Two-stage multishape segmentation of brain structures using image intensity, tissue type, and location information^{a)}

Alireza Akhondi-Asl and Hamid Soltanian-Zadeh^{b)}

Control and Intelligent Processing Center of Excellence (CIPCE), School of Electrical and Computer Engineering, University of Tehran, North Kargar Ave., Tehran 14395-515, Iran; School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran 19395-5746, Iran; and Department of Radiology, Image Analysis Laboratory, Henry Ford Health System, Detroit, Michigan 48202

(Received 25 January 2010; revised 15 May 2010; accepted for publication 20 May 2010; published 29 July 2010)

Purpose: The authors propose a fast, robust, nonparametric, entropy-based, coupled, multishape approach to segment subcortical brain structures from magnetic resonance images (MRIs).

Methods: The proposed method uses three types of information: Image intensity, tissue types, and locations of structures. The image intensity information is captured by estimating the probability density function (pdf) of the image intensities in each structure. The tissue type information is captured by applying an unsupervised tissue segmentation method to the image and estimating a probability mass function (pmf) for the tissue type of each structure. The location information is captured by estimating pdf of the location of each structure from the training datasets. The resulting pmf's and pdf's are used to define an entropy function whose minimum corresponds to a desirable segmentation of the structures. The authors propose a three-step optimization strategy for the segmentation method. In the first step, a powerful automatic initialization method is developed based on tissue type and location information of the structures. In the second step, a quasi-Newton method is used to optimize the parameters of the energy function. To speed up the iterations, derivatives of the energy function with respect to its parameters are analytically derived and used in the optimization process. In the last step, the limitations related to the prior shape model are removed and a level-set method is applied for the fine tuning of the segmentation results.

Results: The proposed method is applied to two different datasets and the results are compared to those of previous methods in literature. Experimental results are presented for lateral ventricles, caudate, thalamus, putamen, pallidum, hippocampus, and amygdala.

Conclusions: The results illustrate superior performance of the proposed segmentation method compared to other methods in literature. The execution time of the algorithm is a few minutes, suitable for a variety of applications. © 2010 American Association of Physicists in Medicine. [DOI: [10.1118/1.3459018](https://doi.org/10.1118/1.3459018)]

Key words: medical image processing, image segmentation, brain structures, shape modeling, location information, tissue type information, level set

I. INTRODUCTION

A major category of the methods proposed for the segmentation of brain structures from magnetic resonance images (MRIs) optimizes an energy function with several parameters that represent the underlying shapes. In the definition of the energy function, earlier methods use the boundary information for the structures of interest.¹⁻³ Later methods use regional information such as intensity histogram (parametric and nonparametric, offline or online) or the intensity variance in an area.⁴⁻⁶ Others combine boundary and regional information.⁷ Recent methods benefit from *a priori* knowledge about the structures of interest.⁸⁻¹¹ This makes the segmentation process robust to the imperfect image conditions. For the methods developed based on the *a priori* information, a registration process is essential to integrate the prior model into the segmentation process. Many of these methods use probabilistic techniques to represent this information. For example, Chupin *et al.*¹² introduced a method driven by hybrid constraints. They defined an energy function based on

global and local data attachments and nonstationary anisotropic Markovian regularization terms. In addition, they applied two additional terms for volume and surface control. Their method has many parameters and is well adapted for their specific datasets.

The anatomical structures in the brain are related to the neighboring structures through their location, size, orientation, and shape. An integration of these relations into the segmentation process improves accuracy and robustness.¹³⁻²³ Litvin *et al.*¹⁸ introduced shape distribution as a new concept for the segmentation of coupled objects. Their model is constructed from distributions of features related to the shape. Their method is 2D and its extension to 3D is not reported yet. Addition of new terms in their energy function leads to challenges in the calculation of the derivatives required for the curve evolution. The deformable models introduced in Refs. 13, 21, and 23 define coupled shape models of multiple structures. Akselrod-Ballin *et al.*¹⁴ proposed a knowledge-based multiscale segmentation method that applied a graph

representation in different levels. They used the probabilistic information derived from an atlas and a likelihood function estimated from the training datasets.

