CONSTRAINED OPTIMIZATION OF NONPARAMETRIC ENTROPY-BASED SEGMENTATION OF BRAIN STRUCTURES

Alireza Akhondi Asl[1,2] and Hamid Soltanian-Zadeh, Senior Member, IEEE[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, Michigan, USA

ABSTRACT

We propose a constrained, three-dimensional, nonparametric, entropy-based, coupled, multi-shape approach to segment subcortical brain structures from magnetic resonance images (MRI). The proposed method uses PCA to develop shape models that capture structural variability. It integrates geometrical relationship between different structures into the algorithm by coupling them (limiting their independent deformations). On the other hand, to allow variations among coupled structures, it registers each structure separately when building the shape models. It defines an entropy-based energy function, which is minimized using quasi-Newton algorithm. To this end, probability density functions (pdf) are estimated iteratively using nonparametric Parzen window method. In the optimization algorithm, constraints are used to improve segmentation quality. These constraints are extracted from training data. Sample results are given for the segmentation of caudate, hippocampus, and putamen, illustrating highly superior performance of the proposed method compared to the most similar methods in the literature.

Index Terms— Image segmentation, brain structures, shape modeling, entropy, nonparametric, constrained optimization.

1. INTRODUCTION

Medical image segmentation is the most important step in visualization, surgical guidance and planning, diagnosis and quantitative measurement [1]. However, 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 low signal-to-noise ratio (SNR), partial volume effects, and field inhomogeneities [2]. An exciting approach for the segmentation is based on the optimization of the energy function using partial differential equations. Kass et al [3] introduced the first work in this category which has been improved by others in recent years. For the shape representation, parametric active contours and geometric active contours have been used. In the definition of the energy function, earlier methods use the boundary information for the structures of interest [4]-[5]. Later methods use regional information such as intensity histogram (parametric and nonparametric, offline or online) or variance of an area [6]-[7]. Recent methods benefit from a priori knowledge about the structures of interest. This makes the segmentation process robust to the imperfect image conditions [8]-[9]. For the methods developed based on the a priori information, a registration process is essential to integrate the prior model into the segmentation process.

In addition, the anatomical structures in the brain are related to the neighboring structures through their location, size, orientation, and shape. An integration of these relations into the segmentation process improves accuracy and robustness as shown in [10]-[11]. In this paper, we extend our previous 3D coupled shape segmentation method by adding linear constraints on the optimization process [12]. In some special cases, segmentation failed because of special conditions of the image datasets. In these cases, final shape parameters are not close to typical values. These typical values can be extracted from training datasets and can be used as important and powerful constraints in the optimization process. In the introduced method, we find some bounds for the shape parameters from the training datasets and use them as the constraints. To solve the resulting constrained optimization problem, we use a sequential quadratic programming (SQP) method. This method solves a quadratic programming (QP) sub problem in each iteration using an active set strategy. As in our previous method, we separate shape and pose variability. This is because inclusion of both of the pose and shape variability in a PCA based model generates huge variability and limits the model's benefits for segmentation. We align each of the shape classes in the training datasets individually to extract co-variations between shape classes without considering their spatial locations (poses). Furthermore, we use completely online pdf estimation, updated during iterations of the optimization algorithm. We use pre-register test datasets for segmentation.

2. SHAPE MODEL

In many segmentation methods, a shape model is used where richer models generate more accurate results. Several shape representation methods are used in the literature that

are based on the kernel integrals [13] and medial shapes [14]. More powerful methods for shape representation are based on distance function, implicit representation, and relationships among different shapes, including pose, orientation, and other geometrical relations [15],[16]. In this paper, we use shape relation and individual transformations for each structure for segmentation. To extract shape relation, we apply principal component analysis (PCA) on the 3D training datasets. To reach accurate results, we individually align each structure of interest in the training datasets.

