estimate pdfs for each $m+1$ region using the following two equations:

$$p_{m+1}(x) = \frac{1}{|\Omega_{m+1}|}\int_{\Omega_{m+1}} K\left(I(x) - I(\hat{x})\right)d\hat{x}$$

$$p_k(x) = \frac{1}{\sum_{i \in S_k}|\Omega_i|}\int_{\sum_{i \in S_k}\Omega_i} K\left(I(x) - I(\hat{x})\right)d\hat{x} \quad k = 1,...,m \qquad (4)$$

Clearly, $p_i(x) = p_k(x)$ for $\forall k \in S_i$, which suggests that a) pixels in regions of the same or similar type would be appropriate for pdf estimation; and b) similar regions have the same pdf. Finally, we have constructed the following energy function:

$$\min J(\phi_1, \phi_2, ..., \phi_m) =$$
$$\sum_{j=1}^{m+1} \int_{\Omega_j} F\left(p_j(x) \times p_{tj}(x)\right)dx \qquad (5)$$

where we set $F(.) = -ln(.)$. This is a decreasing function in which m sign functions should be optimized (ϕ_k). In addition, we can also define: $\Omega_k \equiv \left\{x \,\middle|\, H\left(-\phi_k(x)\right) > 0\right\}$ for $k \in \{1, 2, ..., m\}$ and $\Omega_{m+1} \equiv \left\{x \,\middle|\, \prod_{k=1}^{m} H\left(\phi_k(x)\right) > 0\right\}$ where H is Heaviside function.

3. Constrained Optimization

Our previous work was based, in part, on the assumption that there is no intersection among the regions in the optimization process. This assumption, however, is not always correct. To overcome this error in our analysis, we have added a constraint to the optimization to prevent structures from intersecting. Accordingly, we have applied the following constrained energy function:

$$\min J\left(\phi_1, \phi_2, ..., \phi_m\right) =$$

$$\sum_{j=1}^{m+1} \int_{\Omega_j} F\left(p_j(x) \times p_{tj}(x)\right) dx$$

$$subject \quad to \quad C\left(\phi_1, \phi_2, ..., \phi_m\right) = \tag{6}$$

$$\sum_{j=1}^{m} \sum_{\substack{i=1 \\ i \neq j}}^{m} \int_{\Omega_j} H\left(-\phi_i(x)\right) dx = 0$$

It can be seen that $C \geq 0$, and is equal to zero when there is no evidence of intersection among the regions. Using Lagrange multipliers, the following function can be defined where the third term is added for the smoothness of the regions and μ_j are tuned based on the desired smoothness.

$$E = \sum_{j=1}^{m+1} \int_{\Omega_j} F\left(p_j(x) \times p_{tj}(x)\right) dx$$

$$+\lambda \sum_{j=1}^{m} \sum_{\substack{i=1 \\ i \neq j}}^{m} \int_{\Omega} H\left(-\phi_j(x)\right) H\left(-\phi_i(x)\right) dx \tag{7}$$

$$+\sum_{j=1}^{m} \mu_j \int_{\Omega} \delta\left(\phi_j(x)\right) \left|\nabla \phi_j(x)\right| dx$$

After performing specific mathematical calculations, we arrived at the derivative of the energy function based on the following equation, which can be used to update each sign distance function iteratively.

$$\frac{\partial \phi_j(\hat{x})}{\partial t} = \frac{\partial \phi_j'(\hat{x})}{\partial t} + 2\lambda \left|\nabla \phi_j(\hat{x})\right| \sum_{\substack{l=1 \\ l \neq j}}^{m} H\left(-\phi_l(\hat{x})\right) \tag{8}$$

