HHS Public Access

Author manuscript

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2017 June 15.

Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2017 February 11; 10133: . doi:10.1117/12.2254874.

Machine learning in a graph framework for subcortical segmentation

Zhihui Guo[*,a,c], **Satyananda Kashyap**[b,c], **Milan Sonka**[b,c], and **Ipek Oguz**[d]

[a]Dept. of Biomedical Engineering, Univ. of Iowa, Iowa City, IA, USA 52242

[b]Dept. of Electrical & Computer Engineering, Univ. of Iowa, Iowa City, IA, USA 52242

[c]Iowa Institute for Biomedical Imaging, Univ. of Iowa, Iowa City, IA, USA 52242

[d]Dept. of Radiology, University of Pennsylvania, Philadelphia, PA, USA 19104

Abstract

Automated and reliable segmentation of subcortical structures from human brain magnetic resonance images is of great importance for volumetric and shape analyses in quantitative neuroimaging studies. However, poor boundary contrast and variable shape of these structures make the automated segmentation a tough task. We propose a 3D graph-based machine learning method, called LOGISMOS-RF, to segment the caudate and the putamen from brain MRI scans in a robust and accurate way. An atlas-based tissue classification and bias-field correction method is applied to the images to generate an initial segmentation for each structure. Then a 3D graph framework is utilized to construct a geometric graph for each initial segmentation. A locally trained random forest classifier is used to assign a cost to each graph node. The max-flow algorithm is applied to solve the segmentation problem. Evaluation was performed on a dataset of T1-weighted MRI's of 62 subjects, with 42 images used for training and 20 images for testing. For comparison, FreeSurfer and FSL approaches were also evaluated using the same dataset. Dice overlap coefficients and surface-to-surfaces distances between the automated segmentation and expert manual segmentations indicate the results of our method are statistically significantly more accurate than the other two methods, for both the caudate (Dice: 0.89 ± 0.03) and the putamen (0.89 ± 0.03).

Keywords

Segmentation; subcortical structure; magnetic resonance images (MRI); random forest; graph

1. INTRODUCTION

Quantifying volumetric and shape changes of basal ganglia are essential to the study of many brain diseases. For example, striatal atrophy, assessed by volume measurements of the caudate nucleus and the putamen, is one of the primary biomarkers that can be measured from images for Huntington's Disease.[1] As such, accurate and robust quantification of these

* zhihui-guo@uiowa.edu.

structures is crucial for disease characterization, prediction of clinical onset,[2, 3] as well as assessing the effectiveness of therapeutical interventions in clinical trials.

However, some of the basal ganglia structures have a poor and variable MRI intensity contrast,[4] which makes it difficult to automatically segment them accurately and robustly. Several methods have been proposed for this task, including probability-atlas approaches,[5] multi-atlas registration segmentation methods[6] and artificial neural networks.[7] These methods can often be sensitive to local optima. We propose a novel method, LOGISMOS-RF, based on a machine learning algorithm in the LOGISMOS framework[8] (layered optimal graph image segmentation of multiple objects and surfaces) to segment pairs of the caudate and the putamen. LOGISMOS graph segmentation framework provides the globally optimal solution and is thus insensitive to image noise and weak boundaries. LOGISMOS has been successfully applied to segmentation tasks suffering from weak contrast, such as knee cartilage segmentation,[8, 9] as well as in various neuroimaging applications for both human and rat brains.[10–13] However, the large variability in the appearance of subcortical structures makes it difficult to segment these structures by relying on the image gradient strength alone, and a more sophisticated appearance model is needed. Random forest (RF) classifiers[14] work robustly by aggregating the output of multiple decision trees based on nonlinear relationships and can generate a probability map, which is promising for providing suitable costs for LOGISMOS graph nodes.

