NONPARAMETRIC ENTROPY-BASED COUPLED MULTI-SHAPE MEDICAL IMAGE SEGMENTATION

Alireza Akhoundi-Asl[1,2], Hamid Soltamian-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 Lab. Radiology Dept., Henry Ford Health System, Detroit, MI, USA

ABSTRACT

We propose a 3D nonparametric, entropy-based, coupled, multi-shape approach for the segmentation of subcortical brain structures in magnetic resonance images (MRI). Our method uses PCA to capture structures variability. Because of complex relationships of pose and shape of the coupled structures, we only use their shape and size relation. To this end, we apply separate registrations of the structures. For each structure, we consider a similarity transform using seven parameters. In addition, to generate most accurate results, we estimate probability density functions (pdf) iteratively. The proposed method minimizes an entropy-based energy function using quasi-Newton algorithm. To improve the results, we use analytical derivatives. Sample results are given for the segmentation of putamen, illustrating the impact of coupling on the accuracy of the results.

KEYWORDS: Image segmentation, Biomedical Image processing, Magnetic Resonance Imaging, Brain, Entropy.

1. INTRODUCTION

Medical image segmentation is among the most important problems in many medical applications such as visualization, surgical guidance and planning, diagnosis and quantitative measurement [1]. Many important structures in medical images do not present a clear boundary for segmentation and have variations between different subjects. In addition, imaging methods have limitations such as SNR, partial volume effects, and filed innhomogeneities [2].

These problems decrease medical image segmentation accuracy and make it a complicated problem. Several methods have been introduced in the literature to overcome these troubles. However, there is no general-purpose method for segmentation of all structures. Several methods use gradient information to segment structures [3, 4, 5]. Various methods use regional information such as intensity histogram (parametric and nonparametric, offline or online) or variance of an area to solve segmentation problems [6, 7, 8]. Others combine boundary and regional information to reach accurate results [9]. Because of special properties of anatomical structures, using prior information about shapes of different structures is a very popular and effective method for making segmentation process robust [10, 11, 12]. In this type of methods, a registration process is essential.

Another important property of anatomical structures is the relation of neighboring structures and their coupling. Features such as location, size, orientation, and shape can show the relation between different structures. Considering these relations for coupled structures makes the segmentation process more robust and accurate. A number of works have used these relations to improve the segmentation process [13, 14, 15]. One of the most important works is Tsai et al's work that uses PCA to capture variability of different structures for segmentation [13]. We introduce a new nonparametric entropy based method for 3D coupled multi-structure segmentation based of Tsai's et al. work.

2. METHOD

2.1. Registration

To capture structure variability in the training datasets, registration is critical. Tsai et al used multi-shape alignment to capture structure variability [13]. In this way, variability of structures in their position and shape are considered together. However, we observed that shape and position relations of structures are too complex and PCA cannot sufficiently capture their variability, causing the segmentation process to fail. Therefore, we only consider shape relation between coupled structures and for pose variability we use independent transformation for each structure.

To extract shape relation of coupled structures, for each desired structure, we align shapes of structures on a reference shape labeled by an expert physician. To align 3D images, we use similarity transform that includes a scaling, three rotation, and three translation parameters. The energy function used for minimization is the number of misaligned labels and we employ Amoeba optimizer to find best alignment parameters [16]. Next, we explain a model for shape representation.

2.2. Shape Representation

To represent a shape, a variety of methods can be used but the implicit representation of shapes introduced by Leventon is the most popular method [10]. In this method, a distance function is used for shape representation that is zero on the boundary of a shape and in other points is the Euclidian distance from boundary (negative inside, positive outside). Nevertheless, because of the large sizes of 3D images (for example $256\times256\times128$) and the way we align individual structures of different subjects, we do not need to use the whole image for implicit representation. For each of the structures, we use a rectangular area (say $N_x\times N_y\times N_z$) that covers the desired structures in all of the aligned training datasets. Comparing the results shows that the results are identical in these two methods but less memory is needed for our proposed method (especially for computing principal components).