In each iteration, we can find the Lagrange multiplier and $\dfrac{\partial \phi_j'(\hat{x})}{\partial t}$ based on the following two equations:

$$\lambda = -\frac{\sum_{j=1}^{m} \oint_{\Gamma_j} \left(\frac{\partial \phi_j'(x)}{\partial t} \sum_{\substack{i=1 \\ i \neq j}}^{m} H\left(-\phi_i(x)\right) \right) dx}{2\sum_{j=1}^{m} \oint_{\Gamma_j} \left(\left|\nabla \phi_j(x)\right| \left(\sum_{\substack{i=1 \\ i \neq j}}^{m} H\left(-\phi_i(x)\right) \right)^2 \right) dx} \tag{9}$$

$$\frac{\partial \phi_j'(\hat{x})}{\partial t} = \left|\nabla\left(\phi_j(\hat{x})\right)\right|$$

$$\left\{ F\left(\hat{p}_j(\hat{x})\right) - \prod_{\substack{l=1 \\ l \neq j}}^{m} H\left(\phi_l(\hat{x})\right) \times F\left(\hat{p}_{m+1}(\hat{x})\right) \right.$$

$$-\int_{\{\Omega_k | k \in S_j\}} \frac{F'\left(\hat{p}_j(x)\right)}{\sum_{k \in S_j} |\Omega_k|} \times p_{tj}(x)$$

$$\left(p_j(x) - K\left(I(x) - I(\hat{x})\right) \right) dx \tag{10}$$

$$+\int_{\Omega_{m+1}} \frac{F'\left(\hat{p}_{m+1}(x)\right)}{|\Omega_{m+1}|} \times p_{tm+1}(x)$$

$$\prod_{\substack{l=1 \\ l \neq j}}^{m} H\left(\phi_l(\hat{x})\right) \left(p_{m+1}(x) - K\left(I(x) - I(\hat{x})\right) \right) dx$$

$$\left. +\mu_j div\left(\frac{\phi_j(\hat{x})}{\left|\nabla \phi_j(\hat{x})\right|} \right) \right\}$$

In the next section, we demonstrate that our method is effective for the segmentation of brain structures.

4. Results

To test and evaluate the method, we utilized it for the automatic segmentation and parcellation of brain MRI from the Internet Brain Segmentation Repository (IBSR). It includes T1-weighted volumetric images with slightly different voxel sizes and their corresponding manual segmentation and parcellation. The MR brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at http://www.cma.mgh.harvard.edu/ibsr/. The volumetric images have been positionally normalized into the Talairach orientation (rotation only). In addition, bias field has been corrected for these data. There are two sets of manual segmentations for each subject: manual segmentation of the 34, principle gray and white matter structures of the brain and parcellation results of the 96 structures in the cerebral cortex. In the other words, for each subject, there are 128 manually segmented structures. We have used six structures from the cortical area (left and right frontal pole, frontal orbital cortex, and middle frontal gyrus) as well as left and right cerebral white matter. These structures border one another, a feature that enabled us to establish the effectiveness of the constrained optimization. In addition, our findings suggest that similar left and right structures have the same or similar intensity pdf. To initialize the level set functions, we used the shape averages

based on the sign distance function of the registered label maps. To show the effectiveness of these modifications, we compared a case of optimization that *did not* make use of intensity coupling and constraint to one in which intensity coupling and constraint *were* applied to optimization. Figure 1 depicts segmentation results generated by our method in a series of axial images of a sample dataset. In Table 1, we compare the results of our method to the modified version of the method described in [6-7]. The values presented here, which are based on the Dice coefficient, take into account the joint pmfs of the various tissue types and their respective locations. In every instance, the segmentation performance has improved.

5. Conclusion

We have presented a new segmentation algorithm for the segmentation of brain structures. The method represents a modified version of our previous work described in [5-6]. Specifically, we have modified the pdf estimation of the intensity of the structures to consider the intensity relationships among similar structures. In addition, we have estimated the joint pmf of location and tissue type for each structure; and further, to offset a limitation in the earlier version of this method, we have added a constraint to the optimization process to prevent intersection among different structures. We have successfully demonstrated the effectiveness of this method for the segmentation of selected cortical structures and cerebral white matter. In the next phase of our research, we will apply a modified version of the STAPLE method for segmentation [11], which will include the addition of weights to the raters based on the similarity of the intensity of the atlases and the target image.

