MULTIPLE-ATLAS-BASED AUTOMATIC SEGMENTATION OF HIPPOCAMPUS FOR LATERALIZATION IN TEMPORAL LOBE EPILEPSY

Alireza Akhondi-Asl,[1,2,3] , Kourosh Jafari-Khouzani[3], Hamid Soltanian-Zadeh,[1,2,3]

[1]Control and Intelligent Processing Center of Excellence (CIPCE), School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran
[2]School of Cognitive Sciences, Institute for Studies in Theoretical Physics and Mathematics (IPM), Tehran, Iran
[3]Image Analysis Laboratory, Department of Diagnostic Radiology, Henry Ford Health System, Detroit, Michigan, USA

ABSTRACT

We introduce a 3D segmentation framework which uses principal shapes. The probabilistic energy function of the method is defined based on intensity, tissue type, and location information of the structures using a multiple atlas method. For intensity information, nonparametric probability density function is used which considers intensity relation of different structures. To find a local minimum of the energy function, a two-step optimization strategy is used. In the first step, shape parameters are optimized based on the analytic derivatives of the energy function. In the second step, shapes of the structures are fine-tuned using a level set method. The proposed method is shown to be superior to some popular methods in the literature using a dataset of 64 patients with mesial temporal lobe epilepsy. In addition, the method can be used for lateralization with accuracy close to that of manual segmentation.

Index Terms— Image Segmentation, Atlas-Based, Hippocampus, Epilepsy

1. INTRODUCTION

Medical image segmentation is one of the most important issues in computer aided diagnosis and also one of the most challenging topics in image processing [1]. Many segmentation methods have an energy function, a shape model, and an optimization strategy. Each of these parts plays an important role in the design of an accurate segmentation algorithm. Prior knowledge is incorporated in a variety of fashions in order to design more accurate segmentation algorithms [2-5]. Prior knowledge can be used in the energy function, the shape models, or the optimization algorithm. For example, we have used prior knowledge to extract constraints about the energy function parameters in [6]. Another category of segmentation methods use relation between different structures [7-9].

In all of these three methods, the authors used principal component analysis (PCA) to extract shape relation between different structures in the training datasets. In this framework, each structure can be defined based on the weights of the principal shapes. In this paper, we introduce a 3D segmentation framework based on the PCA which considers shape relation between different structures. It uses principal shapes and their weights to define shapes of different structures. The probabilistic energy function of the method is defined based on intensity, tissue type, and location information of the structures. To extract shape models, tissue type, and location information, a multiple atlas method is employed. For intensity information, the nonparametric probability density function (pdf) is used which considers intensity relation of different structures. To find the local minimum of the energy function, a two-step optimization strategy is used. In the first step, shape parameters are optimized based on the analytic first and second derivatives of the energy function with respect to the parameters. In the second step, shapes of the structures are fine-tuned using a level set method.

2. ENERGY FUNCTION

Suppose that we want to simultaneously segment *m* different structures in a skull stripped image $I(x)$ and we have *n* skull stripped training intensity datasets $I_{Tr}^{i}(x)\, i \in \{1,...,n\}$ and their corresponding training label datasets $L_{Tr}^{i}(x)\, i \in \{1,...,n\}$ in which all of the *m* structures are segmented and labeled with labels $l \in \{0,...,m\}$. In each label image, label 0 is assigned to region outside all of these *m* structures (outside region).

For each voxel in the image we define the probability of a pixel belonging to the region Ω_k with $p(x|x \in \Omega_k) = p_k(x) \times p_{fk}(x) \times p_{sk}(x)$ where $k \in \{1,...,m+1\}$ with *m+1* indicating the outside region. In this definition, $p_{fk}(x)$ explains the information about the tissue type (the intensity ranges), $p_{sk}(x)$ considers the location information, and $p_k(x)$ demonstrates the intensity uniformity of the *k*th structure. To compute $p_{sk}(x)$ and $p_{fk}(x)$ in the first step, we use multiple atlas strategy to register $I_{Tr}^{i}(x)\, i \in \{1,...,n\}$ to *I*(*x*) with a non-rigid registration method. The corresponding transformation is named T^i.There are many methods in the literature that use multiple atlases for segmentation [10-12]. We employ a different method in which local information is used for atlas construction. In the previous methods in the literature, local information is applied for each voxel independently, but here we use a local weight for each structure which is not global in the form of [11] and not local in the form of [10, 12]. We apply T^i to the label volumes

