NeuroImage 46 (2009) 1027–1036

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

A novel framework for segmentation of deep brain structures based on Markov dependence tree

Jue Wu [a,b], Albert C.S. Chung [b,*]

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

ARTICLE INFO

Article history:
Received 23 June 2008
Revised 24 February 2009
Accepted 1 March 2009
Available online 13 March 2009

Keywords:
Deep brain structures
Caudate nucleus
Putamen
Thalamus
Multi-object segmentation
Markov dependence tree

ABSTRACT

The aim of this work is to develop a new framework for multi-object segmentation of deep brain structures (caudate nucleus, putamen and thalamus) in medical brain images. Deep brain segmentation is difficult and challenging because the structures of interest are of relatively small size and have significant shape variations. The structure boundaries may be blurry or even missing, and the surrounding background is full of irrelevant edges. To tackle these problems, we propose a template-based framework to fuse the information of edge features, region statistics and inter-structure constraints for detecting and locating all target brain structures such that initialization by hand is unnecessary. The multi-object template is organized in the form of a hierarchical Markov dependence tree (MDT), and multiple objects are efficiently matched to a target image by a top-to-down optimization strategy. The final segmentation is obtained through refinement by a B-spline based non-rigid registration between the exemplar image and the target image. Our approach needs only one example as training data. We have validated the proposed method on a publicly available T1-weighted magnetic resonance image database with expert-segmented brain structures. In the experiments, the proposed approach has obtained encouraging results with 0.80 Dice score for the caudate nuclei, 0.81 Dice score for the putamina and 0.84 Dice score for the thalami on average.

Introduction

Deep brain structures, such as the caudate nucleus, the putamen and the thalamus, play a critical role in human brain functioning (Packard and Knowlton, 2002). Almost all sensory information sent by the neurons to the cortex goes through the thalamus structure. Putamen is responsible for motor skills and reinforcement learning. The caudate nucleus is known to be an important component of the brain's learning, memory and language comprehension system. Therefore, segmentation of deep brain structures has been an essential technique in human brain image analysis, as deep brain segmentation paves the way for many subsequent quantitative analysis tasks, such as deformable neuroanatomies (Miller et al., 1993), computational anatomy (Grenander and Miller, 1998), the study of cerebral development and aging (Kandel et al., 2000), and brain disease diagnosis (Iosifescu et al., 1997).

Since segmentation of brain structures has numerous practical applications, it has been studied intensively and several methods have been proposed in the medical image analysis community. One common strategy used in the segmentation of brain structures is principal component analysis (PCA), which is utilized to model the shape variations (Yang et al., 2004; Tu et al., 2008; Tsai et al., 2004). For example, the level set based segmentation method, proposed by Yang et al. (2004), modeled the shapes of multiple objects by principal components and considered the constraints of neighboring objects in the maximum a posteriori (MAP) estimation framework. Tu et al. (2008) proposed a hybrid method to learn high and low-level information for brain structures, and each shape model was learned by PCA. The method proposed by Tsai et al. (2004) utilized multiple signed distance functions to represent the variations of multiple shapes by means of several principal modes. However, due to its linearity, PCA may not be able to properly model the multishape variations of several neuro-anatomical structures simultaneously. This is because even though the shape variation of a single structure can be approximated by principal components, PCA can fail as a result of the non-linear shape variation of multiple structures as a whole.

The second popular strategy used in the segmentation of brain structures is fuzzy logic control (Barra and Boire, 2001; Ciofolo and Barillot, 2005; Zhou and Rajapakse, 2005). This strategy can manage the selection of various candidates of possible structures. Barra and Boire (2001) proposed a fuzzy logic based segmentation method for cerebral structures. This method was able to translate the descriptions of expert experiences into a common theoretical frame. It aggregated all the information for various structures and made the decisions to deal with ambiguity and incertitude in order to get final results. Another work that adopted the principles of fuzzy logic was proposed by Ciofolo and Barillot (2005). The competitive level set based

* Corresponding author. Fax: +852 2358 1477.
E-mail address: achung@ust.hk (A.C.S. Chung).

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2009.03.010

segmentation was integrated with the fuzzy control strategy. The fuzzy decision system supervised the simultaneous evolution of several active contours according to an anatomical atlas. Zhou and Rajapakse (2005) suggested a method to create fuzzy templates from a set of training images. Their method registered the fuzzy templates with the test images, and made the decisions about the segmentation by optimizing the certainty in intensity, location, relative position and tissue type. The potential problem with fuzzy logic is that it is difficult to consistently provide highly accurate segmentation of various intracranial structures because the relationship among those structures maintained by fuzzy logic is weak and imprecise. It will be better to model the relations precisely; otherwise the segmentation accuracy can dramatically decrease because the structures are of small size.

Another popular technique used in the segmentation of cerebral structures is level sets or active contours, which model the contour of each structure (Duta and Sonka, 1998; Ciofolo and Barillot, 2005; Cootes and Taylor, 2001; Tsai et al., 2004; Yang et al., 2004; Yang and Duncan, 2004). The active shape model (Cootes and Taylor, 2001) computed a statistical model based on landmark points marked in a training set of images, and matched the model with the boundaries in the input image to synthesize normal anatomy. Duta and Sonka (1998) presented a 2D segmentation approach to improve the active shape model in order to find new examples of previously learned shapes using point distribution models. In the framework of level sets, the method proposed by Yang and Duncan (2004) integrated the gray level of each structure with shape prior information to form a joint distribution based on a set of training images. A problem with the active contour methods is the requirement of precise initialization of the starting contours. Since the brain image abounds with various edges and boundaries formed mainly by the borders between cerebrospinal fluid, gray matter and white matter, the active contours are unlikely to converge to the target structures unless the initialization is good enough. Another concern is that if most of the boundaries of the structures (e.g. the thalamus) are missing, it is difficult to converge to the target structures for the active contour method.

The third common methodology is registration-based subcortical structure segmentation (Miller et al., 1993; Iosifescu et al., 1997; Pohl et al., 2006, 2007a,b; Fischl et al., 2002; Gouttard et al., 2007; Khan et al., 2008; Heckemann et al., 2006; Svarer et al., 2005; Powell et al., 2008; Gousias et al., 2008). In these methods, a training gray level image or an atlas is registered to a target image. The labels associated with the training image or the atlas are thus propagated to the target image and the segmentation is accomplished with the help of image registration. For example, Pohl et al. (2007a) proposed a hierarchical method for brain parcellation, including the segmentation of temporal gyrus, amygdala and hippocampus. Their hierarchy was used to represent relations between tissues and structures and the object in the tree was a further subdivision of its parent object. The solution was obtained by the expectation maximization algorithm combined with spatial prior information from a registered probability atlas to identify the invisible boundary of some structures. In the work of Powell et al. (2008), the authors made use of the artificial neural networks and the support vector machines to classify 25-element feature vectors created from multi-modal brain MR images. One element in the feature vector was obtained from the registration of a template to an image and this element represented the probability that the corresponding voxel belonged to one of the brain structures. Another work based on image registration was proposed to combine the existing software FreeSurfer and large deformation diffeomorphic metric mapping (Khan et al., 2008). They used the labeling results obtained from the probabilistic FreeSurfer as the initialization for the succeeding fine-tuning by the diffeomorphic registration. Intensity normalization was required in the preprocessing. Regions of interest were extracted to enhance the efficiency of the computationally expensive diffeomorphic registration.

Registration-based methods can be highly accurate because sophisticated non-rigid registration techniques can be applied in a meticulous way and the process makes use of context information around each brain structure as much as possible. The main workload of the atlas registration-based methods lies in the construction of an informative brain atlas from a population of subjects. The construction can be arduous because it is data-intensive. It is challenging to fuse large amount of information from the population. Besides, the precise non-rigid registration between the atlas/training image and the various subjects is still an active research topic. The proposed method in the current work can be regarded as a registration-based approach as well. The novelty of the proposed method is that we introduce the idea of detection into structure segmentation. The benefit has two folds. First, we can focus on the specific structures of interest instead of registering the whole brain or a considerable number of structures at the same time (Powell et al., 2008; Pohl et al., 2007a). The registration of the whole brain or a pool of structures offers the context of the target structures but it also brings additional computational costs. Moreover, it tends to deform two close structures in a similar way. Second, the detection techniques make the initialization of the positions of the target structures unnecessary. The initialization can be completed either by hand (Ciofolo and Barillot, 2005; Yang et al., 2004) or by stand-alone softwares (Khan et al., 2008). In the proposed method, the initialization is performed by the structure detection process. Experimental comparison of the proposed work with closely related works (Powell et al., 2008; Khan et al., 2008) can be found in the Discussion section.

