Markov Dependence Tree-Based Segmentation of Deep Brain Structures

Jue Wu and Albert C.S. Chung

Bioengineering Program, and Lo Kwee-Seong Medical Image Analysis Laboratory,
Department of Computer Science and Engineering,
The Hong Kong University of Science and Technology, Hong Kong

Abstract. We propose a new framework for multi-object segmentation of deep brain structures, which have significant shape variations and relatively small sizes in medical brain images. In the images, the structure boundaries may be blurry or even missing, and the surrounding background is a clutter and full of irrelevant edges. We suggest a template-based framework, which fuses the information of edge features, region statistics and inter-structure constraints to detect and locate all the targeted brain structures such that manual initialization is unnecessary. The multi-object template is organized in the form of a hierarchical Markov dependence tree. It makes the matching of multiple objects efficient. Our approach needs only one example as training data and alleviates the demand of a large training set. The obtained segmentation results on real data are encouraging and the proposed method enjoys several important advantages over existing methods.

1 Introduction

Deep brain structures play critical roles in human brain functioning and their segmentation has numerous practical applications [1, 2]. One common strategy used in the segmentation of brain structures is principal component analysis (PCA), which is utilized to model shape variations [3, 4, 5]. Due to its linearity, PCA may not be able to describe the relative positions of multiple objects and their non-linear variations. The second popular brain structure segmentation strategy is fuzzy logic control [6, 7, 8], which can manage the selection of various candidates of possible structures. The problem with fuzzy logic is that it is difficult to give consistently high accuracy to the segmentations of various intracranial structures because the relationship among those structures maintained by fuzzy logic may be weak and imprecise. Another popular technique used in the segmentation of cerebral structures is level sets or active contours [3, 5, 7, 9, 10, 11], which can model the shapes of each structure. One of the concerns about active contours is the requirement of precise initialization of the starting contours. Since the brain images abound with various edges and boundaries, which are formed by interwinding gray matter and white matter, active contours are unlikely to converge to target structures if the initialization is not good enough. Another concern about the active contours is that if most of the

boundaries of a structure (e.g. thalamus) are missing, active contour methods are likely to fail.

One common feature of the above-mentioned methods [3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14] is the need of sufficient training samples. The training process usually requires a considerable number of segmented images with the known ground truths. It will be a difficult situation if a sufficient number of the ground truths is not readily available. In real applications, it is quite labor-intensive for a radiologist to extract a large number of small 3D brain structures manually to obtain training samples.

In this paper, we propose a new template-based framework for deep brain structure segmentation. Instead of using fuzzy logic or active contour framework, we integrate detection strategies with shape matching, manage multi-object relations with a hierarchical tree structure, and evaluate candidates by some specific energy functions using edge/region information and inter-object relationship. We use only one example of ground truth as the shape template, which considerably mitigates the demand of a large amount of ground truths for the training process, which is usually required in other related works. Each object or structure in the template is allowed to deform individually according to a series of transforms. The complicated interaction among multiple structures is simplified by using the Markov dependence tree (MDT). MDT requires that children objects only depend on parent objects and makes efficient the matching of the multi-object template to input images. The proposed framework has the advantages of less train burden, no need of initialization and applicability to missing boundary.

2   Methodology

In the proposed framework, we organize multiple objects by using a Markov dependence tree, where the pose of each object except for the root is only dependent on its parent object. We first need to detect a root object, which is acted as the footstone of the tree structure. For each of the other objects, we match it to the target image under a rigid transform and evaluate the result with an energy function. Part-to-part interaction between parent and child objects is also considered and relative position of interactive parts can be altered. The final segmentation is acquired after we fine-tune the coarse segmentation by affine transform according to the evaluation energy function.