$L^i_{Tr}(x)\ i \in \{1,...,n\}$ to get ${}^{T^i}L^i_{Tr}(x)$. In the next step, for each structure we find a region of interest (ROI) based on the registered datasets and name it R_k which shows the ROI of the *k*th structure. In addition, we define $\left(I(x)\right)_{R_k}$ to show the part of the image *I*(*x*) selected by R_k. Then, we use the following equation to find the similarity between structures of the training datasets and the test image:

$$M^i_k = \mathrm{MI}\left(\left({}^{T^i}I^i_{Tr}(x)\right)_{R_k}, \left(I(x)\right)_{R_k}\right) \quad (1)$$

where MI(*A,B*) show the mutual information between the images *A* and *B*.

If $M^{i_1}_k > M^i_k$ for $i = 1...m, i \neq i_1$ based on R_k for structure *k*, it means that the *k*th structure of the dataset i_1 has highest similarity with the *k*th structure of the test image *I*(*x*) based on the mutual information metric. Next, for each label we construct the following image:

$$\bar{\phi}^s_k(x) = \sum_{i=1}^{n} M^i_k \times S\left(\delta_k\left({}^{T^i}L^i_{Tr}(x)\right)\right) \quad k = 1...m \quad (2)$$

where $\delta_k\left(u(x)\right) = \begin{cases} 1 & if \quad u(x) = k \\ 0 & if \quad u(x) \neq k \end{cases}$ and *S* is the sign distance function (SDF) of the binary. This image has the property that a point with a lower value has higher probability of falling inside the structure of interest. Thus, using $\bar{\phi}^s_k(x)$ we can define $p_{sk}(x)$ for the *k*th structure as:

$$p_{sk}(x) = \begin{cases} \ln\left(\left(e - e^{0.5}\right)\left(\frac{\bar{\phi}^s_k(x)}{\min\left(\bar{\phi}^s_k(x)\right)}\right) + e^{0.5}\right) & if \ \frac{\bar{\phi}^s_k(x)}{\min\left(\bar{\phi}^s_k(x)\right)} > \frac{e^e - e^{0.5}}{e - e^{0.5}} \\ \varepsilon & otherwise \end{cases} \quad (3)$$

This function is 1 when $\bar{\phi}^s_k(x)$ takes its minimum and 0.5 when $\bar{\phi}^s_k(x) = 0$. The parameter ε is used to resolve the numerical problems associated with using $\ln\left(p_{sk}(x)\right)$ in the next steps. To calculate $p_{fk}(x)$, we use a probability mass function (pmf) with a new strategy explained next. We apply clustering twice with 3 and 10 clusters. The 3-class segmentation results capture the global intensity information of the tissues in the brain while the 10-class segmentation captures their local or fine intensity information. To use the results, we model the pmf of the tissue type of each structure, i.e., the probabilities of each of the 10-class clusters being present in the structure, by the model shown in Figure 1 whose parameters are estimated using:

$$\begin{cases} L1 = \arg\min\limits_{1 \le i \le 10} abs\left(C_3(2) - C_{10}(i)\right) \\ L2 = \arg\min\limits_{1 \le i \le 10} abs\left(C_3(3) - C_{10}(i)\right) - 2 \end{cases} \quad (4)$$

In this equation, $C_n(k)$ shows the intensity of the center of the *k*-th class (cluster) in the *n*-class clustering. For the region outside all of the *m* structures (region *m*+1), if all of the other structures have the same pdf, we use 1-*p* but if there are different pdf's, we use *p* = 0.5, as described in:

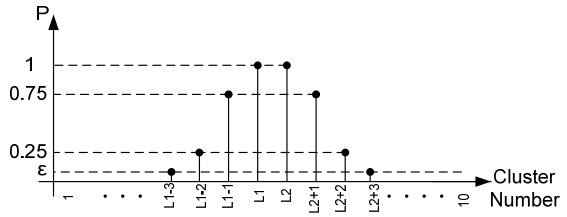

Fig.1. Scaled probability mass function used for integrating the tissue type information in the segmentation process.

$$p_{fm+1}(\mathbf{x}) = \begin{cases} 1 - p_{f1}(\mathbf{x}) & if \ \forall k, k' \neq k \in \{1,...,m\} \quad p_{fk}(\mathbf{x}) = p_{fk'}(\mathbf{x}) \\ 0.5 & otherwise \end{cases} \quad (5)$$