In contrast with earlier approaches using machine learning in a graph-based segmentation context,[8, 9] we divided the mesh surface of the initial segmentation into several regions and trained a random forest classifier for each region, to reduce the intensity entropy of the training data for each random forest classifier. Additionally, while the method proposed by Oguz et al.[12] also focuses on the graph-based subcortical segmentation, that approach is based on multi-atlas label fusion techniques which require substantial computational time.

In brief, LOGISMOS-RF is a novel approach based on the success of LOGISMOS graph framework and RF classifiers to accurately and robustly segment the caudate nucleus and putamen.

2. METHODS

We propose an automated method, LOGISMOS-RF, to segment left and right caudate nucleus and putamen from T1-weighted human brain MRI. LOGISMOS-RF uses a machine learning approach in the LOGISMOS graph-based framework. Fig. 1 shows the workflow of the segmentation algorithm. The first step is normalization of the intensity range and voxel size of the images to reduce variability, both in the training and the testing phases. In the training phase, after image normalization, each mesh surface of the manual segmentations is parcellated into multiple regions. The designed features are extracted to train a RF classifier for each region so that each region has a corresponding RF. From the training set, a mean shape model for each structure is also generated. In the testing phase, the initial segmentation is obtained from the mean shape model generated from the training set and the coarse pixel classification results. The initial segmentation provides a shape prior, based on which the graph is constructed. Then the mesh surface of the initial segmentation is

parcellated with the same rule in the training phase. The previously trained RFs are used to compute the classification probabilities, which determine the graph costs. This segmentation problem is solved within a graph framework to obtain the final surfaces.

2.1 Training

The training step has two aims: 1) train RF classifiers used to provide accurate graph costs during testing, and 2) generate a mean shape model for each structure, which helps create the initial segmentation during testing.

2.1.1 Image Normalization—The goal of the normalization is to reduce the variation of image appearance among different subjects. The normalization includes the intensity range and voxel size. The intensity extremes are removed by truncating the lower and upper 5% of the intensity values, to reduce the effect of extreme intensity values and provide a more uniform intensity distribution for each structure. The remaining inner 90% of the intensity range is rescaled to 0-4096 linearly. Finally, the images are resampled to $1 \times 1 \times 1\ mm^3$ resolution using linear interpolation.

2.1.2 Surface Parcellation—The anatomy surrounding deep brain structures is quite heterogeneous; thus, the intensity profiles near the structure surface can be vastly different for different regions. Fig. 2 illustrates this in two different surface patches of the caudate, one next to the ventricle while the other away from the ventricle. The image intensity gradient directions are reversed between the two patches (bright-to-dark for the region next to the ventricle, dark-to-bright for the other).

Therefore, instead of learning a single appearance model for the entire structure, we parcellate the surface into smaller regions of interest (ROI's) and learn a location-specific appearance model for each ROI. We use a set of heuristics to parcellate the manual segmentation surface for the putamen into 12 ROI's and for the caudate into 12–18 ROI's.

A surface mesh model of the manual segmentation is obtained by using the marching cubes algorithm. Based on this mesh model, we compute the 3D bounding box for each structure. Both the caudate and the putamen undergo the same basic parcellation procedure first. The caudate parcellation is then further refined.

- The basic parcellation of the mesh surface is achieved in 3 sub-steps. First, we divide the surface into 3 regions along the Anterior-Posterior (A-P) axis. Then, we divide the surface into 2 regions along the Inferior-Superior (I-S) axis. Finally, we divide the surface into two regions along the Right-Left (R-L) axis, according to the sign of surface node's normal direction. This leads to a parcellation of the surface into 12 ROI's ($3 \times 2 \times 2$).
- The caudate is located next to the lateral ventricles, which vary substantially in shape and size, especially in datasets with a large age range. Since the ventricles have a strong and distinct appearance model, identifying the regions of the caudate that are neighboring the ventricles is important. Thus, the caudate parcellation is refined such that additional ROI's are defined within the 6 regions of the caudate that are on the medial side (facing the ventricle). If a surface node

is next to the ventricle, as determined by the manual segmentation, then the node is labeled as a different ROI index. Thus, an original ROI on the side near the ventricle may be divided into two sub-ROI's depending on whether there are nodes on the surface that are next to the ventricle, which results in up to 18 ROI's.