2.1   Multi-object Template Construction

As the prior knowledge for segmentation, a tree structure called Markov dependence tree (MDT) is adopted to describe the relationship among the multiple objects of interest. The basic rule is that the pose of every node (object) only depends on its parent's pose except for the root node, which has no parent. This hierarchical structure greatly simplifies the interaction among those objects. For a set of N objects, we need to concern $N(N-1)/2$ relations if we consider all possible pairwise relations. In contrast, we just need to concern $N-1$ relations

if we adopt the relation graph of a tree structure. The energy minimization of the Markov dependence tree (MDT) model can be formalized as,

$$P^* = \arg\min_P E(P) = \arg\min_P(\sum_i M_i(p_i) + \sum_i \sum_{j=\text{parent of } i} T_{i,j}(p_i|p_j)), \quad (1)$$

where $E(P)$ denotes the objective energy of the tree model, P is the set of poses of all parts, $M_i(p_i)$ is a single-part local measure, p_i is the pose of part i, and $T_{i,j}$ measures a parent-to-child compatibility between parts i and j.

In our framework, we construct the template (binary map) from one example of ground truth of the target multiple objects. Only the ground truth of a single data set is needed to form the prior shape model and estimate related parameters, which is an advantage over some existing methods that need a number of training images to either generate the atlas of whole brain or estimate the related parameters of the shape model. For the segmentation of deep brain structures, we construct a three-level hierarchical template. The ventricles are the tree root, caudate nuclei are at the second level, and putamina and thalami are at the third level. This model appears to be reasonable because ventricles, whose boundaries are relatively sharp, are the most prominent object among the structures. Caudate nuclei cling to ventricles, so we regard them as child nodes to ventricles. Putamina and thalami are nearer to caudate nuclei than to ventricles and hence we put them at the third level.

2.2   Single-Part Registration

In this subsection, we focus on the first term of the objective MDT energy, which depicts the measurement of the registration of a single part to the target image. Single-part registration is a data-driven process while pairwise-part matching (the second term) is data-independent. Therefore single-part registration is served as an initial capture or detection of the parts in the image context.

In the proposed method, we make use of the gradient information, intensity variance and distance of edge to register a single part to the image. The energy M_i for single-part registration is formulated as,

$$M_i(p_i) = \alpha e_1(p_i|I) + \beta e_2(p_i|I) + \gamma e_3(p_i|I_E), \quad (2)$$

$$e_1(p_i|I) = -\frac{1}{b_i} \iiint_{B_i} \nabla I(x,y,x)dxdydz, \quad (3)$$

$$e_2(p_i|I) = \frac{1}{a_i} \iiint_{A_i} (I(x,y,z) - \overline{I}(A_i))^2 dxdydz, \quad (4)$$

$$e_3(p_i|I_E) = \text{PHD}(B_i, I_E), \quad (5)$$

where p_i is the pose of part i, I is the input image intensity and I_E refers to the edge map of the image. α, β and γ are the corresponding weights, which do not need to sum up to one. B_i is the boundary of part i, b_i is the area of the boundary B_i and $I(x, y, z)$ is the intensity of the voxel at coordinates (x, y, z). $\overline{I}(A_i)$ is the mean intensity in the image region A_i that is overlapped with part i, and a_i is the volume of the region. PHD, the partial Hausdorff distance, refers to the k-th largest distance of all the points in one point set to the other set, and is defined as $\mathrm{PHD}(B_i, I_E) = K^{th} \max_{a \in B_i} \min_{b \in I_E} D(a, b)$, where $D(\cdot, \cdot)$ denotes the Euclidean distance of two points.

The first term e_1 considers the gradient of the part region boundary. If the mean gradient value (normalized by the boundary area) is large, it is believed that the region has a clear boundary and deserves a low energy. The second term e_2 in Eq. 2 focuses on the gray level variation within the part region. It prefers to a region with smaller intensity variance. The third term in Eq. 2 encourages the partial Hausdorff distance between the boundary of component p_i and the image edge map to be as small as possible.