To calculate $p_k(x)$, we use the Parzen window estimator that is defined in the form of $p_k(x) = \frac{1}{|\Omega|}\int_\Omega K\left(I(x) - I(\hat{x})\right)d\hat{x}$ [13]. In this equation, $K(t)$ is a Gaussian kernel with variance σ which is estimated by the method of [14]. In addition, $|\ |$ is used for the cardinality that is the number of members in the set or the region Ω. Thus, we can estimate pdf's for each *m+1* using the following equations:

$$\begin{aligned} p_k(x) &= \frac{1}{\sum_{i \in S_k}|\Omega_i|}\int_{\sum_{i \in S_k}\Omega_i} K\left(I(x) - I(\hat{x})\right)d\hat{x} \quad k = 1,...,m \\ p_{m+1}(x) &= \frac{1}{|\Omega_{m+1}|}\int_{\Omega_{m+1}} K\left(I(x) - I(\hat{x})\right)d\hat{x} \end{aligned} \quad (6)$$

Finally, we write the energy function as:

$$\begin{aligned} J(\Omega_1,\ldots,\Omega_m) &= \sum_{k=1}^{m+1}\int_{\Omega_k} F\left(p(x|x \in \Omega_k)\right)dx \\ &= \sum_{k=1}^{m+1}\int_{\Omega_k} F\left(p_k(x) \times p_{fk}(x) \times p_{sk}(x)\right)dx \end{aligned} \quad (7)$$

This function depends on the *m* regions and the probability of a voxel being within each region. To find the best segmentation, we should find the regions $(\Omega_1,\ldots,\Omega_m)$. It should be noticed Ω_{m+1} is a function of the other regions and is not an independent region. When all of the regions are as uniform as possible, same type regions have more similar intensity distributions, all of the regions are in the correct tissue as much as possible and all of them are in the proper locations, the energy function takes its optimum. If decreasing *F*(*p*) such as –ln(*p*) is used, The optimization will be a minimization problem. To optimize this function, first a shape (region) representation method is needed. In the next section, we use training datasets to find a shape model which considers shape relation of different structures.

3. SHAPE REPRESENTATION

Sign distance function is one of the most effective shape representation methods in the literature [15] which is used in our framework. There are important information about the shapes and relation of the structures that can be used in segmentation. In [9], we used these relations and showed their effectiveness. Here, we use the same idea.

Suppose that we define a sign distance function (SDF) $\phi_k(x)$ which is negative inside region $k \in \{1,2,...,m\}$. We need to find this SDF's for a correct segmentation. We know that there are relations between shapes of different structures and want to use these relationships for segmentation. We have *n* training label

maps $L^i_{Tr}(x)$ with m non-zero labels $k \in \{1,...,m\}$ representing m structures. To extract shape relations between different structures, for each label we register binary images of all training images to the one with highest similarity to $\left(I(x)\right)_{R_k}$ to find the A^i_k's. Each A^i_k is an affine transform with 12 parameters. It should be noticed that for each structure k we have $12\times(n\text{-}1)$ parameters that should be optimized because for the fixed one (for the structure k it is p_k), $\mathrm{A}^{p_k}_k = I$. We use the location differences of the center of masses of the fix and the remaining images for initialization.

In the next step, we compute the SDF every $^{\mathrm{A}^i_k}\left(\delta_k\left(L^i_{Tr}(x)\right)\right)$ and construct $\varphi^i = \left[\left(\varphi^i_1\right)',...,\left(\varphi^i_m\right)'\right]'$ where $\varphi^i_k = \xi\left(S\left(^{\mathrm{A}^i_k}\left(\delta_k\left(L^i_{Tr}(x)\right)\right)\right)\right)$ and X' is the transpose of X. The function $\xi(f)$ changes the 3D image f with the size of $d_1\times d_2\times d_3$ to a vector of size $d_1d_2d_3$. In our approach, the shape priors are modeled using PCA as some basis shapes (ϑ^i), where deformation model of multiple structures is considered to be the combination of this basis priors ($\phi = \bar{\varphi} + \sum_{i=1}^{l} w_i\vartheta^i$ where $\bar{\varphi} = \frac{1}{n}\sum_{i=1}^{n}\varphi^i$). In this model, we are seeking proper values of w_i that minimize our energy function which depend on functions ϕ_j which are parts of $\phi = \left[\left(\xi(\phi_1)\right)',...,\left(\xi(\phi_m)\right)'\right]'$. Thus, for each set of w_i's we can find $\{\phi_1,...,\phi_m\}$ to compute the energy function. In the next section, we summarize the energy function and design the optimization strategy.