The surface parcellation scheme for the caudate and putamen is illustrated in Fig. 3.

2.1.3 Feature Extraction—For each node on the surface, features are collected in a 5-node patch along the graph column (Fig. 2(a)). The center node is the node that is assigned the extracted features and the remaining 4 nodes are distributed in the two sides along the normal. The distance interval between 2 adjacent nodes is half a voxe. In total, there are 18 features extracted from each patch:

- 5 intensity-based features: the intensity of each node.
- 10 gradient-based features: the signed gradient along the column and the output of the Gaussian gradient magnitude filter ($\sigma = 0.5$) for each node.
- 3 location-based features: the x-y-z coordinates of the center node of the patch.

2.1.4 Training of Multiple RFs—We train an RF classifier for each region with extracted local features for each ROI. In this way, the RFs are location-specific. To train the RFs, both the positive and negative samples are needed from the training set. For each region, the nodes on the surface are treated as the positive samples while the nodes inside or outside the surface are treated as the negative samples. The negative samples are sampled 1 voxel distance far from the surface. For each RF, the number of trees is 100 and the number of randomly selected features for each node of the trees is 3.

2.1.5 Mean Shape Model Generation—The mean shape model is used to generate the initial segmentations. To generate this mean shape model, the T1w images in the training set are registered rigidly with an atlas, and the manual segmentations are averaged to create a probabilistic segmentation in the template space. A threshold of 50% is applied to get a binary segmentation. The marching cubes algorithm is used to generate the surface meshes.

2.2 Testing

2.2.1 Image Normalization—The image normalization is exactly the same here as in the training phase.

2.2.2 Initial Segmentation—To generate initial segmentations for each subject, an atlas-based tissue classification and bias-field correction is first applied to all brain MRI scans using the method in Kim *et al*[15] and Koen *et al*.[16] Voxels that are classified as caudate and putamen are extracted separately to get a coarse segmentation. The initial segmentations for caudate and putamen are created by affinely fitting the mean shape models into these coarse segmentations using the iterative closest points algorithm. The initial segmentations are used as shape priors during the graph construction.

2.2.3 Graph Construction—We construct a geometric node-weighted graph in a LOGISMOS framework.[8] For each vertex on the surface of the initial segmentation, a column, i.e. a stack of graph nodes, is created. We utilize the electric lines of force (ELF), which mimics the Coulomb's law of electrically charged particles, to avoid intersectiions among different columns[8] which may lead to topological defects in the final surface. Each node on the column represents a search location in the image. To cover the potential area that belongs to the desired structure, 50 evenly distributed nodes are chosen along each column with a 0.5 voxel spacing in this study. Two kinds of arcs, intra-column arcs and inter-column arcs are used. The intra-column arcs connect the neighboring nodes of one column and convert the image segmentation task to a max-flow problem on the graph. The inter-column arcs connect neighboring nodes of different columns and encode a smoothness constraint.

The graph structure is illustrated in Fig. 4(a). The smoothness constraint is set as 1 ($\Delta = 1$) between neighboring columns. Fig. 4(b) shows an initial segmentation for the caudate and Fig. 4(c) shows the corresponding graph columns. The costs assigned to each node is determined later in the processing (sec.2.2.6).

2.2.4 Surface Parcellation—The surface parcellation scheme in the testing phase is similar to that in the training phase. However, instead of using the manual segmentation, the initial segmentations are used here. The remaining procedure is the same except the refinement of the parcellation for the caudate. For deciding the caudate regions that are next to the ventricles, we use the cerebrospinal fluid (CSF) segmentation obtained from the tissue classification result in 2.2.2, rather than manual segmentations of the lateral ventricles. If a graph column enters CSF, the corresponding node is labeled as a ventricle-neighboring ROI. Thus, an original ROI may be divided into two sub-ROI's depending on whether there are nodes on the surface with columns entering CSF. This results in up to 18 ROI's for the caudate.