2.3   Pairwise-Part Matching

When one part is registered to the image using single-part registration, we also adjust its pose by means of pairwise-part matching, by which we can eliminate some false detections of part pose in the single-part registration stage via high-level interactions and constraints of the close objects. Note that the matching will only change the pose of child part while the parent pose is fixed. Triple-part or more interactions can be considered but the relation graph will not be tree-structured and thus the computational burden will be increased considerably. The matching of two interactive parts is evaluated by the following energy function $T_{i,j}$,

$$T_{i,j}(p_i|p_j) = \delta e_4(p_i|p_j) + \zeta e_5(p_i|p_j) + \eta e_6(p_i|p_j), \tag{6}$$

$$e_4(p_i|p_j) = D(C_{p_i}, C_{p_j}), e_5(p_i|p_j) = ||\boldsymbol{s}_i - \boldsymbol{s}_j||, e_6(p_i|p_j) = I\{n_{ij}/n_j > \tau\}, \tag{7}$$

where p_i and p_j are the poses of parts i and j, respectively. δ, ζ and η are the corresponding weights of each term. $D(\cdot, \cdot)$ denotes the Euclidean distance of two points. C_{p_i} is the center of mass of part i, which is calculated by, $C_{p_i} = \frac{1}{a_i}(\iiint_{A_i} x dx dy dz, \iiint_{A_i} y dx dy dz, \iiint_{A_i} z dx dy dz)$. a_i is the volume of the image region A_i that is overlapped with part i, assuming density is 1. (x, y, z) are the coordinates of the voxel in region A_i. $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ are the vectors of scaling (each has three elements in 3D) of the parts i and j, respectively. Operator $||\cdot||$ refers to the magnitude of a vector. $I\{event\}$ is an indicator function, i.e., if an event is true, $I = 1$; otherwise $I = 0$. a_j is the volume of part j, n_{ij} is the volume of the overlapping region of parts i and j, and τ is a tolerance factor with respect to the overlapping extent.

The first term e_4 checks the distance between the centers of mass of two parts i and j, and tries to keep the parts in a proper distance. The second term e_5

is pertinent to the scale difference of the two interactive parts. It assures that the two parts have similar levels of scaling. The third term of $T_{i,j}$ deals with the overlapping problem. It penalizes the situation that the overlapping of two parts exceeds certain extent.

2.4 Detection, Segmentation of Root Object and Parameter Estimation

Detection and segmentation of ventricles are not easy tasks and there are not many research works in the literature directly related to it. In the proposed framework, we adopt an automatic method without the need of specifying initial seeds manually. At first, we perform tissue segmentation to classify voxels in the brain image volume into white matter (WM), gray matter (GM), cerebrospinal fluid (CSF) and others. The voxel-based segmentation is implemented through a Gaussian mixture model, which is optimized by the expectation-maximization algorithm. We then register the ventricle shape in the template to the CSF binary segmentation map according to the MDT evaluation function (Eq. 1) and obtain the rough position of the ventricles in the original image. The proposed method further extracts automatically some seeds whose intensity is very dark among the rough ventricle segmentation. The rationale behind this is that CSF has always the lowest intensity compared to WM and GM in deep brain region in T1-weighted MRI. At the final step, we perform a neighborhood connected region growing algorithm based on the selected seeds to extract the ventricles.

In the energy function of MDT (Eq. 1), there are six parameters $(\alpha, \beta, \gamma, \delta, \zeta, \eta)$ that can be adjusted. From the angle of energy minimization, one of the parameters can be fixed and the five others are free parameters. We propose an automatic method to estimate the free parameters. We can calculate the six terms (e_1 to e_6) in the objective energy function based on the single training image and the corresponding true poses of each structure. After the values of the six terms are calculated, we can estimate the weights of each term by means of their ratios as follows, $\alpha : \beta = e_2(p_i^t|I) : e_1(p_i^t|I), \alpha : \gamma = e_3(p_i^t|I_E) : e_1(p_i^t|I), \alpha : \delta = e_4(p_i^t|p_j^t) : e_1(p_i^t|I), \alpha : \zeta = e_5(p_i^t|p_j^t) : e_1(p_i^t|I), \alpha : \eta = 1 : e_1(p_i^t|I)$, where p_i^t and p_j^t are the true poses of parts i and j in the training image I, respectively. I_E is the edge map of I. We further set $\alpha = 1$, and then the six parameters can be determined. Since the importances of each term are roughly the same, we equalize the weights of each term according to the values from the true poses.