One common requirement of the above-mentioned methods (Sonka et al., 1996; Duta and Sonka, 1998; Ciofolo and Barillot, 2005; Pohl et al., 2006, 2007b; Tsai et al., 2004; Yang et al., 2004; Tu et al., 2008; Yang and Duncan, 2004; Gouttard et al., 2007; Cootes and Taylor, 2001; Zhou and Rajapakse, 2005; Powell et al., 2008; Khan et al., 2008) is the need for sufficient training. The training process requires a considerable number of segmented images with the known ground truths. In clinical applications, it is labor intensive for a radiologist to extract a large number of small 3D brain structures manually. If a sufficient number of the ground truths is not readily available, the segmentation performance will be adversely affected. Even though the ground truths of some public databases exist, the differences between the databases make it unsuitable for one database to serve as a training set for other databases except in a few works (Gousias et al., 2008). In addition, the training process can be time-consuming and may take a long time to complete (Powell et al., 2008).

In this paper, we propose a new template-based framework for the segmentation of deep brain structures. Our segmentation framework integrates detection strategies with shape matching. The framework manages multi-object relations with a hierarchical tree structure and evaluates candidates by some specific energy functions using edge/region information and inter-object relationship. Only one example of ground truth is required as the shape template. Each object or structure in the template is allowed to deform individually according to a series of transforms and a top-down optimization strategy. In our framework, the complicated interaction among multiple structures is simplified by using Markov dependence tree (MDT). MDT requires that children objects only depend on parent objects. The rationale behind the hierarchical structure is that not all objects are the same in terms of detection difficulty. We assign the easiest object to detect as the tree root and arrange the other objects hierarchically according to their detection difficulty.

This paper is organized as follows. The Methodology section presents the details of our proposed framework including multi-object template construction, MDT energy optimization strategy, root object detection and segmentation, combined single-part and pair-wise-part measures, and the implementation details. Experiments and validation are given in the Experiments and validation section.

The Discussion section discusses the proposed framework and the Summary section concludes the paper.

Methodology

In this section, we present our framework for multi-object segmentation. The framework is based on a hierarchical detection model and takes into consideration both the high-level shape constraints and the low-level voxel-based properties.

Overview

The rationale of the proposed methodology is grounded on the fact that a complex object can be divided into several parts (or components), and various appearances of the object can be modeled by manipulating each part. We can regard multiple objects with different shapes as one large complex object. In medical applications, multiple anatomic object recognition, detection or segmentation can be achieved by using this strategy.

In our proposed framework, we organize multiple objects using a Markov dependence tree, in which each object except the root is dependent on its parent object. We first need to detect a root object, which acts as the footstone of the tree structure. The other objects are then matched with the target object in the input image volume and the matching result for each object is evaluated with an energy function. Part-to-part interaction is also considered in the energy function and the relative position of interactive parts can be altered. In our hierarchical framework, part-to-part interaction is employed between parent and child objects. Every object is allowed to undertake a rigid transform followed by an affine transform during the matching until a coarse segmentation of multiple objects is obtained. The final segmentation is acquired after we fine-tune the coarse segmentation by using non-rigid B-spline registration between the exemplar image (i.e., the training image) and the target image on a structure-by-structure basis. The flowchart of our framework is drawn in Fig. 1.

Multi-object template construction

A training example is composed of a gray level image and its segmentation (ground truth). As the prior knowledge, a high-level multi-object template is constructed based on the segmentation of the training example. Two elements of the template have to be specified:

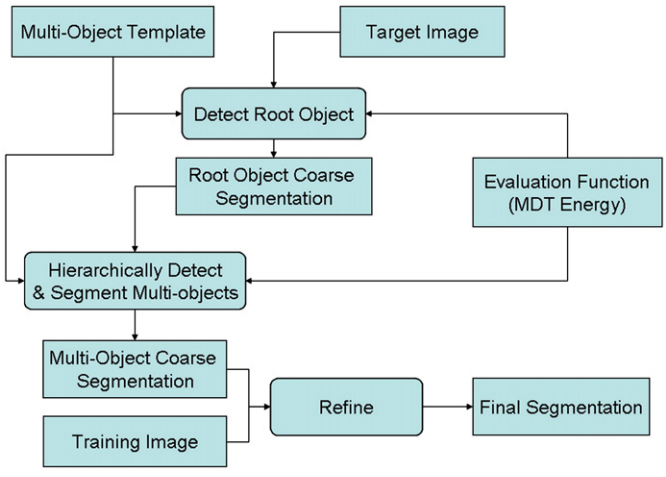

Fig. 1. The flowchart of the proposed framework. The template is a binary image formed by the multi-object segmentation of the training image. The ventricles are regarded as root object. The hierarchical detection and segmentation are based on the single-part and the pairwise-part measures with sequential rigid and affine transforms. The training image is gray-leveled.

one is the approximate shape of each object and the other is the dependency among the objects. A tree structure, which we call Markov dependence tree (MDT), is adopted to describe the relationship among the multiple objects of interest. The basic rule is that the pose[1] of every object depends on its parent's pose except for the root object, which has no parent. The pose of a parent object is independent of the poses of its child objects. This hierarchical structure greatly simplifies the interactions among those objects. For a set of N objects, we need to consider about $N(N-1)/2$ relations if all possible pairwise relations are considered. In contrast, we only need to consider $N-1$ relations if we adopt the relation graph of a tree structure.

The energy of the Markov dependence tree (MDT) model can be formalized as,

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

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

We construct the multi-object template from one example of segmentation results of the target multiple objects. Only the ground truth of a single data set is needed in order to form the prior shape model and estimate related parameters. Instead of requiring a number of training images to either generate the atlas of the whole brain or estimate the related parameters of the shape model, the proposed method chooses a simple way to construct the prior shape model.

For the segmentation of deep brain structures, we construct a three-level hierarchical template, as shown in Fig. 2. The ventricles are the tree root, the caudate nuclei are at the second level, and the putamina and the thalami are at the third level. The ventricles, whose boundaries are relatively sharp, are the most prominent object among the structures. The caudate nuclei cling to the ventricles and have sharper boundaries and more homogeneous intensity than the putamina and the thalami so we assign the caudate nuclei as the child objects to the ventricles and place the putamina and the thalami at the third level of the model. It is a reasonable model from the point of view of detection. We put objects relatively easier to detect on the higher levels (closer to the root) in the tree. As such, the detection of "hard" objects depends on the detection of "easy" objects but not vice versa.

MDT optimization

Once the multi-object template is ready, we can start to map the template to the input image. The mapping results are evaluated based on the MDT energy, as stated in Eq. (1). It is difficult to find the exact minimum of the MDT energy since the searching space is huge if all the parts are allowed to deform simultaneously. In addition, the energy profile is not convex or concave and has a lot of local minima.

Hence, we adopt an optimization strategy suitable for the tree structure. The strategy is based on a top-to-down style from the root object to the leaf objects. We start to optimize one object only after its parent object is optimized and fixed. The optimization is evaluated by both two energy terms that are related to the object, i.e., single-part term M_i and pairwise-part term $T_{i,j}$. When the optimization is finished, the current object is fixed and then we go on to optimize its child objects.

The hierarchical optimization strategy fits with the spirit of using tree structure. We assign the relatively easy-to-detect objects at the

[1] A part pose is the set of points inside the part. In this work, it is represented by a binary image with the same size as the input image.