2.1. Alignment

In knowledge based segmentation methods that use shape variation, alignment is a critical step. It has four important constituents: a transformation matrix, a metric, an interpolator, and an optimizer. We use similarity transform with 7 parameters (1 for scaling, 3 for rotations, and 3 for translations). For the metric, since we use a labeled 3D image that represents the desired structure, the cardinality metric is used. For interpolation of binary images, we use a nearest neighbor interpolator. For optimization, we use the Nelder–Mead (amoeba) method which is a popular direct search method for minimizing real functions with our desired property [17]. Using the above methods, we extract shape variability of the desired structures for model construction as explained in the following section.

2.2. Implicit Parametric Shape Representation

We use a distance map for shape representation that is zero on the boundary of a shape and in other points is the Euclidian distance from the boundary (negative inside, positive outside). After extraction of the distance maps of m desired structures for n different training datasets (ψ_i^k shows the distance map of the kth structure of the ith dataset), we subtract the mean distance map of each structure, computed by averaging of the training datasets ($\bar{\Phi}^k$), from each of the n signed distance maps to remove similar parts in different shapes and show them with $\tilde{\psi}_i^k$. We use these $n \times m$ maps to show variability of different structures in the training dataset. We collect n column vectors of size $m \times N_x \times N_y \times N_z$ and use them to extract n eigenvectors for each of the m structures and show their variability (Φ_i^k). To allow limited, robust shape variability, we use $q \le n$ eigenvectors to represent each shape. In addition, to consider pose differences, we add 7 pose parameters (for local alignment of the structures) to the shape parameters of each structure. Finally, for each structure, we may write

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

where $\mathbf{w}$ is the vector of eigenvectors multipliers and $\mathbf{p}^k$ is the vector containing 7 transformation parameters for the alignment of the kth structure. In this manner, each structure's pose may change while shape classes co-variations are used for coupling. In the next section, we present our proposed entropy-based segmentation method using the shape model described above.

3. SEGMENTATION OF STRUCTURES

After construction of the shape model and shape classes co-variations, an energy function is defined for the segmentation process. In this section, we explain our energy model and optimization method.

3.1. Energy Model

The proposed entropy-based method classifies image voxels to distinguish regions by minimizing a weighted sum of the conditional differential entropies of different structures. To segment m coupled structures with closed boundaries, there are m regions for these structures. We set the area outside of the m structures as $m+1$ and use this notation throughout the paper. Based on the entropy of these regions $\{\Omega_k,\ k=1\cdots m+1\}$, the energy function is defined as

$J(\Omega_1,\ldots,\Omega_{m+1}) = |\Omega|\sum_{j=1}^{m+1} P_{\Omega_j}\hat{H}(\Omega_j)$ where $|\Omega|$ represents the cardinality of a set Ω (number of pixels). When all of the regions are as uniform as possible, the energy function is at its minimum. Nevertheless, there is the important problem of estimating conditional differential entropies.

We estimate the entropy of the kth structure using $\hat{H}(\Omega_k)=\frac{-1}{|\Omega_k|}\int_{\Omega_k}\ln\hat{p}(I(\mathbf{x}),\Omega_k)d\mathbf{x}$. In addition, we use $P_{\Omega_k}=\frac{|\Omega_k|}{|\Omega|}$ based on several previous publications [11]. In entropy estimation, $\hat{p}(I(\mathbf{x}),\Omega_k)$ is the approximate probability density function (pdf) in region k of the 3D image I. Many researchers estimated pdf's off-line. However, we observed dissimilar dynamic ranges of image intensities in different datasets and concluded that an off-line pdf is suboptimal; we use on-line estimation.

We estimate pdf's using the Parzen window method [18], as $\hat{p}(I(\mathbf{x}),\Omega)=\frac{1}{|\Omega|}\int_{\Omega}K(I(\mathbf{x})-I(\hat{\mathbf{x}}))d\hat{\mathbf{x}}$. In this equation, K is the Gaussian kernel with a standard deviation (sigma) as its tuning parameter, which sets the resolution of the pdf estimation process. Choosing low values make pdf estimation sensitive to noise and high values remove useful details from the estimated pdf. In the literature, values between 1and 3 are used. Finally, we write the energy function as: $J(\Omega_1,\ldots,\Omega_{m+1})=J(P)=\sum_{j=1}^{m+1}-\int_{\Omega_j}\ln\hat{p}(I(\mathbf{x}),\Omega_j)d\mathbf{x}$.

where P is vector of $m\times 7+q$ parameters (because each one of the local alignments have 7 parameters).