There are some other minor parameters to be estimated. For instance, the lower and upper thresholds of the neighbor connected region growing can also be set according to the example data since we know the range of intensity of ventricles in the ground truth.

3   Experiments and Validation

In the experiments, we have applied the proposed methodology to the segmentation of deep brain structures. We focus on the segmentation of six gray matter

brain structures, i.e., left and right caudate nuclei, putamina and thalami. These six deep brain structures are often studied in brain anatomy [2, 3, 4, 5, 8, 10, 11, 12, 13, 15]. However, they are difficult to segment automatically due to their blurry boundary and small sizes.

The testing data sets were obtained from the Internet Brain Segmentation Repository (IBSR), which includes 18 high-resolution brain MR (T1-weighted) scans with the expert segmentations of internal structures. The MR brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis at the Massachusetts General Hospital, U.S.A. The resolution of each scan is $256 \times 256 \times 128$ voxels. More information about the data sets can be found at http://www.cma.mgh.harvard.edu/ibsr/.

We took the ground truth of one data set randomly in the database as the shape template and treated the other 17 sets as the target (testing) images. Note that even in the same database, the data sets bear some extents of dissimilarity in terms of voxel size, intensity profiles and intracranial structure morphology because the data sets were attained from different subjects. We implemented the proposed framework under the help of Insight Toolkit (ITK) 3.0.1. using C++ on the platform of Microsoft Visual C++ .NET 2003. The computational time on each data set is around an hour on a 2.13GHz CPU with 1GB memory. Note that all the model parameters are estimated once for ado and the results of different input data sets are based on the same set of parameters.

We have compared the results with the expert segmentations available from the IBSR project, which we considered as the "gold standard". The extent of the overlapping of the experimental results and the ground truth was measured by a widely used index, i.e., Dice similarity coefficient (DSC). It is defined as, $DSC = \frac{2||A \bigcap B||}{||A||+||B||}$, where A and B are two point sets to be compared. $|| \cdot ||$ denotes the number of points in the point set. DSC reflects the extent of the overlapping of two point sets. The larger the measure is, the more overlapping the two sets have. The range of DSC is $[0, 1]$. Besides, we calculated the mean distance of the segmentation results to the ground truth, which is defined as [6, 7], $S_m = \frac{\sum_{a \in A} \min_{b \in B} D(a,b)}{||A||}$, where A and B are the segmentation result and ground truth, respectively, and $D(\cdot, \cdot)$ denotes the Euclidean distance of two points. Mean distance also embodies the difference of two shapes in space. The smaller it is, the more similar the two shapes are. We calculated DSC and mean distance for each structure in the 17 testing data and the overall performance is listed in Tab. 1. From the results, we find that all six structures segmented by the proposed method have a good overlapping (DSC > 0.75) with the expert segmentations.

In addition to the quantitative performance, we present a visual inspection of the experimental results. In Fig. 1, we show the results on some of the IBSR data sets at approx. the same slice location. Although the intensity profile, shape appearance of the structures and location of the brain are different among these images, our method is able to detect the poses of brain structures and perform satisfactory segmentations. From both 2D and 3D inspections, we observe that the results from the proposed method resemble the expert segmentations though

Table 1. The measures of six brain structures of the proposed segmentation results on 17 IBSR data sets as compared to the expert ground truth. The average measures and their standard deviations are shown. LCN(RCN) = left(right) caudate nucleus. LP(RP) = left(right) putamen. LT(RT) = left(right) thalamus.