After extraction of distance functions of m desired structures for n different datasets (ψ_i^k shows the distance function of the kth structure of the ith dataset), we subtract the mean distance function of each structure computed by averaging of the training datasets ($\bar{\Phi}^k$), from each of the *n* signed distance functions. These $n\times m$ mean-offset functions ($\tilde{\psi}_i^k$) are used to show variability of different structures in the training datasets. We collect n column vectors of size $m\times N_x\times N_y\times N_z$ and use them to extract n eigenvectors of each of the m structures to show their variability (Φ_i^k).

Suppose that we use $q\le n$ eigenvectors for shape representation. In addition, to consider pose differences, pose parameters are added to each structure's model separately. Finally, for each shape, we have implicit descriptor in the following form.

$$\Phi^k\left[\mathbf{w},\mathbf{p}^k\right](x,y,z)=\bar{\Phi}^k\left(\tilde{x}_k,\tilde{y}_k,\tilde{z}_k\right)+\sum_{i=1}^{q}w_i\Phi_i^k\left(\tilde{x}_k,\tilde{y}_k,\tilde{z}_k\right) \quad (1)$$

For transformation of points, we use:

$$\begin{bmatrix}\tilde{x}_k\\ \tilde{y}_k\\ \tilde{z}_k\\ 1\end{bmatrix}=M\left(\mathbf{p}^k\right)\begin{bmatrix}x\\ y\\ z\\ 1\end{bmatrix} \quad (2)$$

$$M\left(\mathbf{p}^k\right)=T\left(\mathbf{t}_k+\mathbf{c}_k\right)R\left(\boldsymbol{\theta}_k\right)S\left(s_k\right)T\left(-\mathbf{c}_k\right)$$

where T, R and S are transform, rotation, and scaling matrices, respectively, $\mathbf{t}$ is a vector of translation in 3D and $\mathbf{c}$ is the center of rotation used for scaling and rotation and $R\left(\boldsymbol{\theta}\right)=R_x\left(\theta_x\right)R_y\left(\theta_y\right)R_z\left(\theta_z\right)$ is multiplication of rotations in Cartesian coordinates. The center of transform is fixed and is computed using the mean of the centers of masses of the training datasets.

In this manner, we have the ability to change each structure separately while using their shape relations for coupling. We can use the same parameters for all of the structures although separate parameters generate more realistic results. Before the transformation of shapes constructed using the eigenvectors, their sizes are set to the original image size and their centers of the mass are used to place the rectangular area. In the next subsection, we define our proposed entropy-based segmentation model.

2.3. Entropy-based nonparametric segmentation model

The proposed entropy-based method calssifies image voxels by minimizing a weighted sum of the conditional differential entropies. Based on the entropy of regions $\{\Omega_k,\quad k=1\cdots m+1\}$, the objective function $J\left(\Omega_1,\ldots,\Omega_{m+1}\right)=|\Omega|\sum_{j=1}^{m+1}P_{\Omega_j}\hat{H}\left(\Omega_j\right)$ is used for minimization. The entropy is estimated using the general equation $\hat{H}\left(\Omega_k\right)=\frac{-1}{|\Omega_k|}\int_{\Omega_k}\ln\hat{p}\left(I\left(\mathbf{x}\right),\Omega_k\right)d\mathbf{x}$ [13] and we also have $P_{\Omega_k}=\frac{|\Omega_k|}{|\Omega|}$.

In the entropy estimation, $\hat{p}\left(I\left(\mathbf{x}\right),\Omega_k\right)$ is the approximate probability density function (pdf) in region k of the desired 3D image I. Many researchers used off-line density estimation for pdfs. However, we observed dissimilar behaviors in different datasets and concluded that using an off-line pdf is inappropriate. We estimate pdfs using the parzen window method with the Gaussian kernel as $\hat{p}\left(I\left(\mathbf{x}\right),\Omega\right)=\frac{1}{|\Omega|}\int_{\Omega}K\left(I\left(\mathbf{x}\right)-I\left(\hat{\mathbf{x}}\right)\right)d\hat{\mathbf{x}}$. In this equation, K is the Gaussian kernel with a sigma as tuning parameter [17]. Finally, we write the objective function as:

$$J\left(\Omega_1,\ldots,\Omega_{m+1}\right)=J\left(P\right)=\sum_{j=1}^{m+1}-\int_{\Omega_j}\ln\hat{p}\left(I\left(\mathbf{x}\right),\Omega_j\right)d\mathbf{x} \quad (3)$$