4. OPTIMIZATION

To consider the pose variances in the segmentation process, for each structure an individual affine transform is used. These transforms give flexibility to the regions and can be used for the local alignment. In this case, for each region, we add 12 parameters to the problem. Thus, we have $l+12\times m$ parameters which we put in the vector **P**.

$$\min_{\mathbf{P}} J(\mathbf{P}) = \sum_{j=1}^{m}\int_{\Omega} H\left(-\phi_j(x)\right)F\left(p_j(x)\times p_{fj}(x)\times p_{sj}(x)\right)dx + \int_{\Omega}\prod_{k=1}^{m} H\left(\phi_k(x)\right)F\left(p_{m+1}(x)\times p_{fm+1}(x)\times p_{sm+1}(x)\right)dx \quad (8)$$

In this equation, H is Heaviside function. Any optimization problem needs an initialization. For principal shapes, we use $\mathbf{w}=0$ and for the Affine transforms, we set the translation part of the matrices such that the center of mass of $\phi_k(x)<0$ (with $\mathbf{w}=0$) be the same as the center of mass of $\bar{\phi}^s_k(x)<0$. Other parts of the affine matrices are same as the Identity matrix. To optimize the energy function, many methods use first order derivatives and Quasi-Newton or steepest descent optimization strategies [7, 9]. To use this algorithm, we need to compute first order derivatives of the energy function for the parameters in **P**. For the energy function, it can be shown that the derivative of the function J with respect to a parameter ω is:

$$\frac{\partial J(\mathbf{P})}{\partial\omega} = \sum_{j=1}^{m}\oint_{\Gamma_j}\frac{\partial\phi_j(\hat{x})}{\partial\omega}\times\left\{\prod_{\substack{l=1\\ l\neq j}}^{m} H\left(\phi_l(\hat{x})\right)\times F\left(\hat{p}_{m+1}(\hat{x})\right) - F\left(\hat{p}_j(\hat{x})\right) + \frac{1}{\sum_{k\in S_j}|\Omega_k|}\int_{\{\Omega_k|k\in S_j\}} F'\left(\hat{p}_j(x)\right)\times p_{ij}(x)\times\left(p_j(x) - K\left(I(x)-I(\hat{x})\right)\right)dx - \frac{1}{|\Omega_{m+1}|}\int_{\Omega_{m+1}} F'\left(\hat{p}_{m+1}(x)\right)p_{lm+1}(x)\prod_{\substack{l=1\\ l\neq j}}^{m} H\left(\phi_l(\hat{x})\right)\left(p_{m+1}(x) - K\left(I(x)-I(\hat{x})\right)\right)dx\right\}d\hat{x} \quad (9)$$

In this equation, we use abbreviations $p_{tj}(x) = p_{fj}(x)\,p_{sj}(x)$ and $\hat{p}_j(x) = p_j(x)\,p_{tj}(x)$. In addition, Γ_j is the boundary of the jth region. For computing the desired derivatives we need to compute $\frac{\partial\phi_j(\hat{x})}{\partial\omega}$ where we use the same method in [9].

In the next step and after optimization of the parameters of the energy function using constraint optimization algorithm, to capture details of the structures that cannot be extracted from the principal shapes, we remove shape dependency due to the principal shapes and define the following function:

$$\min J(\phi_1,\phi_2,...,\phi_m) = \sum_{j=1}^{m+1}\int_{\Omega_j} F\left(p_j(x)\times p_{fj}(x)\times p_{sj}(x)\right)dx \quad (10)$$

In this function, there are m level set functions that should be optimized with the same constraint as before. Thus we can use the following energy function where the second term is for the smoothness of the shapes.

$$E = \sum_{j=1}^{m+1}\int_{\Omega_j} F\left(p_j(x)\times p_{fj}(x)\times p_{sj}(x)\right)dx + \sum_{j=1}^{m}\mu_j\int_{\Omega}\delta\left(\phi_j(x)\right)\left|\nabla\phi_j(x)\right|dx \quad (11)$$

Using the following equation, we can update each shape iteratively where t shows time.