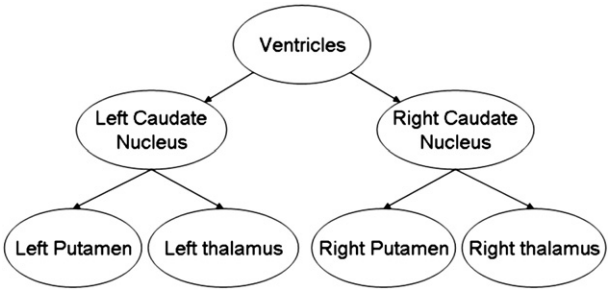

Fig. 2. A three-level Markov dependence tree representing the relations between deep brain structures.

high level (close to root) in the MDT. Since the caudate nucleus is easier to detect than the putamen or the thalamus and harder to detect than the ventricle, the pose of the caudate nucleus depends on that of the ventricle instead of those of the putamen and the thalamus. This dependency is reflected in the tree as the form of unilateral dependence of children on their parent. Therefore, during the optimization, once a parent object is matched according to image content and its parent, it is fixed and will not be affected by the results of the matching of its children. This strategy is reasonable as well as computationally efficient.

We make use of the Powell's method as the optimizer to minimize the energy terms that are related to each object. Powell's method is an efficient direction-set-based method that does not need to evaluate the gradient of object function. More details about the Powell's method can be found in (Press et al., 1992).

Detection and segmentation of root object

The root object is the cornerstone of the whole multi-object template. Therefore, we select an object that is relatively easier to segment as the root object. In the segmentation of deep brain structures, the ventricles are selected as the root object since they have relatively sharp boundaries. The brain ventricles are full of cerebrospinal fluid and have different intensity from the surrounding white matter and gray matter.

However, the detection and segmentation of the ventricles are not easy tasks and there are not many directly related research works in the literature. In the proposed framework, we adopt an automatic method without specifying initial seeds. 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 mainly focus on the segmentation results of CSF, which permeates the ventricles. Then we register the ventricle shape in the template to the CSF binary segmentation map according to the single-part measurement M_i in the MDT evaluation function (Eq. (1)). After this, we know the rough position of the ventricles in the original image. To obtain more accurate boundary of the ventricles, the proposed method further extracts automatically some seeds with the lowest intensity among the rough ventricle segmentation. The reason is that CSF has the lowest intensity compared with WM and GM in the deep brain region in T1-weighted MR images. At the final stage, we perform a neighborhood connected region growing algorithm (Ibanez and Schroeder, 2005) based on the selected seeds to extract the ventricles. The flowchart of the segmentation of the root object is shown in Fig. 3.

Single-part measures

In this subsection, we focus on the first term of the objective MDT energy (Eq. (1)), which depicts the measurement of the registration of a single part in the target image. The single-part measures are data-driven while the pairwise-part measures (the second term) are data-independent. Note that the single-part and the pairwise-part measures are simultaneously in effect in the energy optimization.

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

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

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 volume. α, β and γ are the corresponding weights, which do not need to sum up to one. The three terms, e_1, e_2 and e_3, are discussed below.

The first term e_1 considers the gradient of the part region boundary. If the mean gradient magnitude (normalized by the boundary area) is large, it is believed that the region has a clear boundary and deserves a low energy. The term is formulated as,

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

where 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 coordinate (x, y, z).

The second term e_2 in Eq. (2) focuses on the gray level variations within the part region. It prefers a region with smaller intensity variance and is expressed by,

$$e_2(p_i|I) = \frac{1}{a_i}\int\int_{A_i}\int \left(I(x,y,z) - \bar{I}(A_i)\right)^2 dxdydz, \tag{4}$$

where $\bar{I}(A_i)$ 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 A_i.

The third term in Eq. (2) encourages the partial Hausdorff distance between the boundary of part p_i and the image edge map to be as small as possible.

$$e_3(p_i|I_E) = \mathrm{PHD}(B_i, I_E), \tag{5}$$

where B_i is the boundary of the pose of part i. PHD, the partial Hausdorff distance (Huttenlocher et al., 1993), refers to the k^{th} largest

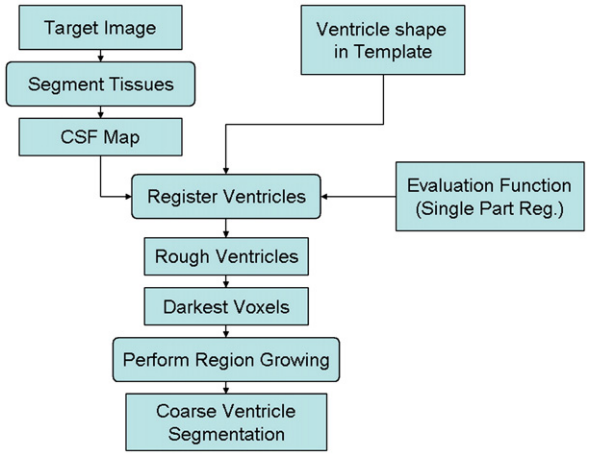

Fig. 3. The flowchart of detection and segmentation of the root object. Tissue segmentation is performed by finite mixture model optimized by the expectation maximization algorithm. The voxels with the lowest intensity are extracted from the rough ventricle segmentation and then they are used as seeds for the region growing with constraint of neighborhood connection.

distance of all the points in one point set to the other set, and is defined as,

$$\mathrm{PHD}(B_i, I_E) = k^{\mathrm{th}} \max_{a \in B_i} \min_{b \in I_E} D(a,b), \tag{6}$$

where $D(\cdot,\cdot)$ denotes the Euclidean distance of two points.

Pairwise-part measures

When one part is registered to the input image volume using the single-part measures, at the same time we also adjust its pose by means of the pairwise-part measures. The reference part for the current part is its parent in the Markov dependence tree. This matching is data-independent because it will only consider the relationship between the parent and the child parts rather than the image content. Through the pairwise-part measures, we can eliminate some false detections of part poses via the high-level interactions and constraints from the close objects. Note that the matching will only change the pose of the 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 increase considerably. The matching of two interactive parts is evaluated by the following energy function $T_{i,j}$,

$$T_{i,j}\left(p_i \mid p_j\right) = \delta e_4\left(p_i \mid p_j\right) + \zeta e_5\left(p_i \mid p_j\right) + \eta e_6\left(p_i \mid p_j\right), \tag{7}$$

where p_i and p_j are the poses of parts i and j, respectively. δ, ζ and η are corresponding weights of each term, which do not need to sum up to one.

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 short distance. It is defined as,

$$e_4\left(p_i \mid p_j\right) = D\left(C_{p_i}, C_{p_j}\right), \tag{8}$$

where $D(\cdot,\cdot)$ denotes the Euclidean distance of two points. C_{pi} is the center of mass of part i, which is calculated by,

$$C_{p_i} = \frac{1}{a_i}\left(\int\int\int_{A_i} x\,dx\,dy\,dz, \int\int\int_{A_i} y\,dx\,dy\,dz, \int\int\int_{A_i} z\,dx\,dy\,dz\right), \tag{9}$$

where a_i is the volume of the image region A_i that is overlapped with part i, assuming uniform density is 1. (x, y, z) are the coordinates of the voxel in region A_i.

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.

$$e_5\left(p_i \mid p_j\right) = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|, \tag{10}$$

where $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ are the vectors of scaling (which has three elements in 3D) of the parts i and j, respectively. Operator $\|\cdot\|$ refers to the magnitude of a vector.

The third term of $T_{i,j}$ deals with the overlapping problem. It penalizes the situation that the overlapping of two parts exceeds a certain extent.

$$e_6\left(p_i \mid p_j\right) = I\left\{a_{ij} / a_j > \tau\right\}, \tag{11}$$

where $I\{\text{event}\}$ is an indicator function, i.e., if the event is true, $I=1$; otherwise $I=0$. a_j is the volume of part j, a_{ij} is the volume of the overlapping region between parts i and j, and τ is a tolerance factor with respect to the overlapping extent.

During the matching process of all the objects in the Markov dependence tree, we adopt two transforms sequentially for every object. First, we perform a simple 3D similarity transform and its degree of freedom is 7. The transform equation of similarity transform can be expressed by,

$$\boldsymbol{V}' = \mathbf{T}_{\mathrm{t}}\mathbf{T}_{\mathrm{r}}S\boldsymbol{V}, \tag{12}$$

where $\boldsymbol{V} = (x\ y\ z\ 1)^{\mathrm{T}}$ and $\boldsymbol{V}' = (x'\ y'\ z'\ 1)^{\mathrm{T}}$ are the coordinate vectors of a voxel before transform and after transform, respectively. $\mathbf{T}_{\mathrm{t}}$ and $\mathbf{T}_{\mathrm{r}}$ are the standard translation and rotation transform matrices, respectively. S is the standard isotropic scaling transform matrix.