6. Acknowledgements

7. References

[1] M. Rouainia and N. Doghmane, "Segmentation of Magnetic Resonance Images : A State of The Art," *2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications, Vols 1-5,* pp. 1116-1122, 2008.
[2] A. Akhondi-Asl and H. Soltanian-Zadeh, "Effect of number of coupled structures on the segmentation of brain structures," *Journal of Signal Processing Systems,* vol. 54, pp. 215-230, 2009.
[3] A. Akhondi-Asl and H. Soltanian-Zadeh, "Nonparametric Entropy-Based Coupled Multi-Shape Medical Image segmentation," in *ISBI*, 2007, pp. 1200-1203.
[4] A. Akhondi-Asl and H. Soltanian-Zadeh, "Constrained optimization of nonparametric entropy-based segmentation of brain structures," in *ISBI*, 2008, pp. 41–44.
[5] A. Akhondi-Asl and H. Soltanian-Zadeh, "Two-stage multishape segmentation of brain structures using image intensity,

tissue type, and location information," *Medical physics,* vol. 37, pp. 4501-4516, 2010.
[6] A. Akhondi-Asl, K. Jafari-Khouzani, K. Elisevich, and H. Soltanian-Zadeh, "Hippocampal volumetry for lateralization of temporal lobe epilepsy: Automated versus manual methods," *NeuroImage,* 2010.
[7] A. Akhondi-Asl, K. Jafari-Khouzani, and H. Soltanian-Zadeh, "Multiple-atlas-based automatic sementation of hippocampus for lateralization in temporal lobe epilepsy," in *ISBI*, 2010, pp. 836-839.
[8] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. Viergever, and B. van Ginneken, "Multi-atlas-based segmentation with local decision fusion—Application to cardiac and aortic segmentation in CT scans," *Medical Imaging, IEEE Transactions on,* vol. 28, pp. 1000-1010, 2009.
[9] E. Parzen, "On the estimation of a probability density function and the mode," *Annals of Mathematical Statistics,* vol. 33, pp. 1065–1076, 1962.
[10] B. M. Silverman, *Density estimation for statistics and data analysis*. London: Chapman and Hall, 1986.
[11] S. Warfield, K. Zou, and W. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *Medical Imaging, IEEE Transactions on,* vol. 23, pp. 903-921, 2004.
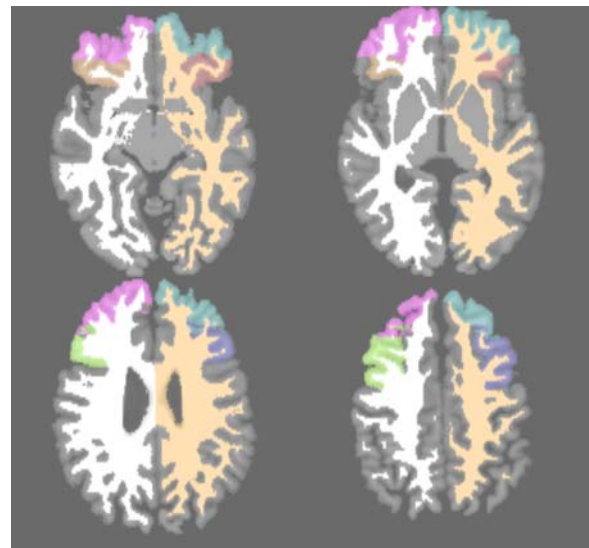
Figure 1. Segmentation results generated by our method in a series of axial images of a sample dataset.

Table 1. Segmentation results of the proposed method and method described in [6-7] for 8 brain structures and the results of the method for the IBSR datasets.

Method	L frrontal pole	R frontal pole	L frontal orbital cortex	R frontal orbital cortex	L cerebral white matter	R cerebral white matter	L middle frontal gyrus	R middle frontal gyrus
Method of [6-7]	0.83 ±0.04	0.82 ±0.04	0.73 ±0.07	0.71 ±0.06	0.91 ±0.02	0.91 ±0.02	0.75 ±0.08	0.74 ±0.05
Our Method	0.85 ±0.04	0.84 ±0.04	0.75 ±0.07	0.73 ±0.06	0.92 ±0.01	0.92 ±0.01	0.76 ±0.08	0.78 ±0.05