Tu and Toga²² developed a hybrid method that applied a multiclass classifier for the learning/computing of the multiclass discriminative models and a learned edge field to constrain the region boundaries. Bazin and Pham¹⁵ introduced a segmentation method that used a statistical and topological atlas generated from the training data along with some pre-existing general atlases. Development of the topological atlas requires manual editing and is thus semiautomatic. The authors applied the energy function used in the FANTASM method.²⁴ Corso *et al.*¹⁶ proposed a graph shifts algorithm using a dynamical hierarchical representation of the image. The terms in the energy function were learned from the training data. Wu and Chung²⁵ introduced a method based on Markov dependence tree. In this method, a framework for segmentation of multiple brain structures is introduced that uses edge, region, and partial Hausdorff distance. The partial Hausdorff distance is used to define constraints that minimize the distance between the boundaries of objects and the image edge maps.

In this paper, we propose a 3D segmentation method, which considers coupling information of the shapes of the related structures in a nonparametric entropy-based energy function. We use principal component analysis (PCA) to extract principal shapes of different structures.^{13,21,23} The proposed method is robust, fast, and accurate with a small number of parameters to set. It integrates information obtained from different sources in the energy function. In addition, while using the prior knowledge for the extraction of the shape model, it allows flexibility of fine-tuning by relaxing constraints related to the shape model in the final stage of the algorithm. There is another category of segmentation methods in the energy function to limit shape variation and gain flexibility.^{9,10} However, these methods are sensitive to the weights used for different energy terms. If the weights of the shape constraints, extracted from the training datasets to limit the shape variation around a mean shape, are large, there will be almost no flexibility. In other words, the shape of the segmented structure is limited to the mean shapes extracted from the training datasets. On the other hand, if these weights are small, the segmentation loses robustness because the information extracted from the training datasets is not used. Moreover, these types of methods are sensitive to the initialization.

We consider tissue type information in the energy function by constructing of a probability mass function (pmf) using a fuzzy membership function. In addition, we consider position and shape of the structures to define a probability density function (pdf) for the location of each structure. The energy function is built based on this information and nonparametric pdf's for the intensity of the structures. In our method for shape representation, we use the Euclidean sign distance function as a powerful shape representation method.¹⁰ The energy function has two types of parameters, which are the weights of the principal shapes and the trans-

Fig. 1. Steps of the proposed method in the training and test phases.

formation parameters of each shape for the local alignment. To find the minimum of the energy function, first we use a powerful initialization strategy using tissue type and location information. Then, we apply a quasi-Newton based algorithm to optimize the parameters. In the third step, we utilize the segmentation result of the previous step as the initialization of a level-set to remove the limitations of the shape deformation due to the principal shapes. Figure 1 shows the general steps of the proposed method.

The main contributions of this work are threefold: (1) An integration of the location and tissue type information with a prior knowledge of the structures in a new energy function, (2) a multistage optimization process, and (3) a powerful initialization approach. To see the effectiveness of each part of the algorithm, we apply the proposed method in four steps. In the first step, we use the training datasets for the pdf estimation and use the coupling of the shape and pose. We name this method “PCA-Only.” This is a modified version of our method in Ref. 13 that improves the registration, the selection of the number of the principal components, and the optimization procedures. In the second step, we use a modified initialization, update the intensity pdf's iteratively, and also used location and tissue type information by adding the pdf for the location and the pmf for the tissue type information. Because of the modification in the probability function relative to the PCA-Only, we name it “Modified-Probability.” In the third step, we change the coupling method in the Modified-Probability to consider only the

TABLE I. Brief comparison of four different implemented methods based on the utilized features.

	Shape coupling	Initialization	Online pdf estimation	Location and tissue type	Fine-tuning	Pose coupling
Fine-Tuned	✓	✓	✓	✓	✓	×
Shape-Coupled	✓	✓	✓	✓	×	×
Modified-Probability	✓	✓	✓	✓	×	✓
PCA-Only	✓	×	×	×	×	✓

shape coupling and thus name it “Shape-Coupled.” Finally, in the last step, we add the fine-tuning algorithm to the Shape-Coupled and name it “Fine-Tuned.” For convenience, these names are used throughout the paper. Table I compares these four methods based on their features. The rest of the paper is organized as follows. Section II explains the proposed shape model and its properties. In Sec. III, we explore the energy function and the proposed method for its optimization. Experimental data, procedures, and results are presented in Sec. IV. Discussion and conclusions are presented in Secs. V and VI, respectively.