where P is vector of $m\times7+q$ parameters. To minimize this objective function, we use quasi-Newton algorithm with BFGS method for Hessian matrix estimation [18]. In the literature, methods such as steepest descent are more popular but we obtained more accurate results with the quasi-Newton algorithm [13]. In this algorithm, we need gradient with respect to the parameters. We can estimate gradients using numerical methods but analytical computation is more robust and generates more accurate results. To compute the gradients analytically, we use Heaviside function $U\left(\Phi^k\right)=\begin{cases}1 & if\ \Phi^k\ge0\\ 0 & if\ \Phi^k<0\end{cases}$. There are two types of parameters $\mathbf{w}$ and $\mathbf{p}^k$ and for the ith component of $\nabla_{\mathbf{w}}$ and $\nabla_{\mathbf{p}^k}$, we compute derivatives as follows.

$$\nabla w_i J = \sum_{j=1}^{m} \left\{ \iint_{\Gamma_j} \nabla w_i \Phi^j(\mathbf{x}) \left(\hat{p}_j(\mathbf{x}) \ln \hat{p}_j(\mathbf{x}) - \hat{p}_{m+1}(\mathbf{x}) \ln \hat{p}_{m+1}(\mathbf{x}) \right) d\mathbf{x} \right.$$
$$+ \frac{1}{|\Omega_j|} \int_{\Omega_j} \left(\frac{1}{\hat{p}_j(\mathbf{x})} \iint_{\Gamma_j} \nabla w_i \Phi^j(\hat{\mathbf{x}}) K\left(I(\mathbf{x}-\hat{\mathbf{x}})\right) d\hat{\mathbf{x}} \right) d\mathbf{x} \quad (4)$$
$$\left. - \frac{1}{|\Omega_{m+1}|} \int_{\Omega_{m+1}} \left(\frac{1}{\hat{p}_{m+1}(\mathbf{x})} \iint_{\Gamma_j} \nabla w_i \Phi^j(\hat{\mathbf{x}}) K\left(I(\mathbf{x}-\hat{\mathbf{x}})\right) d\hat{\mathbf{x}} \right) d\mathbf{x} \right\}$$

$$\nabla p_i^k J = \iint_{\Gamma_j} \nabla p_i^k \Phi^k(\mathbf{x}) \left(\hat{p}_k(\mathbf{x}) \ln \hat{p}_k(\mathbf{x}) - \hat{p}_{m+1}(\mathbf{x}) \ln \hat{p}_{m+1}(\mathbf{x}) \right) d\mathbf{x}$$
$$+ \frac{1}{|\Omega_k|} \int_{\Omega_k} \left(\frac{1}{\hat{p}_k(\mathbf{x})} \iint_{\Gamma_k} \nabla p_i^k \Phi^k(\hat{\mathbf{x}}) K\left(I(\mathbf{x}-\hat{\mathbf{x}})\right) d\hat{\mathbf{x}} \right) d\mathbf{x} \quad (5)$$
$$- \frac{1}{|\Omega_{m+1}|} \int_{\Omega_{m+1}} \left(\frac{1}{\hat{p}_{m+1}(\mathbf{x})} \iint_{\Gamma_k} \nabla p_i^k \Phi^k(\hat{\mathbf{x}}) K\left(I(\mathbf{x}-\hat{\mathbf{x}})\right) d\hat{\mathbf{x}} \right) d\mathbf{x}$$

In these equations, for simplicity we used $\hat{p}_k(\mathbf{x}) = \hat{p}\left(I(\mathbf{x}), \Omega_k\right)$. For $\nabla p_i^k \Phi^k(\mathbf{x})$ and $\nabla w_i \Phi^k(\mathbf{x})$ we have:

$$\nabla w_i \Phi^k(\mathbf{x}) = \Phi_i^k$$
$$\nabla p_i^k \Phi^k(\mathbf{x}) = \begin{bmatrix} \frac{\partial \Phi^k(\tilde{x}_k, \tilde{y}_k, \tilde{z}_k)}{\partial \tilde{x}_k} \\ \frac{\partial \Phi^k(\tilde{x}_k, \tilde{y}_k, \tilde{z}_k)}{\partial \tilde{y}_k} \\ \frac{\partial \Phi^k(\tilde{x}_k, \tilde{y}_k, \tilde{z}_k)}{\partial \tilde{z}_k} \\ 0 \end{bmatrix}^T \times \frac{\partial M(\mathbf{p}^k)}{\partial p_i^k} \quad (6)$$