2.2.5 Feature Extraction—The feature design is the same as that in the training. Features are extracted from a sliding window of a 5-node patch along the graph columns.

2.2.6 Probability Map—For the testing, the probability p of each graph node belonging to the final segmentation surface is estimated using the trained location-specific RF classifiers, each node has a probability value. Thus, a probability map is generated. Based on this probability map, the cost of each graph node is determined as $(1 - p)$.

2.2.7 Segmentation—Finally, we solve the segmentation problem by standard min-cut/max-flow optimization algorithm,[17] which provides the globally optimal solution of the objective function under the defined constraints, in low-order polynomial time.

3. EXPERIMENTAL RESULTS

A dataset of T1-weighted MRI scans for 62 subjects (aged 5–96) is used to evaluate the proposed method, LOGISMOS-RF. 49 of these images have a resolution of $1 \times 1 \times 1\ mm^3$, and 13 images have a resolution of $0.9375 \times 0.9375 \times 1.5\ mm^3$. 42 images are chosen for

training purpose while the remaining 20 images are for testing. We calculated Dice coefficient (Dice Coef) and averaged unsigned automated segmentation surface to manual segmentation surface distance (Avg Surf Dist).

For comparison, FreeSurfer[5] (version 5.1) and FSL[4] (version 6.0) were also evaluated using the same dataset. FreeSurfer uses an atlas-based probabilistic approach via Markov Random Fields. FSL implements an Active Appearance Model in a Bayesian framework.

Quantitative results are shown in Table 1. LOGISMOS-RF was significantly more accurate than the two compared methods for both the caudate and the putamen ($p \ll 0.001$), except for the Dice coefficient comparison with FSL which did not reach statistical significance. Note that FSL crashed for 6 subjects in the testing set, only the remaining 14 subjects were used for t-test assessment between FSL and LOGISMOS-RF.

Qualitative results from all three methods are illustrated in Fig. 5, with ground truth and intersection of segmentation overlaid on the same slice.

4. DISCUSSION

The method we proposed in this paper, LOGISMOS-RF, offers statistically significant improvement in segmentation accuracy over the two compared methods, for both caudate and putamen. FreeSurfer gave the worst results among all three methods. We observed that FreeSurfer often suffered from a leaking problem in weak boundaries. FSL offered relatively accurate results, but it crashed for 6 out of 20 subjects in the testing set. Note that the FSL and FreeSurfer are designed to work on T1-weighted images.

LOGISMOS-RF improved the results both quantitatively and qualitatively, based on the observations that it segmented well the thin tail of the caudate which is typically a challenging region, had no obvious leaking problem for weak boundaries and it never crashed. The improved performance is likely based on several factors. First, the LOGISMOS graph framework in this study leverages the mean shape model as a shape prior and ensures a global optimal solution with respect to the defined cost function and smoothness constraints, which makes it very robust to weak boundaries and image noise. Second, after surface parcellation, the training set for each RF classifier is relatively homogeneous and is more specific than a classifier that considers the whole surface, which leads to a suitable cost function for the graph-based segmentation. This location-specific training and robust graph-based optimization approach are likely to contribute most to the improvement. The parcellation of the surface into ROI's is based on a set of heuristics in the present work; exploring a data-based parcellation scheme (e.g., clustering feature patches), which might further improve performance, remains as future work.

Validation on pathological datasets also remains as future work. The proposed approach is expected to sustain high segmentation accuracy in such datasets, provided the availability of suitable training sets that reflect the underlying distribution.

Acknowledgments

This work was supported by National Institutes of Health grants (R01-EB004640, R01-NS094456). We are grateful to Dr. Hans Johnson for providing the dataset.