II. SHAPE MODEL

In many segmentation methods, a shape model is used where richer models generate more accurate results. A shape model can be more sophisticated than a shape representation, but for the sake of simplicity, we use them as synonyms here. Several shape representation methods are used in literature.^{24–29} Cootes *et al.*²⁷ introduced a point based method. Hong *et al.*²⁸ used kernel integrals for shape representation and Pizer *et al.*²⁹ used medial shape representation. Powerful methods for shape representation are based on distance function, implicit representation, and relationships among different shapes including pose, orientation, and other geometrical relations.^{30,31}

Colliot *et al.*³⁰ used fuzzy forces to describe spatial relation of different structures and Bahman-Bijari *et al.*³¹ used symmetry properties of structures for segmentation. Scherrer *et al.*³² used a unified Markov random field framework for the segmentation of the tissue types as well as specific structures. They use location constraints provided by fuzzy descriptor as *a priori* anatomical knowledge to improve the segmentation results. Ciofolo and Barillot³³ introduced a segmentation method using level set and a fuzzy decision system. In their method, the initial contours of multiple structures evolve using the fuzzy decision system that combines *a priori* knowledge extracted from an anatomical atlas with the intensity pdf of the image and relative positions of the contours.

In this paper, shape relation and individual transformations are used to develop shape models of the brain structures. To this end, PCA is applied on the 3D training datasets. Each training dataset has a label map that shows manual segmentation of the structures by the expert. In this image, each structure has a specific label (fixed value/intensity). Before applying PCA, the label maps are aligned to an atlas. The atlas has two aligned images: Intensity image and the

label map that assigns to each voxel a label of a specific structure. There are two ways to align the label maps of the training datasets to the label map of the atlas. The label map of a specific structure can be aligned individually or the labels of all of the structures can be aligned collectively. When the label map that includes all of the structures is used for alignment, in effect, both of the shape and pose information are used for the coupling. This alignment is used in the PCA-Only and Modified-Probability. On the other hand, when the label maps of the structures are aligned individually, then in effect, only the shape information is used. This alignment is used in the Shape-Coupled and Fine-Tuned. The alignment method is described next.

II.A. Alignment of training datasets for shape variability extraction

In knowledge-based segmentation methods that use shape variation, alignment is a critical step. It reduces shape variability due to misalignment of different subjects. A good review of the basic registration methods can be found in Ref. 34. For 3D objects, several transformations such as similarity, affine, and Euler can be used. The Euler transformation is applicable when there is no scale difference between the moving and the target images. The affine transformation is applicable when there is a small misalignment. In addition, it has 12 parameters, making the optimization process complex. Therefore, we first apply a similarity transformation with seven parameters (one scaling, three rotations, and three translations). Then, we apply an affine transformation for the fine-tuning of the registration process.

For the metric, several options are available in literature. They vary based on the types of the images used as the moving and target (reference) images. Since our goal is to align the labels, the metric in our implementation is a cardinality metric that computes the number of pixels that are not matched between the moving and target images. This metric is equal to zero when all of the pixels are matched. It can be used for both of the cases described above.

For interpolation of binary images, we use a nearest neighbor interpolator. For the optimization, since the proposed metric is nondifferentiable, we use an optimizer that does not need derivatives. This is the Nelder–Mead (amoeba) method, which is a popular direct search method for minimizing unconstrained real functions with our desired property.³⁵

Now, suppose that there are n 3D labeled images available for the training, each with m structures labeled as

$\{l_1, l_2, \dots, l_m\}$. Based on the methodology used for the coupling, we align the label maps. For each structure, we define a 3D binary image that represents the desired structure with one (foreground) and other pixels with zero (background). Therefore, for each structure, there are n 3D binary images that are aligned. The results are used to extract the variability of the shape of each structure in the training datasets.

To achieve an accurate and robust registration, a good initialization is needed. We use the difference between the centers of mass of the moving and reference images as the initial translation parameters, the angle between the main principal axes of the two images as the initial rotation parameter, and 1 for the initial scale parameter. Using these initializations, the method does not have any problems with the small sizes of the structures. We do not make any assumptions about the images and do not use any regions of interest (ROIs) at this point.

Using the above methods, we extract shape variability of the desired structures for model construction as explained in Sec. II B. It should be noticed that all of the other parts of the segmentation algorithm are the same for both of the coupling methods described above.

II.B. Implicit parametric shape representation

As stated in Ref. 10, implicit parametric shape representation has advantages such as computational efficiency, accuracy, capturing wide range of shape variability, and handling topological changes. We use a distance map for shape representation that is zero on the boundary of a structure and in other points is the Euclidian distance from the boundary (negative inside, positive outside). For simultaneously segmenting m structures that are coupled, we embed their boundaries as the zero level sets of the corresponding signed distance functions $\{\psi_1, \psi_2, \dots, \psi_m\}$. We have n aligned binary images for each one of the m desired structures.