where $\frac{\partial M(\mathbf{p}^k)}{\partial p_i^k}$ is computed using derivatives of each matrix element. Finally, we start with initial parameters and update parameters with the following equations:

$$\begin{aligned} s_k &= P_{k+1} - P_k \\ q_k &= \nabla J(P_{k+1}) - \nabla J(P_k) \\ H_{k+1} &= H_k + \frac{q_k q_k^T}{q_k^T s_k} - \frac{H_k^T s_k^T s_k H_k}{s_k^T H_k s_k} \\ d_k &= -H_k^{-1} \cdot \nabla J(P_k) \\ P_{k+1} &= P_k + \alpha d_k \end{aligned} \quad (7)$$

where α is a scalar step size parameter.

3. RESULTS

In this section, we will show some results of applying the proposed methods to real MRI data. Real MRI data obtained from the Internet Brain Segmentation Repository (IBSR) is used for training and test of the system [19]. Six datasets of $256 \times 256 \times 128$ voxels are used for training and two datasets of same voxel size is considered for test. Pixel size of datasets is not identical and this makes the test process more reliable. Each test dataset is registered using normalized mutual information (NMI) and the same transform is used for the alignment of the training datasets.

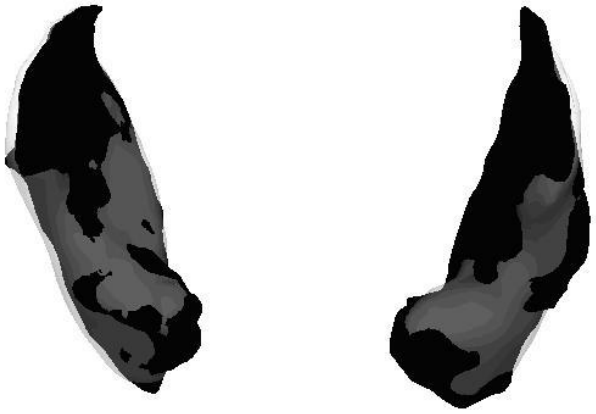

Figure 1. Left and right putamen segmentation result (white with low opacity) over physician segmented (black).

Table 1. Dice coefficient for left and right puatmen for two datasets with and without coupled information.

	Left alone	Right alone	Left coupled	Right coupled
1	82	84	89	92
2	72	73	84	84

On the base of the structures of interest, a region of interest (ROI) is chosen to make the algorithm faster. In addition, pdf estimation of background (region out of structures) is more realistic and makes segmentation process more accurate. For initialization of structures, we registered all training datasets on the reference without considering any special structure. Then average of center of mass of structures in all of the registered training datasets is used as the position of center of the mass of the initial shape of structures. Because of different scaling on parameters, a normalization of parameters is needed. For example, we compute rotation with radian and translation by pixel. Therefore, we have to choose a smaller unit for rotation. Choosing units for scaling and principal components is very important in the same manner. In order to evaluate the results, we use the Dice coefficient [20] computed as:

$$C_D = \frac{\left(2|A \cap B|\right)}{\left(|A| + |B|\right)} \quad (8)$$

In this equation, A and B are the sets of voxels of volumes to be compared. Because of their unclear edges, structures such putamen is hard to segment. Figure 1 shows the results of segmentation of the left and right putamen (with lower opacities) and physician labeled structures. Table 1 shows Dice coefficient

computed for the 2 test datasets with and without considering coupling information. It is obvious that using coupled information makes segmentation more accurate. For two structures (left and right putamen) on a 3 GHz windows XP workstation (1 GB RAM), computation time is about 6 minutes. For visualization of structures we used 3D Slicer [21] and for registration of training datasets we used ITK library [22]. Other parts of the works are done with MATLAB workstation [23].