3.2. Constrained Optimization

In our previous work, we used Quasi-Newton algorithm with BFGS method for Hessian matrix estimation [12]. In this paper, we use constrained optimization to transform the segmentation problem into an easier sub problem that can then be solved by an iterative process. The Kuhn-Tucker (KT) equations are the most popular method to solve constraint methods [19]. If the objective function and constraints are convex, then KT equations are both necessary and sufficient for a global solution point. We can state the general constrained problem as:

$$\underset{x}{\text{minimize}}\ J(P)\ \ subject\ to\ \begin{cases} C_i(P)=0 & i=1,\cdots,m \\ C_i(P)\le 0 & i=m+1,\cdots,N \end{cases} \quad (2)$$

where C_i shows the *i*th constraint, *m* of this constraints are equality and *N-m+1* of them are non-equality. The Kuhn-Tucker equations can be stated as:

$$\begin{aligned} &\nabla J(P^*)+\sum_{i=1}^{m}\lambda_i^*.\nabla C_i(P^*)=0 \\ &\lambda_i^* C_i(P^*)=0 \quad i=1,\cdots,m \ and \ \ C_i(P)=0 \quad i=1,\cdots,m \\ &\lambda_i^*\ge 0 \quad i=m+1,\cdots,N \quad and \ \ C_i(P)\le 0 \quad i=m+1,\cdots,N \end{aligned} \quad (3)$$

where P^* is a global solution vector. Constraints used in our problem are lower and upper bounds that can be shown with non-equality linear constraints. Equation (3) is quadratic programming sub-problem. To solve it, we used a sequential quadratic programming where in each major iteration, the Hessian matrix is updated using BFGS method as follows.

$$\begin{aligned} q_k &= \nabla J(P_{k+1})+\sum_{i=1}^{m}\lambda_i.\nabla C_i(P_{k+1})-\left(\nabla J(P_k)-\sum_{i=1}^{m}\lambda_i.\nabla C_i(P_k)\right) \\ H_{k+1} &= H_k+\frac{q_k q_k^T}{q_k^T s_k}-\frac{H_k^T s_k^T s_k H_k}{s_k^T H_k s_k} \end{aligned} \quad (4)$$

Where $s_k = P_{k+1}-P_k$. Then, the following problem is solved using an active set strategy method described in [20].

$$\begin{aligned} &\underset{x}{\text{minimize}}\ \ q(x)=\frac{1}{2}x^T H_k x+\nabla J(P_k)^T x \\ &\nabla C_i(P_k)^T x+C_i(P_k)=0 \quad i=1,\cdots,m \\ &\nabla C_i(P_k)^T x+C_i(P_k)\le 0 \quad i=m+1,\cdots,N \end{aligned} \quad (5)$$

Finally, we use a line search to find $P_{k+1}=P_k+\alpha d_k$. The gradients can be estimated using numerical methods but analytical computation is more robust and generates results that are more accurate. There are two types of parameters $\mathbf{w}$ and $\mathbf{p}^k$ and for the *i*th component of $\nabla_{\mathbf{w}}$ and $\nabla_{\mathbf{p}^k}$. We compute the derivatives as follows.