After extracting n distance maps for each one of the m desired structures (ψ_k^j shows the distance map of the k th structure of the i th dataset), we subtract the mean distance map of each structure, computed by averaging of the training datasets ($\bar{\Psi}_k$), from each of the n signed distance maps to remove similar parts in different shapes and show them with $\tilde{\psi}_k^j$,

$$\bar{\Psi}_k = \frac{1}{n} \sum_{i=1}^n \psi_k^i, \quad \tilde{\psi}_k^j = \psi_k^j - \bar{\Psi}_k. \quad (1)$$

We use these $n \times m$ maps to show the variability of different structures in the training dataset. We collect n column vectors in matrix $S = \{\tilde{\psi}^1, \tilde{\psi}^2, \dots, \tilde{\psi}^n\}$ with size $mN_xN_yN_z \times n$, where N_x and N_y are the number of pixels in the x and y directions in each slice and N_z is the number of slices. Each column is built from a concatenation of the vectors that are the lexicographical order of $N_x \times N_y \times N_z$ matrices $\{\tilde{\psi}_1^j, \tilde{\psi}_2^j, \dots, \tilde{\psi}_m^j\}$. We need the eigenvectors of SS^T to define the eigenshapes. However, SS^T is a very large matrix. To

minimize computational complexity, we find the eigenvalues of S^TS and call its eigenvectors $\mathbf{d}^i$. Then, we find the eigenvectors of SS^T using the transformation $\phi^i = S\mathbf{d}^i$. This generates up to n different eigenshapes for each of the m structures, denoted by ϕ_k^i . Each one of these eigenshapes includes the relationships between different structures and takes the coupling into account. This is because the data from all structures are used in the definition of the eigenshapes. When the label maps of a training dataset are aligned individually, the shape relation of the structures is encoded in the model (Shape-Coupled and Fine-Tuned). On the other hand, when they are aligned collectively, both of their shape and pose relations are encoded (PCA-Only and Modified-Probability).

To allow limited, robust shape variability, we use $q \leq n$ eigenvectors to represent each shape. The number of the principal components is calculated using

$$q = \min \left\{ k \mid 1 \leq k \leq n, \sum_{i=1}^k \lambda_i \geq f \sum_{i=1}^n \lambda_i \right\}, \quad (2)$$

where λ_i 's are the eigenvalues and f determines the threshold. The eigenvalues show variations in the model in the directions of the eigenshapes and are ordered decreasingly.

In addition, to consider pose differences, we add 12 local alignment parameters to the shape parameters of each structure. This is used in Shape-Coupled and Fine-Tuned. Our experiments show that using similarity transforms with seven parameters or a single global transformation is insufficient; different shape models need independent local alignments with more flexibility. For PCA-Only and Modified-Probability, we use single global transformation with 12 parameters. Finally, for each structure, we may write

$$\begin{aligned} \phi_k(\mathbf{x}) &= \phi_k[\mathbf{w}, \mathbf{p}_k](\mathbf{x}) = \phi_k[\mathbf{w}, \mathbf{p}_k](x, y, z) \\ &= \bar{\Psi}_k(\tilde{x}_k, \tilde{y}_k, \tilde{z}_k) + \sum_{i=1}^q w_i \phi_k^i(\tilde{x}_k, \tilde{y}_k, \tilde{z}_k), \end{aligned} \quad (3)$$

where $\mathbf{w}$ is the vector of eigenvectors multipliers and $\mathbf{p}_k$ is the vector containing 12 transformation parameters for the alignment of the k th structure according to the following equations:

$$\begin{bmatrix} \tilde{x}_k \\ \tilde{y}_k \\ \tilde{z}_k \\ 1 \end{bmatrix} = T(\mathbf{c}_k)A(\mathbf{p}_k)T(-\mathbf{c}_k) \begin{bmatrix} x_k \\ y_k \\ z_k \\ 1 \end{bmatrix} = M_k \begin{bmatrix} x_k \\ y_k \\ z_k \\ 1 \end{bmatrix}, \quad (4)$$