References

1. Paulsen J, Long J, Johnson H, Aylward E, Ross C, Williams J, Nance M, Erwin C, Westervelt H, Harrington D, Bockholt H. Clinical and biomarker changes in premanifest Huntington disease show trial feasibility: a decade of the PREDICT-HD study. Frontiers in Aging Neuroscience. 2014; 6(APR)
2. Long JD, Paulsen JS. Investigators, P.-H., and of the Huntington Study Group, C. Multivariate prediction of motor diagnosis in Huntington's disease: 12 years of PREDICT-HD. Movement Disorders. 2015; 30(12):1664–1672. [PubMed: 26340420]
3. Paulsen JS, Long JD, Ross CA, Harrington DL, Erwin CJ, Williams JK, Westervelt HJ, Johnson HJ, Aylward EH, Zhang Y, et al. Prediction of manifest Huntington's disease with clinical and imaging measures: a prospective observational study. The Lancet Neurology. 2014; 13(12):1193–1201. [PubMed: 25453459]
4. Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage. 2011; 56(3):907–922. [PubMed: 21352927]
5. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. Neuron. 2002; 33(3):341–355. [PubMed: 11832223]
6. Wang H, Suh JW, Das SR, Pluta JB, Craige C, Yushkevich PA. Multi-atlas segmentation with joint label fusion. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013; 35(3):611–623. [PubMed: 22732662]
7. Powell S, Magnotta VA, Johnson H, Jammalamadaka VK, Pierson R, Andreasen NC. Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures. Neuroimage. 2008; 39(1):238–247. [PubMed: 17904870]
8. Yin Y, Zhang X, Williams R, Wu X, Anderson DD, Sonka M. LOGISMOS–layered optimal graph image segmentation of multiple objects and surfaces: cartilage segmentation in the knee joint. IEEE Transactions on Medical Imaging. 2010; 29(12):2023–2037. [PubMed: 20643602]
9. Kashyap, S., Oguz, I., Zhang, H., Sonka, M. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2016. Automated Segmentation of Knee MRI Using Hierarchical Classifiers and Just Enough Interaction Based Learning: Data from Osteoarthritis Initiative; p. 344-351.
10. Oguz I, Sonka M. LOGISMOS-B: layered optimal graph image segmentation of multiple objects and surfaces for the brain. IEEE Trans Med Imaging. 2014; 33(6):1220–1235. [PubMed: 24760901]
11. Oguz, I., Sonka, M. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2014. Robust cortical thickness measurement with LOGISMOS-B; p. 722-730.
12. Oguz, I., Kashyap, S., Wang, H., Yushkevich, P., Sonka, M. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2016. Globally Optimal Label Fusion with Shape Priors; p. 538-546.
13. Oguz I, Zhang H, Rumple A, Sonka M. RATS: rapid automatic tissue segmentation in rodent brain MRI. Journal of Neuroscience Methods. 2014; 221:175–182. [PubMed: 24140478]
14. Breiman L. Random forests. Machine learning. 2001; 45(1):5–32.
15. Kim EY, Johnson HJ. Robust multi-site MR data processing: iterative optimization of bias correction, tissue classification, and registration. Frontiers in Neuroinformatics. 2013; 7:29. [PubMed: 24302911]

16. Van Leemput K, Maes F, Vandermeulen D, Suetens P. Automated model-based tissue classification of MR images of the brain. IEEE Transactions on Medical Imaging. 1999; 18(10):897–908. [PubMed: 10628949]

17. Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004; 26(9):1124–1137. [PubMed: 15742889]