$$\frac{\partial\phi_j(\hat{x})}{\partial t} = \left|\nabla\left(\phi_j(\hat{x})\right)\right|\left\{F\left(\hat{p}_j(\hat{x})\right) - \prod_{\substack{l=1\\ l\neq j}}^{m} H\left(\phi_l(\hat{x})\right)\times F\left(\hat{p}_{m+1}(\hat{x})\right) - \int_{\{\Omega_k|k\in S_j\}}\frac{F'\left(\hat{p}_j(x)\right)}{\sum_{k\in S_j}|\Omega_k|}p_{ij}(x)\left(p_j(x) - K\left(I(x)-I(\hat{x})\right)\right)dx + \int_{\Omega_{m+1}}\frac{F'\left(\hat{p}_{m+1}(x)\right)}{|\Omega_{m+1}|}p_{lm+1}(x)\prod_{\substack{l=1\\ l\neq j}}^{m} H\left(\phi_l(\hat{x})\right)\left(p_{m+1}(x) - K\left(I(x)-I(\hat{x})\right)\right)dx + \mu_j div\left(\frac{\phi_j(\hat{x})}{\left|\nabla\phi_j(\hat{x})\right|}\right)\right\} \quad (12)$$

5. RESULT

To test and evaluate the proposed method, we have applied it to the MRI of 64 patients with mesial temporal lobe epilepsy (mTLE) acquired on a General Electric 1.5T Signa System (GE Medical Systems, Milwaukee, WI) [16]. The hippocampi were manually segmented by an expert. Comparison of the hippocampus segmentation results generated using our method and the expert segmentation is shown in Figure 2. It is known that the mTLE may cause changes in the size and shape of the hippocampus. Thus correct segmentation of hippocampus in MRI of mTLE patients is a complicated task. The method is also compared with other previous methods (Table 1) which shows the superiority of our method. In addition, we compared lateralization output using right/left volume ratio of automatic methods and the manual

segmentation in Table 2. It can be seen that except HAMMER other methods have almost same standard lateralization percentage.

Table 1. Mean and Standard Deviation (Std) and Minimum and Maximum of Dice similarity, relative absolute difference (RD), Hausdorff distance (H95), average symmetric surface distance (AD), of our, HAMMER, and FreeSurfer methods for the segmentation of hippocampuses in MRI of 64 mTLE cases.

		Dice	RD	H95	AD
Our Method	Mean±Std	0.71±0.08	0.06±0.35	4.17±2.91	0.89±0.48
	Minimum	0.52	-0.46	1.87	0.42
	Maximum	0.81	1.29	14.37	2.65
HAMMER [17]	Mean±Std	0.65±0.08	-0.12±0.19	4.20±2.53	1.04±0.65
	Minimum	0.39	-0.45	2.07	0.51
	Maximum	0.76	0.74	21.08	5.27
FreeSurfer [18]	Mean±Std	0.63±0.13	-0.28±0.30	6.00±4.05	1.62±2.81
	Minimum	0.00	-0.62	2.14	0.46
	Maximum	0.79	1.41	28.74	20.08

Table 2. Comparison of lateralization accuracy of mTLE patients extracted using 4 segmentation methods.

Segmentation Method	Manual	HAMMER	Free Surfer	Our Method
T1 Volume R/L Ratio	0.78	0.73	0.78	0.77

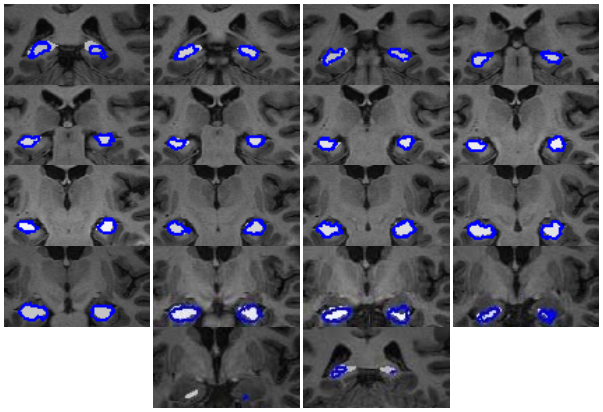

Fig. 2. Comparison of the segmentation results for the hippocampus generated our method (blue) and the expert segmentation (white) in a series of coronal images for a sample temporal lobe epilepsy dataset.