Next we adopt affine transform to deform each object based on the initialization obtained from the results of similarity transform. The matching process with affine transform is conceptually the same as that with similarity transform because they refer to the same evaluation energy function (Eq. (1)) and go through the same matching process. The degree of freedom of affine transform is 12 for 3D objects. The transform equation can be expressed by,

$$\boldsymbol{V}' = \mathbf{T}_{\mathrm{t}}\mathbf{T}_{\mathrm{r}}\mathbf{T}_{\mathrm{sh}}\mathbf{T}_{\mathrm{sc}}\boldsymbol{V}, \tag{13}$$

where $\mathbf{T}_{\mathrm{t}}$, $\mathbf{T}_{\mathrm{r}}$, $\mathbf{T}_{\mathrm{sh}}$, $\mathbf{T}_{\mathrm{sc}}$ are the translation, rotation, shearing, anisotropic scaling transform matrices, respectively.

Refinement by non-rigid registration

After we obtain the coarse segmentation as a result of template-to-image matching with low-dimensional transforms, a non-rigid registration with a high-dimensional transform is performed to achieve more freedom of deformation in order to refine the segmentation. We adopt B-spline as the deformation field and Mutual Information as the similarity metric for the registration between the training image and the target image. The non-rigid registration method is basically the same as free-form deformation (FFD) (Rueckert et al., 1999) but other non-rigid registration can also be used in place of FFD. B-spline based registration was found to have better performance than affine or fluid registrations when applied to smaller cerebral structures (Crum et al., 2004).

We perform the non-rigid registration on a structure-by-structure basis instead of matching all the structures together. This allows us to fine-tune the shape of each structure and keep the computational costs low. The initialization of the registration is acquired from the coarse segmentation from the previous step. In the previous step, we obtain the resultant transform matrices (similarity and affine) so that we can deform the corresponding structure in a region of interest (ROI) of the training image accordingly. This deformed ROI image is served as the moving image for the non-rigid registration. The ROI with the same size and position in the target image is served as the fixed image. After the non-rigid registration is done, we deform the coarse segmentation according to the deformation field resulting from the registration and thus obtain the final segmentation of each structure.

Estimation of weights

In the energy function of MDT (Eq. (1)), there are six parameters (α, β, γ, δ, ζ, η) that can be adjusted. From the angle of energy minimization, one of the parameters can be fixed and the other five are free parameters. We propose an automatic method to estimate the free parameters based on the prior multi-object template. Note that the template is obtained from a training gray level image and its ground truth is known. We can calculate the six terms (e_1 to e_6) in the

objective energy function based on the 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,

$$\begin{aligned}
&\alpha:\beta = e_2\left(p_i^t|I\right):|e_1\left(p_i^t|I\right)|,\\
&\alpha:\gamma = e_3\left(p_i^t|I_E\right):|e_1\left(p_i^t|I\right)|,\\
&\alpha:\delta = e_4\left(p_i^t|p_j^t\right):|e_1\left(p_i^t|I\right)|,\\
&\alpha:\zeta = e_5\left(p_i^t|p_j^t\right):|e_1\left(p_i^t|I\right)|,\\
&\alpha:\eta = 1:|e_1\left(p_i^t|I\right)|,
\end{aligned} \tag{14}$$

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 importance of each term is roughly the same, we try and equalize the weights of each term according to the values obtained from the true poses. For example, based on the ground truth, the value of mean gradient of one part (e_1) may be much larger than the value of intensity variance of that part (e_2). If the weights of e_1 and e_2, α and β, are unadjusted (i.e., both are set to 1), then e_1 will dominate e_2 in the optimization process. However, using the above estimation strategy, it can be seen that all six terms have similar effect in the minimization process. 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 the ventricles in the ground truth.

Implementation issues

The proposed framework can be readily implemented by various Insight ToolKit (ITK) algorithms, which are available in open source codes.[2] For example, we implemented the main framework of the proposed method by means of a model-based registration in ITK, i.e., a model-to-image registration. We used binary images from the training ground truth to represent the model, which depicts the shape of each structure. ITK offers various transforms to deform model/template and we chose 3D similarity transform and affine transform. The B-spline based registration was implemented by the combination of B-spline deformable transform, LBFGSB optimizer and Mutual Information metric in ITK. ITK also provides some simple algorithms that can be used in our framework, such as neighborhood connected region growing for the ventricle segmentation and Canny edge detector for the edge map construction. More details about the above-mentioned algorithms can be found in the ITK Software Guide (Ibanez and Schroeder, 2005).

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., the left and right caudate nuclei, the putamina, and the thalami. These six deep brain structures are often studied in brain anatomy (Cootes and Taylor, 2001; Pohl et al., 2006, 2007b; Sonka et al., 1996; Tsai et al., 2004; Yang et al., 2004; Tu et al., 2008; Zhou and Rajapakse, 2005; Iosifescu et al., 1997; Fischl et al., 2002; Yang and Duncan, 2004). However, they are difficult to segment automatically due to their small sizes and blurry boundaries. For a normal person, the largest structure among these six structures is the thalamus and has the volume around 6000–8000 mm^3 (viz. 6–8 ml), as compared with the normal adult human brain volume, which is over 1200 ml. The image volumes (number of voxels multiplied by voxel size) in the experiments are 8815–12,583 ml. Due to their small sizes, it is challenging to detect and locate the deep brain structures in a 3D image.

The testing data set was obtained from the Internet Brain Segmentation Repository (IBSR),[3] which includes 18 high-resolution brain T1-weighted MR scans with expert segmentations of the internal structures. The MR brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available publicly. The resolution of each scan is $256\times256\times128$ voxels. The voxel size is $0.9375\times0.9375\times1.5$ mm^3 for 8 scans, $1.0\times1.0\times1.5$ mm^3 for 6 scans and $0.8371\times0.8371\times1.5$ mm^3 for 4 scans. The T1-weighted volumetric images have been positionally normalized into the Talairach orientation (rotation only) and have been processed by the CMA bias field correction routines. More information about the data sets can be found in the IBSR website.

We took the ground truth of one data set in the database as the shape template and treated the other 17 sets as the target (testing) image volumes. 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 since the data were obtained from different subjects.

We implemented the proposed framework with 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 1 to 2 hours on a 2.13 GHz CPU with 1 GB memory.

The main parameters of the proposed method can be set according to the methods proposed in the Estimation of weights subsection. The variance and threshold of the Canny edge detector were fixed at 1.0 and 2.0, respectively. The parameter K for partial Hausdorff distance was set to $0.1*||B_i||$. We used a third order B-spline with grid size of $5\times5\times5$. τ was set to 5%. Note that all the model parameters are set once for all experiments and the results of different input data are based on the same set of parameters.

We have compared our results with the expert segmentations available from the IBSR database, which are considered the "gold standard". The extent of the overlapping between the experimental results and the ground truth was measured by a widely used index, Dice similarity coefficient (DSC) (Dice, 1945). It is defined as,

$$\mathrm{DSC} = \frac{2||A\cap B||}{||A|| + ||B||}, \tag{15}$$

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 between two point sets. When it equals 1, the two sets are perfectly overlapped. The smaller the measure is, the less overlapping the two sets have. The range of DSC is [0, 1].

Since we adopt B-spline registration at the last stage (i.e., refinement) of the proposed method, we are interested in the results of direct application of B-spline registration on the same data sets. In direct B-spline registration, due to the local deformation of B-spline, we register the gray level training image with the testing image first by the global (rigid and affine) transforms. We then apply the same B-spline registration as described in the proposed method (see the Refinement by non-rigid registration subsection). The comparison results are plotted in Fig. 4. Apart from the direct B-spline results and the final results of the proposed method, we also show the intermediate results before refinement in the proposed method. From this comparison, it is observed that the results of the proposed method are significantly better than those of the direct B-spline registration especially for smaller structures like the caudate nucleus

[2] Insight ToolKit, available online at: www.itk.org.