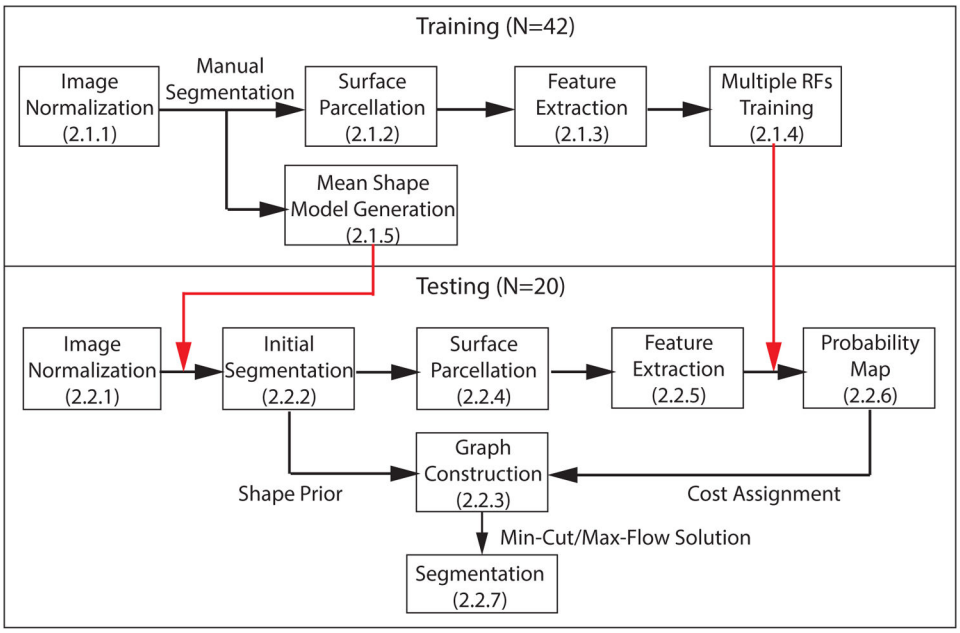
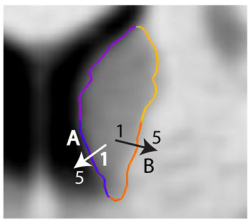


Figure 1.
The work flow of the proposed method. The section numbers for steps are shown in parenthesis.



(a)

Node Index	Average Intensity (left caudate)	
	A (facing the ventricle)	B (away from the ventricle)
1	1810	2174
2	1702	2346
3	1600	2519
4	1526	2685
5	1502	2823

(b)

Figure 2.

Different appearance models are needed for different regions of the surfaces. (a) An MRI slice zoomed in on the left caudate, overlaid with the manual segmentation surface. Region A (blue) is facing the ventricle. Region B (red) is facing the white matter. The white and black arrows indicate the direction of graph columns. (b) The average intensity across the training set for nodes in ROI's A vs. B. A has a gradient from bright (caudate) to dark (CSF), while B has a gradient from dark (caudate) to bright (WM).

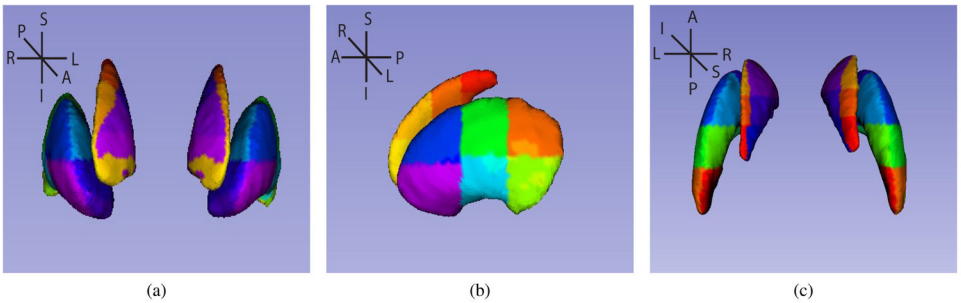


Figure 3.

An example of parcellation result for pairs of caudate and putamen in three viewpoints. (a) View from Anterior. (b) View from Left. (c) View from Superior. Different color represents different ROI after parcellation.