$$\begin{aligned} \nabla w_i J &= \sum_{j=1}^{m}\Bigg\{\oint_{\Gamma_j}\nabla w_i\Phi^j(\mathbf{x})\left(\ln\hat{p}_j(\mathbf{x})-\ln\hat{p}_{m+1}(\mathbf{x})\right)d\mathbf{x} \\ &+\frac{1}{|\Omega_j|}\int_{\Omega_j}\left(\frac{1}{\hat{p}_j(\mathbf{x})}\oint_{\Gamma_j}\nabla w_i\Phi^j(\hat{\mathbf{x}})K\left(I(\mathbf{x}-\hat{\mathbf{x}})\right)d\hat{\mathbf{x}}\right)d\mathbf{x} \\ &-\frac{1}{|\Omega_{m+1}|}\int_{\Omega_{m+1}}\left(\frac{1}{\hat{p}_{m+1}(\mathbf{x})}\oint_{\Gamma_j}\nabla w_i\Phi^j(\hat{\mathbf{x}})K\left(I(\mathbf{x}-\hat{\mathbf{x}})\right)d\hat{\mathbf{x}}\right)d\mathbf{x}\Bigg\} \end{aligned} \quad (6)$$

$$\begin{aligned} \nabla p_i^k J &= \oint_{\Gamma_k}\nabla p_i^k\Phi^k(\mathbf{x})\left(\ln\hat{p}_k(\mathbf{x})-\ln\hat{p}_{m+1}(\mathbf{x})\right)d\mathbf{x} \\ &+\frac{1}{|\Omega_k|}\int_{\Omega_k}\left(\frac{1}{\hat{p}_k(\mathbf{x})}\oint_{\Gamma_k}\nabla p_i^k\Phi^k(\hat{\mathbf{x}})K\left(I(\mathbf{x}-\hat{\mathbf{x}})\right)d\hat{\mathbf{x}}\right)d\mathbf{x} \\ &-\frac{1}{|\Omega_{m+1}|}\int_{\Omega_{m+1}}\left(\frac{1}{\hat{p}_{m+1}(\mathbf{x})}\oint_{\Gamma_k}\nabla p_i^k\Phi^k(\hat{\mathbf{x}})K\left(I(\mathbf{x}-\hat{\mathbf{x}})\right)d\hat{\mathbf{x}}\right)d\mathbf{x} \end{aligned} \quad (7)$$

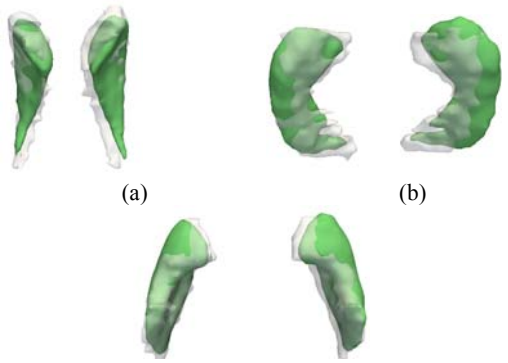
4. EXPERIMENTAL RESULTS

In this section, we present the results of applying the proposed method to the real MRI data. This data is obtained from the Internet Brain Segmentation Repository (IBSR) and used for the training and testing of the proposed method [21]. Datasets are T1-weighted volumetric images with different pixel sizes. Expert physicians have segmented 43 structures for 18 datasets. We define the ROI as 1.05 times of the smallest cube that covers all of the desired structures in different training datasets. To evaluate the results, we use the Dice coefficient [22]. In addition, as an alternative evaluation measure, we use the segment Hausdorff distance [23]. For optimization and extraction of the principal shapes, we use MATLAB [20]. All programs are run on a 3.2 GHz Windows XP workstation with 2 GB RAM. We use ten datasets (randomly chosen) for training and the remaining eight datasets for testing of the proposed segmentation algorithm. We segment each of the left and right structures individually using the principal shapes extracted from the training datasets. We also segment the left and right structures or multiple structures by the proposed coupling method. To evaluate the proposed method in comparison to the most similar work in the literature, we implement and apply the method of Tsai et al [11] and our previous method [12] which does not apply any constraints. Evaluation of the results using dice coefficient and segment Hausdorff distance for the caudate, hippocampus, and putamen for eight test datasets are shown in Table 1. we can see that coupling of the left and the right structures gives better segmentation result and are more accurate with respect to the segmentation in alone manner with both of the Dice coefficient and segment Hausdorff distance measures in almost all cases. Our methods are more accurate than the Tsai et al's method in almost all cases, while the constrained method is the most accurate method in overall. In Figure 1, sample 3D segmentation results for the three structures are shown.