[3] Internet Brain Segmentation Repository, available online at: www.cma.mgh.harvard.edu/ibsr/.

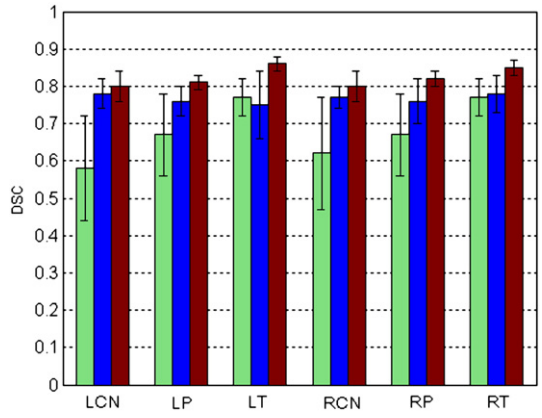

Fig. 4. The accuracy comparison of segmentation results in terms of Dice similarity coefficient (DSC). The average results obtained from the direct B-spline registration, intermediate and final segmentations of the proposed method are represented by green, blue, and red bars, respectively. The standard deviations are shown in error bars on top. The test data is 17 sets in IBSR.

and the putamen. This is because direct B-spline method registers two whole brain scans locally and may not deform effectively a small local structure without distinct features when significant deformation occurs across different subjects. On the contrary, the proposed method targets directly the small structures and focuses on the refinement within the region of interest. As compared with the intermediate results, the final results also improve the segmentation accuracy thanks to the B-spline registration with a good initialization. The B-spline registration allows more freedom of the shape deformation than the rigid or affine registration in the previous stages. Therefore, it can match two structures more precisely.

We found that the mean DSC and the standard deviation for the ventricles were 0.87 and 0.04, respectively. Segmentation of the ventricles is important when we study the volume of the ventricles in brain diseases (Chou et al., 2008). Since the current work focuses on deep brain structural segmentation of normal brains, we do not focus on the segmentation results of the ventricles in this paper. We randomly chose nine data sets in the IBSR database to serve as the template one by one and performed the same segmentation experiment on the remaining 17 sets nine times. We can subdivide the result pool by fixing the testing subject or the template. The subdivided results are listed in Tables 1 and 2. Table 1 lists the structure-wise DSC results (means, standard deviations, best cases and worst cases) for the individual testing subjects based on all templates. It is noted that, for each testing subject, it was not used as the template in its corresponding experiments. Table 2 lists the structure-wise DSC results (means, standard deviations, best cases and worst cases) for the individual templates. For each template, it was applied on the remaining 17 testing subjects. Judging from the results listed in the tables, we find that the proposed method is tolerant of different templates and produces accurate segmentations consistently. For a given testing subject, the results based on different templates are similar and within a reasonable variance (standard deviation is not larger than 0.05 for the DSC values). For a given template, the segmentation results of different testing data sets are also similar (standard deviation is not larger than 0.05). From the results listed in Tables 1 and 2, we find that all six types of structures segmented by the proposed method have good overlaps (mean DSC>0.80) with the expert segmentations. When DSC is larger than 0.7, the segmentation result has a good overlap with ground truth for objects of small sizes. We can perform a simple calculation to justify this. Suppose we have a small cubic object with the size of $10\times10\times10$ voxels. If the segmentation is also a 1000-voxel cube but is shifted by 1 voxel along each of the three dimensions, the DSC for this segmentation is 0.729. In this case, the segmentation is close to the ground truth but the DSC is near 0.73. Similarly, the volume of the deep brain structures is small and a small misalignment with ground truth will cause the DSC to drop dramatically. In the literature, the segmentation of even larger objects (e.g. brain tissues) with DSC over 0.7 is regarded as good (Cuadra et al., 2005).

In addition to the quantitative evaluation, we present a visual inspection of the experimental results. In Fig. 5, we show the results of direct B-spline segmentations, intermediate and final segmentations of the proposed method. Two data sets are chosen. One (left) is above average and the other (right) is below average in terms of the resulting accuracy of the proposed method. Although the intensity profile, shape appearance of the structures and location of the brain are different among these images, the three methods are able to perform reasonable segmentations. As a whole, in terms of 3D morphology, the three results are similar because they all originate from the same binary template (prior shape), and both the direct and refining B-spline use Mutual Information as similarity measure. The

Table 1
Each row lists the structure-wise DSC results (mean ± standard deviation, the best and worst cases) of one fixed testing subject based on all other templates, except for itself as the template.

Testing	RCN			LCN			LP			RP			LT			RT		
	Mean ± SD	Best	Worst	Mean ± SD	Best	Worst	Mean ± SD	Best	Worst	Mean ± SD	Best	Worst	Mean ± SD	Best	Worst	Mean ± SD	Best	Worst
1	0.77 ± 0.05	0.82	0.70	0.79 ± 0.02	0.81	0.73	0.82 ± 0.01	0.84	0.79	0.83 ± 0.04	0.89	0.78	0.84 ± 0.0 2	0.87	0.81	0.84 ± 0.03	0.89	0.78
2	0.82 ± 0.03	0.86	0.77	0.81 ± 0.02	0.84	0.78	0.81 ± 0.02	0.83	0.78	0.83 ± 0.03	0.87	0.79	0.85 ± 0.01	0.87	0.84	0.84 ± 0.03	0.90	0.81
3	0.80 ± 0.06	0.87	0.69	0.79 ± 0.04	0.89	0.73	0.81 ± 0.02	0.85	0.77	0.83 ± 0.01	0.85	0.81	0.86 ± 0.02	0.89	0.84	0.83 ± 0.03	0.87	0.77
4	0.81 ± 0.03	0.87	0.77	0.80 ± 0.02	0.83	0.77	0.81 ± 0.03	0.86	0.78	0.81 ± 0.03	0.84	0.77	0.86 ± 0.02	0.88	0.82	0.85 ± 0.04	0.91	0.79
5	0.83 ± 0.05	0.92	0.75	0.81 ± 0.03	0.85	0.77	0.83 ± 0.04	0.90	0.79	0.81 ± 0.04	0.88	0.75	0.86 ± 0.02	0.88	0.83	0.87 ± 0.03	0.91	0.82
6	0.80 ± 0.04	0.85	0.73	0.79 ± 0.03	0.87	0.76	0.81 ± 0.02	0.84	0.78	0.83 ± 0.03	0.89	0.79	0.86 ± 0.02	0.89	0.84	0.85 ± 0.02	0.88	0.82
7	0.80 ± 0.04	0.86	0.73	0.79 ± 0.04	0.83	0.72	0.82 ± 0.03	0.86	0.78	0.80 ± 0.03	0.85	0.74	0.86 ± 0.02	0.90	0.83	0.84 ± 0.03	0.88	0.78
8	0.79 ± 0.03	0.82	0.75	0.78 ± 0.02	0.81	0.74	0.81 ± 0.02	0.84	0.79	0.84 ± 0.02	0.86	0.81	0.85 ± 0.03	0.91	0.80	0.84 ± 0.02	0.86	0.81
9	0.80 ± 0.03	0.84	0.75	0.81 ± 0.03	0.84	0.77	0.80 ± 0.01	0.83	0.79	0.80 ± 0.04	0.88	0.77	0.85 ± 0.02	0.87	0.82	0.83 ± 0.04	0.89	0.75
10	0.79 ± 0.04	0.84	0.73	0.80 ± 0.02	0.84	0.78	0.79 ± 0.04	0.85	0.72	0.82 ± 0.03	0.87	0.79	0.86 ± 0.01	0.87	0.84	0.84 ± 0.04	0.92	0.78
11	0.79 ± 0.04	0.85	0.73	0.80 ± 0.03	0.86	0.77	0.82 ± 0.03	0.85	0.77	0.83 ± 0.02	0.87	0.79	0.86 ± 0.02	0.88	0.83	0.84 ± 0.02	0.86	0.81
12	0.78 ± 0.04	0.83	0.72	0.80 ± 0.02	0.84	0.78	0.81 ± 0.02	0.84	0.78	0.82 ± 0.02	0.85	0.79	0.84 ± 0.02	0.87	0.82	0.85 ± 0.02	0.88	0.83
13	0.78 ± 0.04	0.85	0.73	0.80 ± 0.04	0.86	0.73	0.82 ± 0.03	0.87	0.78	0.84 ± 0.04	0.90	0.79	0.85 ± 0.03	0.88	0.82	0.85 ± 0.02	0.88	0.82
14	0.83 ± 0.04	0.88	0.75	0.80 ± 0.04	0.87	0.75	0.82 ± 0.03	0.86	0.78	0.83 ± 0.03	0.86	0.79	0.86 ± 0.03	0.92	0.81	0.84 ± 0.03	0.89	0.78
15	0.80 ± 0.03	0.84	0.74	0.80 ± 0.03	0.83	0.72	0.81 ± 0.02	0.85	0.77	0.83 ± 0.03	0.88	0.78	0.87 ± 0.02	0.89	0.83	0.85 ± 0.02	0.88	0.82
16	0.80 ± 0.04	0.90	0.76	0.79 ± 0.03	0.86	0.76	0.81 ± 0.03	0.87	0.78	0.80 ± 0.03	0.86	0.75	0.85 ± 0.02	0.88	0.83	0.83 ± 0.03	0.90	0.81
17	0.79 ± 0.03	0.83	0.75	0.79 ± 0.03	0.85	0.75	0.81 ± 0.02	0.85	0.79	0.82 ± 0.03	0.86	0.75	0.85 ± 0.02	0.88	0.82	0.86 ± 0.03	0.91	0.82
18	0.82 ± 0.03	0.88	0.78	0.80 ± 0.02	0.84	0.77	0.82 ± 0.01	0.85	0.80	0.81 ± 0.03	0.85	0.76	0.85 ± 0.02	0.88	0.83	0.84 ± 0.02	0.87	0.81
Average	0.80 ± 0.04	0.87	0.72	0.80 ± 0.02	0.84	0.76	0.81 ± 0.02	0.86	0.77	0.82 ± 0.03	0.88	0.77	0.85 ± 0.02	0.89	0.82	0.84 ± 0.03	0.89	0.79