where T and A are the translation and affine transformation matrices, respectively, and $\mathbf{c}$ is the fixed center of rotation, computed using the midposition of the ROI for each structure. This choice of the center of rotation improves the local alignment of the model to the data. In this manner, each structure's pose may change, while shape class covariations are used for the coupling. For the PCA-Only and Modified-Probability, $\mathbf{p}_k$ and $\mathbf{c}_k$ are the same for all of the structures and the center of mass is computed using the mid position of the ROI of the segmentation. Also, we define $\mathbf{P}$ as the vector that contains all of the parameters. The number of the param-

eters equals $m \times 12 + q$ for the Shape-Coupled and Fine-Tuned (each of the local alignments has 12 parameters) and equals $12 + q$ for the PCA-Only and Modified-Probability (they have only one global transformation). In Sec. III, we present the proposed entropy-based segmentation method using the shape model described above.

III. SEGMENTATION PROCESS

After construction of the shape models and their covariances, we register the image that we want to segment to the intensity image of the atlas. To this end, the image is registered using normalized mutual information and linear interpolation. To improve the results, the skull is stripped from the images using the BET method.³⁶ The other parts of the registration protocol are the same as before. An energy function is defined for the segmentation process as explained below.

III.A. Energy model

An energy function integrates a variety of information sources for segmentation. To this end, multiple sources of information should be used without adding unwanted complexity to the method. In addition, initialization is critical for optimization of the energy function. In our method, three types of information are used: Image intensity, tissue types, and locations of structures. The image intensity information is captured by estimating the pdf of the image intensities in each structure. This information is used in all of the methods. The tissue type information is captured by applying an unsupervised tissue segmentation method to the image and estimating a pmf for the tissue type of each structure. The location information is captured by estimating pdf of the location of each structure from the training datasets. The tissue type and location information are used in the Modified-Probability, Shape-Coupled, and Fine-Tuned. As explained next, the resulting pmf's and pdf's are used to define an entropy function whose minimum corresponds to a desirable segmentation of the structures.

To segment m coupled structures with closed boundaries, there are m regions for these structures. We set the area outside of the m structures as $m+1$ and use this notation throughout the paper. Based on the estimated entropy of these regions, $EN(\Omega_k)$ $\{\Omega_k, k=1, \dots, m+1\}$ an energy function is defined as $J(\Omega_1, \dots, \Omega_{m+1}) = |\Omega| \sum_{k=1}^{m+1} P_{\Omega_k} EN(\Omega_k)$, where $|\Omega|$ is the cardinality of a set Ω (number of pixels) and $P_{\Omega_k} = |\Omega_k|/|\Omega|$. When all of the regions are as uniform as possible, are in the correct tissue as much as possible, and also are in the correct location, the energy function is at its minimum. The entropy of the k th structure is estimated using $EN(\Omega_k) = (-1/|\Omega_k|) \int_{\Omega_k} \ln(p_k(\mathbf{x}) \times p_{fk}(\mathbf{x}) \times p_{sk}(\mathbf{x})) d\mathbf{x}$. Here, $p_k(\mathbf{x})$ is an estimate of the pdf of image intensities in the k th region of the 3D image I , $p_{fk}(\mathbf{x})$ is the pmf of the tissue type of the k th region, and $p_{sk}(\mathbf{x})$ is the pdf of the location of the k th region. In the PCA-Only, we utilize the pdf of image intensities ($p_k(\mathbf{x})$) and we do not consider the other two prob-

ability functions. Thus, we can consider $p_{fk}(\mathbf{x})$ and $p_{sk}(\mathbf{x})$ equal to those of the PCA-Only. In the other three methods (Modified-Probability, Shape-Coupled, and Fine-Tuned), we also make use of the other two probabilities. Some researchers estimated the pdf's of image intensities offline, i.e., they used the pixel intensities in the training datasets to construct the pdf's of image intensities and used the results for the segmentation of the testing datasets without adapting them to the datasets being segmented.^{19,21} In the PCA-Only approach, we estimate the pdf of the image intensities offline. However, we observed dissimilar dynamic ranges of image intensities in different datasets and concluded that an offline pdf is suboptimal. Therefore, we use online estimation, i.e., we estimate the intensity pdf's in each iteration of the algorithm. To estimate the intensity pdf's in each iteration, we use a powerful and robust method (the Parzen window method³⁷) with a Gaussian kernel,