5. CONCLUSION

We presented a new method for the segmentation of brain subcortical structures using their shapes relation extracted using PCA. We used separate shape class registrations to extract each class variation independent of the other shape classes. To this end, we removed variations due to the shape poses by considering each shape representation locally. In addition, we allowed independent placements of the shape models to allow more flexibility. The energy function used for segmentation took into account entropy of different

shapes. With an automatic initialization of the structures and use of the SQP method, a local minimum of the energy function was found. To achieve accurate results, probability density functions were calculated in each iteration and gradients were computed analytically. Our method has low sensitivity to the parameters and is robust. Experimental results illustrate robustness and quality of the results generated by the proposed framework**.**



(a) (b)

(c)

Fig 1. (a). Final caudates, hippocampus, and putamen segmented with constraints in the coupled manner (green) compared to the ideal segmentation (low opacity).

Table 1. Mean Values of Dice coefficients and segment Hausdorff distances for the segmentation obtained by the Tsai et al [11], our unconstrained (normal) [12], and constrained method for the selected structures without and with the coupling information.

		Dice Coefficient			Segment Hausdorff Distance (mm)		
		Tsai et al [16]	Normal [12]	Constrained	Tsai et al [16]	Normal [12]	Constrained
Caudate	Alone	0.55	0.58	0.72	5.97	5.84	5.53
	Coupled	0.65	0.74	0.76	5.82	5.46	5.35
Hippocampus	Alone	0.59	0.61	0.66	6.29	6.30	6.20
	Coupled	0.59	0.62	0.70	6.25	6.17	6.00
Putamen	Alone	0.76	0.81	0.81	5.17	5.02	5.01
	Coupled	0.78	0.82	0.82	5.06	4.95	4.91

6. 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 Resonance in Medicine, Vol. 36, No. 3, pp. 494–497, 1996.

[3] M. Kass, A. Witkin, D. Terzopoulos, "Snakes: Active Contour Models," *First International Conference on Computer Vision*, pp. 259-268, 1987.

[4] 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.

[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] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] A. R. Akhondi-Asl, H. Soltanian-Zadeh, "Nonparametric Entropy-Based Coupled Multi-Shape Medical Image segmentation," *IEEE ISBI'2007,* pp. 1200-1203, 2007.

[13] B. W. Hong, E. prados, S. Soatto, and L. vese, "Shape Representation based on Integral Kernels: Application to Image Matching and Segmentation," *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* Vol. 1 pp. 833-840, 2006.

[14] S. Pizer, D. Fritsch, P. Yushkevich, V. Johnson, and E. Chaney, "Segmentation, registration, and measurement of shape variation via image object shape," *IEEE Trans. Med. Imaging*, vol. 18, pp. 851–865, Oct. 1999.

[15] O. Colliot, O. Camara, and I. Bloch. "Integration of fuzzy spatial relations in deformable models - Application to brain MRI segmentation," *Pattern Recognition*, Vol. 39, No. 8, pp. 1401-1414, August 2006.

[16] P. B. Bijari, A. R. Akhoundi-Asl, H. Soltanian-Zadeh, "Interactive Coupled Object Segmentation Using symmetry and Distance Constraint," *Third Cairo International Biomedical Engineering Conference,* Cairo, 2006.

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

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

[19] H. W. Kuhn, A. W. Tucker "Nonlinear programming," *Proceedings of 2nd Berkeley Symposium*, pp. 481-492, 1951.

[20] http://mathworks.com

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

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

[23] C. Guerra, and V. Pascucci, "3D segment matching using the Hausdorff distance," *Seventh International Conference on Image Processing And Its Applications,* Vol. 1, pp. 18-22, 1999.