	LCN	LP	LT	RCN	RP	RT
DSC	0.78 ± 0.04	0.76 ± 0.04	0.75 ± 0.09	0.77 ± 0.03	0.76 ± 0.06	0.78 ± 0.05
S_m (mm)	0.34 ± 0.12	0.53 ± 0.15	0.73 ± 0.51	0.34 ± 0.09	0.40 ± 0.15	0.56 ± 0.21
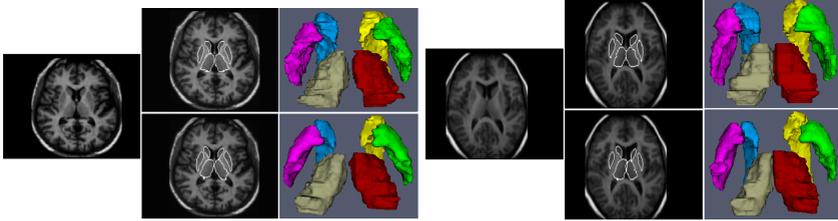
Fig. 1. (*Color Images*) Two sets of data, their ground truth and segmentation results around the middle slice. For each group, the left image is the original data. For each subgroup, the up row on the right is the ground truth and 3D surface visualization. The down row is the segmentation.

the boundaries are smoother. But abrupt change in segmented object boundaries is a common problem of 3D object segmentation if it is performed manually on a slice by slice basis (see an example of the thalami in 3D view of the second data set in Fig. 1). The proposed method seems to alleviate this problem.

4   Discussion and Conclusion

It is worth pointing out that it is very difficult to compare the proposed method with other related works experimentally due to the unavailability of the source codes of the related works. So we only compare the proposed method with other related works conceptually if the quantitative results on the same database were not reported in those works. The work of [7] reported quantitative results of the segmentation of the deep brain structures on the same IBSR data sets (failed in 3 sets and the following results of theirs were from the other 14 sets). The thalamus segmentation of the proposed method is slightly better than the reference work [7] (DSC is the same (0.77 vs. 0.77) and S_m is much shorter (0.65 vs. 1.79)), while our segmentation accuracies of putamina and caudate nuclei are significantly higher (DSC: 0.76 vs. 0.65 for caudate nuclei and 0.75 vs. 0.70 for putamina. S_m: 0.34 vs. 1.71 for caudate nuclei and 0.47 vs. 1.46 for putamina). The reason is that the proposed method has a more flexible architecture to model the interrelations among neuroanatomical structures for dealing with the small and subtle objects in deep brains than that in [7]. For the other related works [3, 5, 16], they did not report the quantitative results of the segmentation. An

advantage of the proposed method over those methods is that we circumvents the intractable blurry boundaries of brain structures, such as thalami, by emphasizing the relative position to other deep brain structures. On the contrary, the level-set-based methods used in [3, 5, 16] may not be able to perform the segmentation of thalami when facing objects with weak edges and gradients. Even the use of the idea of "active contours without edges" [17] cannot significantly improve the methods [3, 11] to find the missing boundaries of thalami because the thalami are located in the clutter environment in brains. Some voxels outside the thalami are darker than the thalami, e.g. CSF and background. Other voxels outside the thalami are brighter, e.g. white matter. The average intensity of the outside region is close to the average intensity of the inside region, which may not evolve the contours correctly.

We introduce the idea of detection into the segmentation of deep brain structures. This is the first time, to the best of our knowledge. The detection techniques greatly improve the automation of the proposed method. There is no need to initialize the algorithm by hand. In contrast, the level set based methods [3, 5, 7, 10, 11] need to initialize the active contours inside or close to every structure of interest. On the other hand, registration-based methods [2, 12, 13, 15, 18] can obviate the problem of contour initialization but unfortunately, it is challenging to build an informative and effective atlas for brain structures and warp the atlas to input brain images using non-rigid registration.