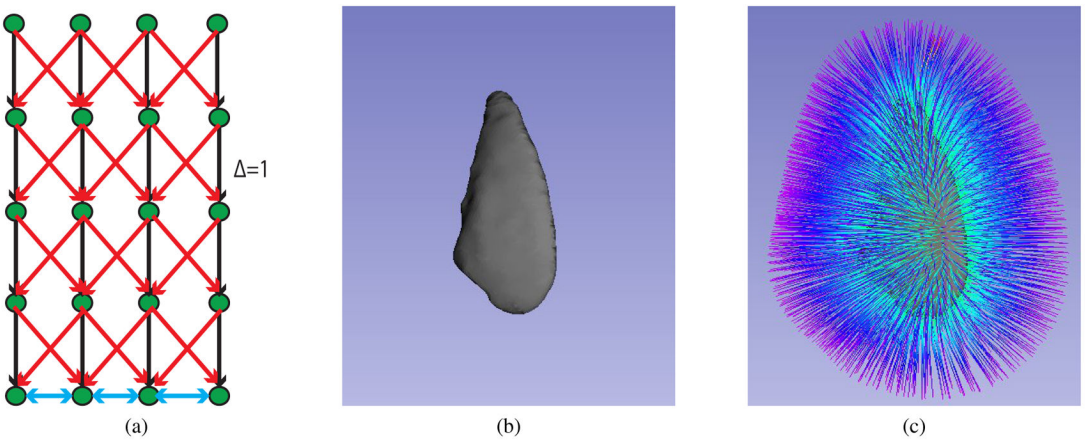


Figure 4.

The graph framework construction. (a) Simplified graph structure. The green circles represent nodes on the columns. Each vertical array of nodes represent a column. The red arrows denote the smoothness constraint, the cyan ones represent the base graph to determine the support domain for graph search, and the black arrows represent intra-column arcs. (b) An initial segmentation for the caudate. (c) The graph columns based on (b). Each column has 50 nodes (numbered 0–49) and the color shows the number.

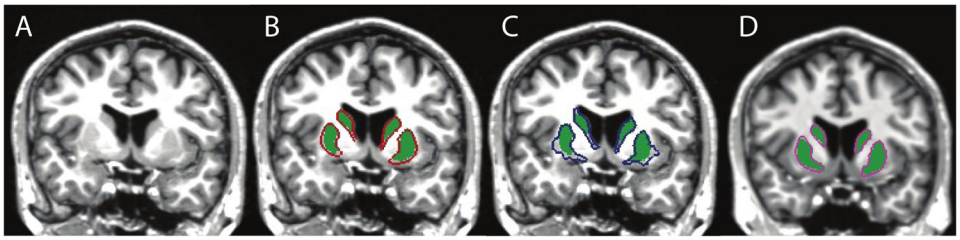


Figure 5.
Qualitative segmentation results. Green regions show the ground truth, whereas the outlines are the automated segmentation results. A: original slice. B: FSL. C: FreeSurfer. D: LOGISMOS-RF.

Table 1

Comparison of Dice Coef and Avg Surf Dist for caudate and putamen of three automated segmentation results.

	Methods	Dice Coef		Avg Surf Dist		Success Rate
		Mean ± SD	p value	Mean ± SD	p value	
Caudate	FSL	0.84 ± 0.03	≪ 0.001*	0.68 ± 0.09	≪ 0.001*	70%†
	FreeSurfer	0.83 ± 0.03	≪ 0.001*	1.18 ± 0.32	≪ 0.001*	100%
	LOGISMOS-RF	**0.89 ± 0.03**	\	**0.44 ± 0.04**	\	100%
Putamen	FSL	0.88 ± 0.01	0.19	0.56 ± 0.05	≪ 0.001*	70%†
	FreeSurfer	0.79 ± 0.03	≪ 0.001*	1.25 ± 0.23	≪ 0.001*	100%
	LOGISMOS-RF	**0.89 ± 0.03**	\	**0.42 ± 0.05**	\	100%

The p value column indicates the paired t-test between LOGISMOS-RF and the method in the row. The symbol † indicates that the number of subjects using for t test between FSL and LOGISMOS-RF is 14. The symbol * indicates statistical significance ($p<0.05$). The **bold** numerical value shows the best result among the results from all three different methods.