LCN(RCN) = left(right) caudate nucleus. LP(RP) = left(right) putamen. LT(RT) = left(right) thalamus.

Table 2
Each row lists the structure-wise DSC results (mean ± standard deviation, the best and worst cases) of the remaining 17 testing subjects based on one fixed template.

Template	RCN			LCN			LP			RP			LT			RT		
	Mean ± SD	Best	Worst	Mean ± SD	Best	Worst	Mean ± SD	Best	Worst	Mean ± SD	Best	Worst	Mean ± SD	Best	Worst	Mean ± SD	Best	Worst
1	0.80 ± 0.04	0.85	0.73	0.80 ± 0.04	0.85	0.72	0.81 ± 0.02	0.85	0.79	0.82 ± 0.02	0.88	0.79	0.86 ± 0.02	0.89	0.83	0.85 ± 0.02	0.88	0.81
2	0.81 ± 0.03	0.87	0.74	0.81 ± 0.05	0.89	0.72	0.82 ± 0.05	0.90	0.72	0.83 ± 0.04	0.89	0.74	0.87 ± 0.03	0.92	0.82	0.84 ± 0.04	0.91	0.75
3	0.79 ± 0.03	0.84	0.73	0.79 ± 0.02	0.83	0.76	0.82 ± 0.02	0.86	0.77	0.83 ± 0.02	0.89	0.79	0.85 ± 0.02	0.88	0.82	0.83 ± 0.02	0.86	0.78
4	0.79 ± 0.04	0.85	0.70	0.79 ± 0.01	0.82	0.77	0.82 ± 0.02	0.85	0.78	0.81 ± 0.04	0.88	0.75	0.85 ± 0.02	0.88	0.82	0.84 ± 0.02	0.89	0.77
5	0.80 ± 0.05	0.90	0.72	0.79 ± 0.02	0.82	0.77	0.81 ± 0.02	0.84	0.77	0.82 ± 0.03	0.90	0.78	0.86 ± 0.02	0.88	0.80	0.84 ± 0.02	0.88	0.80
6	0.80 ± 0.04	0.86	0.74	0.79 ± 0.02	0.82	0.76	0.82 ± 0.02	0.85	0.79	0.81 ± 0.03	0.86	0.75	0.84 ± 0.02	0.87	0.81	0.84 ± 0.03	0.89	0.78
7	0.80 ± 0.05	0.92	0.69	0.79 ± 0.02	0.82	0.75	0.80 ± 0.02	0.83	0.77	0.82 ± 0.03	0.88	0.77	0.86 ± 0.02	0.89	0.83	0.86 ± 0.03	0.91	0.81
8	0.81 ± 0.05	0.88	0.72	0.81 ± 0.03	0.86	0.77	0.81 ± 0.02	0.86	0.78	0.81 ± 0.03	0.87	0.75	0.85 ± 0.02	0.88	0.83	0.84 ± 0.02	0.87	0.78
9	0.80 ± 0.03	0.85	0.74	0.81 ± 0.01	0.84	0.79	0.81 ± 0.02	0.85	0.78	0.83 ± 0.03	0.89	0.77	0.85 ± 0.02	0.89	0.83	0.86 ± 0.03	0.92	0.79
Average	0.80 ± 0.04	0.86	0.74	0.80 ± 0.03	0.85	0.75	0.81 ± 0.02	0.85	0.78	0.82 ± 0.03	0.87	0.78	0.85 ± 0.02	0.88	0.83	0.84 ± 0.03	0.89	0.80

LCN(RCN) = left(right) caudate nucleus. LP(RP) = left(right) putamen. LT(RT) = left(right) thalamus.

intermediate results have the smoothest boundaries. The final results are the closest to the ground truth because they are not over-smoothed as compared with the intermediate ones and they preserve better relative poses than the direct B-spline method. From 2D inspection, we can see that the direct B-spline method cannot simultaneously locate all the structures precisely.

Discussion

The proposed framework enjoys several significant advantages over some existing multi-object segmentation methods.

1. The training burden is lessened. Unlike the existing methods (Sonka et al., 1996; Duta and Sonka, 1998; Ciofolo and Barillot, 2005; Pohl et al., 2006, 2007b; Tsai et al., 2004; Yang et al., 2004; Tu et al., 2008; Yang and Duncan, 2004; Gouttard et al., 2007; Cootes and Taylor, 2001; Zhou and Rajapakse, 2005; Powell et al., 2008; Khan et al., 2008) which require a number of ground truths in order to perform sufficient training before running the segmentation, our proposed method requires less training workload by using the detection strategy in the MDT-based model. The shape template, as prior knowledge in the proposed framework, stems from one example and hence the training set is a singleton. This advantage can abate the manual work in clinical applications. If there is no ground truth available for a certain population, the proposed method is quick and convenient to apply because only one manual segmentation is needed.
2. Initialization is unnecessary. In our proposed method, there is no need to initialize the objects at some specific locations because our algorithm automatically detects the root object by matching the root in the MDT with the result of brain tissue segmentation. The proposed method can detect the target structures within the brain images without the need of initialization. In contrast, the level set based methods (Duta and Sonka, 1998; Ciofolo and Barillot, 2005; Cootes and Taylor, 2001; Tsai et al., 2004; Yang et al., 2004; Yang and Duncan, 2004) require the initial curves to be placed inside or very close to the target structures. Registration-based methods (Pohl et al., 2006, 2007b; Fischl et al., 2002; Iosifescu et al., 1997; Khan et al., 2008; Heckemann et al., 2006; Gouttard et al., 2007; Gousias et al., 2008) can obviate the problem of contour initialization but unfortunately, it is difficult to build an informative and effective atlas for brain structures and warp the atlas to input brain images using non-rigid registration.
3. It is applicable to missing boundaries. In addition to the edge/region information, the proposed framework further considers the interactions between objects. Under the constraint of MDT, the thalami are assigned to leaf objects and after their parent objects are located, they can be segmented successfully. The performance of the active contour methods (Tsai et al., 2004; Yang et al., 2004; Yang and Duncan, 2004) may be adversely affected when these methods are employed to segment the thalamus, whose boundary is difficult to locate even by human eyes.
4. It enjoys greater flexibility. We can change the current terms in the evaluation energy, the transform matrices or the B-spline non-rigid registration to other alternatives. For example, we can replace the current shape descriptor (dense binary images) with other complicated deformable models, such as simplex mesh (Delingette, 1999), elastic models (Pitiot et al., 2002) and spherical wavelets (Nain et al., 2007).