$$p_j(\mathbf{x}) = p_j[\mathbf{w}, \mathbf{p}_j](\mathbf{x})$$

$$= \frac{\int_{\Omega} H(-\phi_j[\mathbf{w}, \mathbf{p}_j](\hat{\mathbf{x}})) K(I(\mathbf{x}) - I(\hat{\mathbf{x}})) d\hat{\mathbf{x}}}{\int_{\Omega} H(-\phi_j[\mathbf{w}, \mathbf{p}_j](\hat{\mathbf{x}})) d\hat{\mathbf{x}}},$$

$$p_{m+1}(\mathbf{x}) = p_{m+1}[\mathbf{P}](\mathbf{x})$$

$$= \frac{\int_{\Omega} \prod_{k=1}^m H(\phi_k[\mathbf{w}, \mathbf{p}_k](\hat{\mathbf{x}})) K(I(\mathbf{x}) - I(\hat{\mathbf{x}})) d\hat{\mathbf{x}}}{\int_{\Omega} \prod_{k=1}^m H(\phi_k[\mathbf{w}, \mathbf{p}_k](\hat{\mathbf{x}})) d\hat{\mathbf{x}}}, \quad (5)$$

where H is the Heaviside (unit step) function. This is the approach that we use in the other three methods. To translate the integrals over the changing regions into an integral over the entire image, the Heaviside functions

$$H(s) = \begin{cases} 0 & \text{if } s < 0 \\ 1 & \text{if } s \geq 0 \end{cases}$$

are used. In these equations, K is a Gaussian kernel with the standard deviation (sigma) that sets the resolution of the estimated pdf. Care must be taken in selecting appropriate values for sigma. Small sigma values make pdf estimation sensitive to noise and large values remove useful details from the estimated pdf. In the literature, values between 1 and 3 are used; we apply the method of Silverman³⁸ to choose it adaptively,

$$\sigma = 0.9AN^{1/5}, \quad \text{where } A = \min \left\{ s, \frac{R}{1.34} \right\}, \quad (6)$$

where s is the empirical standard deviation of the samples in each region, R is the interquartile range of the intensities, and N is the number of samples.

To estimate the pmf's of the tissue types of the structures,

we apply a fuzzy clustering method named FANTASM (Ref. 24) to segment the brain MRI into white matter (WM), gray matter (GM), and cerebrospinal fluid regions. This method is an extension of the standard fuzzy clustering, which is robust to the image intensity nonuniformities and noise. It also considers the information of the neighboring voxels in the clustering process. In its application to the T1-weighted images, the resulting clusters are ordered such that the center of the first cluster has the lowest value (representing darkest regions of the image) and the last cluster center has the highest value (representing the brightest region of the image). Our experiments show that each structure may partially belong to different tissue types and also the ranges of the intensities may be different in different structures. For example, amygdala and hippocampus are in the gray matter but they have different intensity ranges. To capture this information, we use a new strategy where we apply the clustering twice with 3 and 10 clusters. The 3-class segmentation results capture the global intensity information of the normal tissues in the brain while the 10-class segmentation captures their local or fine intensity information. Our experiments have shown that the 10-class clustering captures details of the structures' intensity information and is sufficiently robust to the intensity inhomogeneities and artifacts.

To use the results, we model the pmf of the tissue type of each structure, i.e., the probabilities of the 10 clusters in the structure as

$$p_{fk}(\mathbf{x}) = \begin{cases} \varepsilon, & l(\mathbf{x}) \leq L1 - 3, \quad l(\mathbf{x}) \geq L2 + 3 \\ \frac{1}{4}, & l(\mathbf{x}) = L1 - 2, \quad l(\mathbf{x}) = L2 + 2 \\ \frac{3}{4}, & l(\mathbf{x}) = L1 - 1, \quad l(\mathbf{x}) = L2 + 1 \\ 1, & L1 \leq l(\mathbf{x}) \leq L2, \end{cases} \quad (7)$$

where $l(\mathbf{x})$ shows the cluster number in 10-class segmentation. It should be noted that the probabilities in each voxel only depend on the cluster numbers. In addition, $L1$ and $L2$ are computed based on the following equation:

TABLE II. Shift values used for the construction of tissue type pmf's for different structures.