In the literature [4, 5, 8, 11, 12], the segmentation experiments were all performed under a condition that the training set is not smaller than the testing set. The extreme is the "leave-one-out" experiment [5, 8, 11]. In the pool of data sets (usually more than ten data sets), all except for one data set are used as training data and the remained one as testing. This leave-one-out experiment is fine for cross validation but is not enough for practical validation. In practical use, if most of the data are segmented manually and only a few data are processed by an automatic computer program, it is not meaningful for labor-saving. Therefore, the less need of hand segmentation, the better for practical use. However, if the training set is largely reduced, those works that were only validated by leave-one-out experiments may not be stable or reliable because they need to acquire prior information as much as possible. This dilemma makes important our method, which only needs a single example of ground truth to complete a set of satisfactory segmentations.

References

1. Grenander, U., Miller, M.: Computational anatomy: An emerging discipline. Quarterly of Applied Mathematics 56, 617–694 (1998)
2. Iosifescu, D., Shenton, M., Warfield, S., Kikinis, R., Dengler, J., Jolesz, F., McCarley, R.: An automated registration algorithm for measuring mri subcortical brain structures. NeuroImage 6, 13–25 (1997)
3. Yang, J., Staib, L., Duncan, J.: Neighbor-constrained segmentation with level set based 3-d deformable models. IEEE Trans. on Medical Imaging 23(8), 940–948 (2004)

4. Tu, Z., Narr, K., Dollar, P., Dinov, I., Thompson, P., Toga, A.: Brain anatomical structure segmentation by hybrid discriminative/generative models. IEEE Trans. on Medical Imaging (accepted for publication)

5. Tsai, A., Wells, W., Tempany, C., Grimson, E., Willsky, A.: Mutual information in coupled multi-shape model for medical image segmentation. Medical Image Analysis 8(4), 429–445 (2004)

6. Barra, V., Boire, J.: Automatic segmentation of subcortical brain structures in mr images using information fusion. IEEE Trans. on Medical Imaging 20(7), 549–558 (2001)

7. Ciofolo, C., Barillot, C.: Brain segmentation with competitive level sets and fuzzy control. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 333–344. Springer, Heidelberg (2005)

8. Zhou, J., Rajapakse, J.: Segmentation of subcortical brain structures using fuzzy templates. NeuroImage 28, 915–924 (2005)

9. Duta, N., Sonka, M.: Segmentation and interpretation of mr brain images: An improved active shape model. IEEE Trans. on Medical Imaging 17(6), 1049–1062 (1998)

10. Cootes, T., Taylor, C.: Statistical models of appearance for medical image analysis and computer vision. In: SPIE Medical Imaging, vol. 4322, pp. 236–248 (2001)

11. Yang, J., Duncan, J.: 3d image segmentation of deformable objects with joint shape-intensity prior models using level sets. Medical Image Analysis (8), 285–294 (2004)

12. Pohl, K., Fisher, J., Grimson, W., Kikinis, R., Wells, W.: A bayesian model for joint segmentation and registration. NeuroImage 31, 228–239 (2006)

13. Pohl, K., Fisher, J., Bouix, S., Shenton, M., MacCarley, R., Grimson, E., Kikinis, R., Wells, W.: Using the logarithm of odds to define a vector space on probabilistic atlases. Medical Image Analysis 11(5), 465–477 (2007)

14. Gouttard, S., Styner, M., Joshi, S., Smith, R., Cody, H., Gerig, G.: Subcortical structure segmentation using probabilistic atlas priors. In: SPIE Medical Imaging, vol. 6512, 65122J–1 – 65122J–11 (2007)

15. Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355 (2002)

16. Litvin, A., Karl, W.: Coupled shape distribution-based segmentation of multiple objects. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 345–356. Springer, Heidelberg (2005)

17. Chan, T., Vese, L.: Active contours without edges. IEEE Trans. on Image Processing 10(2) (2001)

18. Rohlfing, T., Brandt, R., Menzel, R., Maurer, C.R.: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage 21(4), 1428–1442 (2004)