It is very difficult to compare the proposed method with other related works experimentally for several reasons. The source codes of the related works are usually not available to public. Moreover, it is hard to implement the related methods independently. Even though the related methods are perfectly replicated, there are still some critical problems such as how to perform the relevant intensive training, estimate and tune a series of parameters, or construct a complicated brain atlas. These problems are challenging and hence we only compare the proposed method with other related works conceptually if the quantitative results on the same database are not reported in those works.

The work of Ciofolo and Barillot (2005) reported quantitative results of the segmentation of the deep brain structures on the IBSR database. It reported the DSC results for the caudate nucleus, the putamen and the thalamus as 0.65, 0.70 and 0.77, respectively with 8-minute computational time on average. Obviously the proposed method outperformed their work as the proposed method has a more flexible architecture to model the interrelation among neuro-anatomical structures to deal with the small and subtle objects in deep brains than the fuzzy control in Ciofolo and Barillot (2005).

In a recent paper (Powell et al., 2008), the authors reported highly accurate segmentation results of deep brain structures (caudate: 0.91, putamen: 0.92, thalamus: 0.93) on a database which is not available publicly. However, their method required a sufficient and time-consuming training process before the artificial neural network or support vector machine could be employed. The training took one day to complete. Moreover, the high accuracy obtained by their method was achieved by the utilization of multiple channels in MRI (i.e., T1 and T2), which could make easy the segmentation of deep brain structures. In the proposed method, we only rely on the information obtained from the T1-weighted MR images. Another difference between the work of Powell et al. (2008) and ours is that the former segmentation is based on the result of tissue segmentation. We do not perform tissue segmentation prior to structure segmentation to avoid the effect of imperfectness in the tissue segmentation.

In another paper on segmentation of subcortical structures (Khan et al., 2008), the reported results for healthy right caudate nucleus, putamen and thalamus (a different database from IBSR) were 0.81, 0.83 and 0.86, respectively. The accuracies were similar to ours but

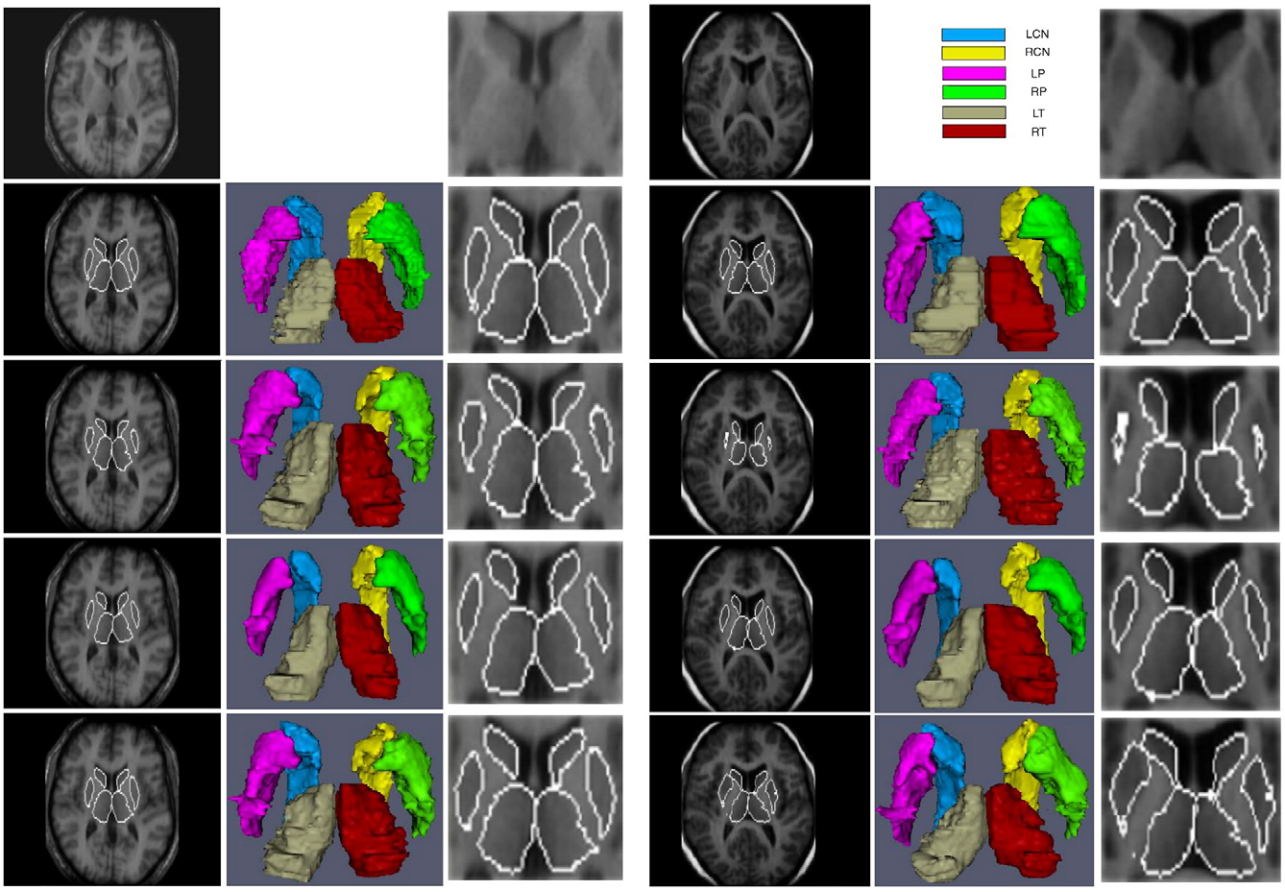

Fig. 5. Two sets of data (the first row), their ground truth (the second row), results obtained from the direct B-spline registration (the third row), intermediate results (the fourth row) and final results (the fifth row) using the proposed method. For each ground truth or result, we show a 2D slice, 3D surface reconstruction and 2D close-ups.

their computational cost (10 to 15 h for FreeSurfer, and over one hour for LDDMM using multi-processors) was significantly higher than ours (1 to 2 h). Moreover, as the initialization, FreeSurfer required a training set of 41 manually labeled brains to construct the atlas. The proposed method, on the other hand, requires only one training brain.

Some of the related methods were tested on some of the structures studied here. For example, the work of Gousias et al. (2008) reported the accuracies for left caudate nucleus from 0.87 to 0.90 based on a different database from ours. The method in Gouttard et al. (2007) obtained the accuracies for caudate and putamen at 0.76 and 0.78 based on IBSR data sets. Heckemann et al. reported the accuracies for right caudate nucleus and right thalamus at 0.84 and 0.86 based on a private database. Other related works on multi-object segmentation (Litvin and Karl, 2005; Tsai et al., 2004; Yang et al., 2004) did not report the quantitative results of the segmentation of deep brain structures. One advantage of the proposed method over those methods is that we circumvent the intractable blurry boundaries of brain structures, such as the thalami, by emphasizing the relative position to other deep brain structures. On the contrary, the level set based methods used in Litvin and Karl (2005), Tsai et al. (2004), and Yang et al. (2004) may not be able to perform the segmentation of the thalami when facing objects with weak edges and gradients. Even the use of the idea of "active contours without edges" (Chan and Vese, 2001) cannot significantly improve the methods (Yang et al., 2004; Yang and Duncan, 2004) to find the missing boundaries of the thalami, which are located in the complicated 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.

The proposed method can be particularly useful in the case that the ground truths of a specific data set are not available. The data set may be obtained from a population with special characteristics related to aging or a specific kind of pathology. The ground truths from the other populations can be inappropriate to serve as the training examples for this data set due to the absence of the distinctive features. For the proposed method, only a single ground truth for one datum in this data set is needed to obtain from the expert. The proposed method offers an alternative to the conventional methods requiring many training examples.