Structure	l_1	l_2
Thalamus	+1	-1
Caudate	+1	-2
Hippocampus	0	-2
Amygdala	-1	-2
Putamen	+1	-1
Pallidum	+3	0

$$\begin{cases} L1 = \arg \min_{1 \leq i \leq 10} \text{abs}(C_3(2) - C_{10}(i)) + l_1 \\ L2 = \arg \min_{1 \leq i \leq 10} \text{abs}(C_3(3) - C_{10}(i)) + l_2. \end{cases} \quad (8)$$

In this equation, $C_n(k)$ shows the intensity of the center of the k th class (cluster) in the n -class clustering. $L1$ and $L2$ are not fix numbers; they are calculated for each dataset by Eq. (8). In other words, the closest cluster number in the 10-class segmentation to the WM or GM centers are not fixed; they vary in different datasets. In addition, the model is designed for the T1-weighted images used to study anatomy of the structures; for other MR images that have different contrasts, different models may be needed.

Based on the brain anatomy, each structure is located in a specific tissue type and based on the MRI intensities, it has a specific intensity range. This information is used to estimate the parameters l_1 and l_2 that are listed in Table II for different structures. Because of specific features of ventricles, $L1=1$ and $L2=\arg \min_{1 \leq i \leq 10} \text{abs}(C_3(2) - C_{10}(i)) - 2$ are considered for the ventricles. For the region outside all of the m structures (region $m+1$), if all of the other structures have the same pdf, we use $1-p$, but if there are different pdf's, we use $p=0.5$, as described in Eq. (9),

$$p_{fm+1}(\mathbf{x}) = \begin{cases} 1 - p_{f1}(\mathbf{x}) & \text{if } \forall k, \quad k' \neq k \in \{1, \dots, m\} \quad p_{fk}(\mathbf{x}) = p_{fk'}(\mathbf{x}) \\ 0.5 & \text{otherwise.} \end{cases} \quad (9)$$

To capture the *a priori* information about the locations of the structures, we construct a pdf for the location of each structure. We previously registered the image of interest (a test dataset) to the intensity image of the atlas. Thus, alignment of the intensity images of the training datasets can be used to find location information of the structures. To this end, the intensity images of the training datasets are registered using the same algorithm described above to the intensity image of the atlas. Again, the skull is stripped from the images using the BET method.³⁶ The same registration protocol is applied to all of the training datasets. Then, the re-

sulting transformations are applied to the binary images (labels) of the structures. After registration of all of the training datasets, the strategy explained for shape model extraction is applied to find the mean shape of each structure $\bar{\Phi}_k$ [this notation is used to show the difference between this mean shape, extracted from the registered intensity images, and the one extracted from the registered label images explained before ($\bar{\Psi}_k$)].

Next, the $\bar{\Phi}_k(\mathbf{x})$ are used to construct the pdf's for the locations of the structures. Clearly, $\bar{\Phi}_k(\mathbf{x})$ is negative for the

voxels inside the mean structure, zero on the boundary, and positive outside. If for a point, this value is more negative than another one, it means that it is more inside the structure and/or is inside the structure in a larger number of the train-

ing datasets than the other one. In addition, the points on the boundary are neither inside nor outside. Thus, the following decreasing function of $\bar{\Phi}_k(\mathbf{x})$ which is 1 for the minimum of $\bar{\Phi}_k(\mathbf{x})$ and is 0.5 for $\bar{\Phi}_k(\mathbf{x})=0$ (boundary voxels) is defined:

$$p_{sk}(\mathbf{x}) = \begin{cases} \ln \left((e - e^{0.5}) \left(\frac{\bar{\Phi}_k(\mathbf{x})}{\min_{\mathbf{x}}(\bar{\Phi}_k(\mathbf{x}))} \right) + e^{0.5} \right) & \text{if } \ln \left((e - e^{0.5}) \left(\frac{\bar{\Phi}_k(\mathbf{x})}{\min_{\mathbf{x}}(\bar{\Phi}_k(\mathbf{x}))} \right) + e^{0.5} \right) > \varepsilon \\ \varepsilon = 10^{-4} & \text{otherwise,} \end{cases}$$

$$p_{sm+1}(\mathbf{x}) = \min_{1 \leq k \leq m} (1 - p_{sk}(\mathbf{x})). \quad (10)$$

Here, for the region outside all of the structures, we find $1 - p$ for all of the structures and use their minimum for constructing the pdf of this region. ε is used to avoid $\ln(0)$ in Eq. (11). The resulting functions can be considered as scaled pdf's.