4. CONCLUSION

We presented a new method for segmentation of brain subcortical structures using their shapes relation. Energy function used for segmentation takes into account entropy of different shapes. With an automatic initialization of structures and using quasi-Newton algorithm, local minimum of the energy function is found.

To achieve more accurate results, probability density functions are calculated iteratively and gradients are computed analytically. Experimental results illustrate robustness and quality of the results generated by the proposed framework.

5. REFERENCES

[1] J.S. Suri, S. K. Setarehdan, and S. Singh, *Advanced Algorithmic Approaches to Medical Image Segmentation, State of the Art Applications in Cardiology, Neurology, Mammography and Pathology,* Springer-Verlag, February 2002.

[2] A. Macovski, "Noise in MRI," *Magnetic Resonace in Medicine*, Vol. 36, No. 3, pp. 494–497, 1996.

[3] E. D. Angelini, Y. Jin, and A. F. Laine, *State-of-the-Art of Levelset Methods in Segmentation and Registration of Medical Imaging Modalities,* The Handbook of Medical Image Analysis - Volume III: Registration Models, Kluwer Academic/Plenum Publishers, New York,, 2005.

[4] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic Active Contours," *International Journal of Computer Vision,* Vol. 22, No. 1, pp. 61–79, 1997.

[5] M. Jacob, T. Blu, and M. Unser , "Efficient energies and algorithms for parametric snakes," *IEEE Transactions on Image Processing,* Vol. 13, No. 9, pp. 1231-1244, 2004.

[6] X. Bresson, *Image Segmentation with Variational Active Contours,* PhD Thesis, University of Lausanne, July 2005.

[7] M. W. Woolrich, and T. E. Behrens, " Variational Bayes Inference of Spatial Mixture Models for Segmentation," IEEE Transaction On Medical Imaging, Vol. 25, No. 10, pp. 1380-1391, October 2006.

[8] A. Huang, G. M. Nielson, A Razdan, G. E. Farin, D. P. Baluch, and D. G. Capco, " Thin Structure Segmentation and Visualization in Three-Dimensional Biomedical Images: A Shape-Based Approach," IEEE Transaction On Visualization and Computer Graphics, Vol. 12, No. 1, pp. 93-102, January/February 2006.

[9] N. A. Paragios, "Geodesic Active Regions and Level Set Methods," *PhD Thesis, University of Nice,* January 2000.

[10] M. E. Leventon, W. E. L. Grimson, and O. Faugeras, "Statistical Shape Influence in Geodesic Active Contours," *IEEE Internatinal Conference on Computer Vision and Pattern Recognition,* Vol. 1, pp. 1316-1323, 2000.

[11] S. Jehan-Besson, A. Herbulot, M. Barlaud, and G. Aubert "Shape Gradient for Image and Video Segmentation," *Mathematical Models in Computer Vision: The Handbook, Springer,* 2005.

[12] K. M. Pohl, J. Fisher, R. Kikinis, W. E. L. Grimson, and W. M. Wells, "Shape Based Segmentation of Anatomical Structures in Magnetic Resonance Images," *International Conference on Computer Vision,* Vol. 3765, pp. 489-498, 2005.

[13] 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, No. 4, pp. 429-445, December 2004.

[14] J. Yang , L. H. Staib, and J. S. Duncan, "Neighbor-Constrained Segmentation with Level Set Based 3D Deformable Models," *IEEE Transactions on Medical Imaging,* Vol. 23, No. 8, pp. 940-948, August 2004.

[15] A. Litvin, and W.C. Karl, "Coupled Shape Distribution-Based Segmentation of Multiple Objects," *Boston University, Boston, USA, Tech. Rep. ECE-2005-01,* March 2005.

[16] J. A. Nelder, and R. Mead, "A Simplex Method for Function Minimization". *The Computer Journal*, Vol. 7, pp. 306–313, 1965.

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

[18] K. J. Bathe, and A. P. Cimento, "Some Practical Procedures for the Solution of Nonlinear Finite Element Equations," *Computer Methods in Applied Mechanics and Engineering,* Vol. 22, pp. 59-85, 1980.

[19] http://www.cma.mgh.harvard.edu/ibsr/

[20] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology,* Vol. 26, pp. 297–302, 1945.

[21] http://www.slicer.org

[22] http://www.itk.org

[23] http://www.mathworks.com