6. CONCLUSION

We have presented a new method for segmentation of the brain subcortical structures using their shape relationships and tissue type and location information using multiple atlas strategy. The energy function used for segmentation is based on the entropy of different structures. Tissue type information is used to improve robustness and accuracy of the segmentation process. In addition, location information is used to further improve the segmentation process. With a powerful automatic initialization of the structures which considers tissue type and location information and the use of the quasi-Newton algorithm, a local minimum of the energy function is found. In the final step, a level-set based optimization is applied. To achieve accurate results, the intensity pdf's are calculated in each iteration of the algorithm and gradients are computed analytically. Experimental results have illustrated superiority of the proposed framework to the other methods in the literature for the segmentation of hippocampus in MRI of mTLE patients. In addition, our proposed method can be used for lateralization with accuracy close to that of the manual segmentation.

7. REFERENCES

[1] D. L. Pham, C. Y. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annual Review of Biomedical Engineering,* vol. 2, pp. 315-+, 2000.

[2] P. K. Yan and A. A. Kassim, "Medical image segmentation using minimal path deformable models with implicit shape priors," *Ieee Transactions on Information Technology in Biomedicine,* vol. 10, pp. 677-684, 2006.

[3] M. Chupin, A. R. Mukuna-Bantumbakulu, D. Hasboun, E. Bardinet, S. Baillet, S. Kinkingnehun, L. Lemieux, B. Dubois, and L. Garnero, "Anatomically constrained region deformation for the automated segmentation of the hippocampus and the amygdala: Method and validation on controls and patients with Alzheimer's disease," *Neuroimage,* vol. 34, pp. 996-1019, 2007.

[4] S. Dambreville, Y. Rathi, and A. Tannenbaum, "A framework for image segmentation using shape models and kernel space shape priors," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 30, pp. 1385-1399, 2008.

[5] C. Tejos, P. Irarrazaval, and A. Cardenas-Blanco, "Simplex Mesh Diffusion Snakes: Integrating 2D and 3D Deformable Models and Statistical Shape Knowledge in a Variational Framework," *International Journal of Computer Vision,* vol. 85, pp. 19-34, 2009.

[6] A. Akhondi-Asl and H. Soltanian-Zadeh, "Constrained optimization of nonparametric entropy-based segmentation of brain structures," in *5th IEEE International Symposium on Biomedical Imaging - From Nano to Macro*, Paris, FRANCE, 2008, pp. 41-44.

[7] A. Tsai, W. Wells, C. Tempany, E. Grimson, and A. Willsky, "Mutual information in coupled multi-shape model for medical image segmentation," *Medical Image Analysis,* vol. 8, pp. 429-445, Dec 2004.

[8] J. Yang, L. H. Staib, and J. S. Duncan, "Neighbor-constrained segmentation with level set based 3-D deformable models," *Ieee Transactions on Medical Imaging,* vol. 23, pp. 940-948, 2004.

[9] A. Akhondi-Asl and H. Soltanian-Zadeh, "Effect of Number of Coupled Structures on the Segmentation of Brain Structures," *Journal of Signal Processing Systems for Signal Image and Video Technology,* vol. 54, pp. 215-230, 2009.

[10] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. A. Viergever, and B. van Ginneken, "Multi-Atlas-Based Segmentation With Local Decision Fusion-Application to Cardiac and Aortic Segmentation in CT Scans," *Ieee Transactions on Medical Imaging,* vol. 28, pp. 1000-1010, 2009.

[11] T. Rohlfing, D. B. Russakoff, and C. R. Maurer, "Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation," *Ieee Transactions on Medical Imaging,* vol. 23, pp. 983-994, 2004.

[12] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de-Solorzano, "Combination strategies in multi-atlas image segmentation: application to brain MR data," *IEEE Trans Med Imaging,* vol. 28, pp. 1266-77, 2009.

[13] E. Parzen, "On the estimation of a probability density function and the mode," *Annals of Mathematical Statistics,* vol. 33, pp. 1065–1076, 1962.

[14] B. M. Silverman, *Density estimation for statistics and data analysis*. London: Chapman and Hall, 1986.

[15] R. Malladi, J. A. Sethian, and B. C. Vemuri, "SHAPE MODELING WITH FRONT PROPAGATION - A LEVEL SET APPROACH," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 17, pp. 158-175, Feb 1995.

[16] http://radiologyresearch.org/eigentool.htm.

[17] http://bric.unc.edu/ideagroup/tools/developed-tools

[18] surfer.nmr.mgh.harvard.edu