The limitation of the proposed method lies in the template originating from a single manual segmentation. It is possible that the singularity in the manual segmentation can propagate to the automated segmentation. In this sense, the template trained from multiple manual segmentations is advantageous since the anomaly from a single subject can be smoothed out. When a larger training set is available, it may be beneficial to perform an atlas selection to make use of all training information (Rohlfing et al., 2004). Alternatively, we can fuse the segmentation results of the test image which are obtained from the registrations with each training image (Heckemann et al., 2006). Further improvement on the consistency and accuracy of the current method can be achieved by the decision fusion or atlas selection strategy.

Summary

This paper has proposed a novel framework for segmentation of multiple brain structures in a complex background. The framework is integrated with detection techniques and the relationship among multiple structures (objects) is organized by using a Markov

dependence tree. We have validated the proposed method on the medical segmentation of deep brain structures and showed its accuracy and efficiency. As compared with other related works, the proposed method has the following advantages: automatic initialization, training from a single example, segmentation of objects with missing boundary, and flexible multi-object segmentation framework. The current framework is general and can be naturally applied to other multi-object detection and segmentation applications in the future. Future work may include the incorporation of more brain structures, upgrade to more interactions in the graphical model and tests across different databases.

Acknowledgment

The authors would like to acknowledge the financial support from the K S Lo Foundation, Hong Kong.

References

Barra, V., Boire, J., 2001. Automatic segmentation of subcortical brain structures in mr images using information fusion. IEEE Trans. Med. Imag. 20 (7), 549–558.
Chan, T., Vese, L., 2001. Active contours without edges. IEEE Trans. on Image Process. 10 (2), 266–277.
Chou, Y., Lepore, N., de Zubicaray, G., Carmichael, O., Becker, J., Toga, W., Thompson, P., 2008. Automated ventricular mapping with multi-atlas fluid image alignment reveals genetic effects in Alzheimer's disease. NeuroImage 40, 615–630.
Ciofolo, C., Barillot, C., 2005. Brain segmentation with competitive level sets and fuzzy control. Information Processing in Medical Imaging, Vol. LNCS 3565, pp. 333–344.
Cootes, T., Taylor, C., 2001. Statistical models of appearance for medical image analysis and computer vision. SPIE Medical Imaging, Vol. 4322, pp. 236–248.
Crum, W., Rueckert, D., Jenkinson, M., Kennedy, D., Smith, S., 2004. A framework for detailed objective comparison of non-rigid registration algorithms in neuroimaging. MICCAI, pp. 679–686.
Cuadra, M., Cammoun, L., Butz, T., Cuisenaire, O., Thiran, J., 2005. Comparison and validation of tissue modelization and statistical classification methods in t1weighted mr brain images. IEEE Trans. Med. Imag. 24, 1548–1565.
Delingette, H., 1999. General object reconstruction based on simplex meshes. Int. J. Comput. Vis. 32 (2), 111–146.
Dice, L., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297–302.
Duta, N., Sonka, M., 1998. Segmentation and interpretation of mr brain images: an improved active shape model. IEEE Trans. Med. Imag. 17 (6), 1049–1062.
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., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.
Gousias, I., Rueckert, D., Heckemann, R., Dyet, L., Boardman, J., Edwards, A., Hammers, A., 2008. Automatic segmentation of brain mris of 2-year-olds into 83 regions of interest. NeuroImage 40, 672–684.
Gouttard, S., Styner, M., Joshi, S., Smith, R., Cody, H., Gerig, G., 2007. Subcortical structure segmentation using probabilistic atlas priors. SPIE Medical Imaging, Vol. 6512, pp. 65111j–65122j.
Grenander, U., Miller, M., 1998. Computational anatomy: an emerging discipline. Q. Appl. Math. 56, 617–694.
Heckemann, R., Hajnal, J., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain mri segmentation combining label propagation and decision fusion. NeuroImage 33, 115–126.
Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W., 1993. Comparing images using the hausdorff distance. IEEE Trans. Pattern Anal. Mach. Intell. 15 (9), 850–863.
Ibanez, L., Schroeder, W., 2005. The ITK Software Guide. Kitware, Inc.
Iosifescu, D., Shenton, M., Warfield, S., Kikinis, R., Dengler, J., Jolesz, F., McCarley, R., 1997. An automated registration algorithm for measuring mri subcortical brain structures. NeuroImage 6, 13–25.
Kandel, E., Schwartz, J., Jessell, T., 2000. Principles of Neural Science, 4th Edition. McGraw-Hill, New York.
Khan, A., Wang, L., Beg, M., 2008. Freesurfer-initiated fully-automated subcortical brain segmentation in mri using large deformation diffeomorphic metric mapping. NeuroImage 41, 735–746.
Litvin, A., Karl, W., 2005. Coupled shape distribution based segmentation of multiple objects. Information Processing in Medical Imaging, Vol. LNCS 3565, pp. 345–356.
Miller, M., Christensen, G., Amit, Y., Grenander, U., 1993. Mathematical textbook of deformable neuroanatomies. Proc. Natl. Acad. Sci. U. S. A. 90, 11944–11948.
Nain, D., Haker, S., Bobick, A., Tannenbaum, A., 2007. Multiscale 3-d shape representation and segmentation using spherical wavelets. IEEE Trans. Med. Imag. 26 (4), 598–618.
Packard, M., Knowlton, B., 2002. Learning and memory functions of the basal ganglia. Annu. Rev. Neurosci. 25, 563–593.
Pitiot, A., Toga, A., Thompson, P., 2002. Adaptive elastic segmentation of brain mri via shape-model-guided evolutionary programming. IEEE Trans. Med. Imag. 21 (8), 910–923.
Pohl, K., Fisher, J., Grimson, W., Kikinis, R., Wells, W., 2006. A bayesian model for joint segmentation and registration. NeuroImage 31, 228–239.
Pohl, K., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R., Kikinis, R., Grimson, W., Shenton, M., Wells, W., 2007a. A hierarchical algorithm for mr brain image parcellation. IEEE Trans. Med. Imag. 26, 1201–1212.
Pohl, K., Fisher, J., Bouix, S., Shenton, M., MacCarley, R., Grimson, E., Kikinis, R., Wells, W., 2007b. Using the logarithm of odds to define a vector space on probabilistic atlases. Med. Image Anal. 11 (5), 465–477.
Powell, S., Magnotta, V., Johnson, H., Jammalamadaka, V., Pierson, R., Andreasen, N., 2008. Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures. NeuroImage 39, 238–247.
Press, W., Flannery, B., Teukolsky, S., Vetterling, W., 1992. Numerical Recipes in C2nd Edition. Cambridge University Press.
Rohlfing, T., Brandt, R., Menzel, R., Maurer, C., 2004. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage 21, 1428–1442.
Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D., 1999. Non-rigid registration using free-form deformations: application to breast mr images. IEEE Trans. Med. Imag. 18 (8), 712–721.
Sonka, M., Tadikonda, S., Collins, S., 1996. Knowledgebased interpretation of mr brain images. IEEE Trans. Med. Imag. 15 (4), 443–452.
Svarer, C., Madsen, K., Hasselbalch, S., Pinborg, L., Haugbol, S., Frokjar, V., Holm, S., Paulson, O., Knudsena, G., 2005. Mr-based automatic delineation of volumes of interest in human brain pet images using probability maps. NeuroImage 24, 969–979.
Tsai, A., Wells, W., Tempany, C., Grimson, E., Willsky, A., 2004. Mutual information in coupled multi-shape model for medical image segmentation. Med. Image Anal. 8 (4), 429–445.
Tu, Z., Narr, K., Dollar, P., Dinov, I., Thompson, P., Toga, A., 2008. Brain anatomical structure segmentation by hybrid discriminative/generative models. IEEE Trans. Med. Imag. 27 (4), 495–508.
Yang, J., Duncan, J., 2004. 3d image segmentation of deformable objects with joint shape-intensity prior models using level sets. Med. Image Anal. (8), 285–294.
Yang, J., Staib, L., Duncan, J., 2004. Neighbor-constrained segmentation with level set based 3-d deformable models. IEEE Trans. Med. Imag. 23 (8), 940–948.
Zhou, J., Rajapakse, J., 2005. Segmentation of subcortical brain structures using fuzzy templates. NeuroImage 28, 915–924.