Figure 2 shows the probability map of the image intensity ($p_k(\mathbf{x})$), the tissue type ($p_{fk}(\mathbf{x})$), and the location ($p_{sk}(\mathbf{x})$) of the thalamus in a sample slice along with their multiplication ($p_k(\mathbf{x}) \times p_{fk}(\mathbf{x}) \times p_{sk}(\mathbf{x})$) and the segmentation result generated by the Fine-Tuned (gray line) and the expert segmentation (white). Brighter points show higher probabilities. It should be noted that none of the probabilities predicts the corresponding structure accurately and thus cannot be used for the segmentation by itself. However, multiplication of the probabilities localizes the structure of interest accurately. In other words, each term captures a specific attribute of a structure only while their multiplication captures all of the attributes. To summarize, the pdf's and pmf's estimated by the above methods are used in the following energy function (also introduced earlier in this section):

$$J(\Omega_1, \dots, \Omega_{m+1}) = J(P) = \sum_{k=1}^{m+1} - \int_{\Omega_k} \ln(p_k(\mathbf{x}) \times p_{fk}(\mathbf{x}) \times p_{sk}(\mathbf{x})) d\mathbf{x}. \quad (11)$$

The following equation demonstrates the detailed dependency of the energy function to the different parameters:

$$J = \sum_{j=1}^m \int_{\Omega} H(-\phi_j[\mathbf{w}, \mathbf{p}_j](\mathbf{x})) F(p_j[\mathbf{w}, \mathbf{p}_j](\mathbf{x}) \times p_{fj}(\mathbf{x}) \times p_{sj}(\mathbf{x})) d\mathbf{x} + \int_{\Omega} \prod_{k=1}^m H(\phi_k[\mathbf{w}, \mathbf{p}_k](\mathbf{x})) F(p_{m+1}[\mathbf{P}](\mathbf{x}) \times p_{fm+1}(\mathbf{x}) \times p_{sm+1}(\mathbf{x})) d\mathbf{x}, \quad (12)$$

where $p_{sk}(\mathbf{x})$ and $p_{fk}(\mathbf{x})$ do not depend on the parameters in

the vector $\mathbf{P}$. In this equation, $F(p) = -\ln(p)$. However, in the Modified-Probability, Shape-Coupled, and Fine-Tuned, $p_k(\mathbf{x})$ depends on the parameters according to Eq. (5); for the PCA-Only, it is fixed.

III.B. Energy optimization

To optimize the energy function, we have three steps. An initialization, a quasi-Newton algorithm to optimize the parameters of the shapes, and a level set algorithm for fine tuning. First, we present the initialization.

As explained in Sec. I, initialization is very important for a segmentation algorithm. We use the center of mass of the negative portions of the $\bar{\Phi}_k$ to place the center of mass of the k th shape model of the brain structures, $\bar{\Psi}_k$. Then, we fine-tune the initialization using the location and tissue type information (their estimated pdf and pmf). This is done by optimizing the following energy function using its derivative with respect to $\mathbf{d}$:

$$E_j(\mathbf{d}) = - \int_{\Omega_j} p_{fj}(\hat{\mathbf{x}} + \mathbf{d}) \times p_{sj}(\hat{\mathbf{x}} + \mathbf{d}) d\hat{\mathbf{x}},$$

$$\frac{\partial E_j}{\partial \mathbf{d}_i} = \int_{\Gamma_j} \frac{\partial \phi_j(\hat{\mathbf{x}} + \mathbf{d})}{\partial \mathbf{d}_i} \times p_{fj}(\hat{\mathbf{x}} + \mathbf{d}) \times p_{sj}(\hat{\mathbf{x}} + \mathbf{d}) d\hat{\mathbf{x}}, \quad (13)$$

where $\mathbf{d}$ is the translation vector rearranged as a translation matrix in Eq. (4) and $\int_{\Gamma_j}$ is used to denote an integration over the boundary of the region Ω_j . The method to compute the derivatives will be described after the second step. We use this step in the Modified-Probability, Shape-Coupled, and Fine-Tuned.

In the next step, to minimize the energy function, we use a quasi-Newton algorithm along with the Broyden–Fletcher–Goldfarb–Shanno method for Hessian matrix estimation.³⁹ Although methods such as steepest descent are also popular in literature,²¹ we obtained superior results using the quasi-Newton algorithm. This method needs more memory and is

more complex than the steepest descent algorithm but finds the optimal point faster. It requires the gradient of the objective function with respect to the parameters. The gradients can be estimated using numerical methods but analytical computation is faster and generates more accurate results. As such, we computed them analytically resulting in the following equations for the derivatives of the energy function with respect to the parameters $\mathbf{w}$ and $\mathbf{p}_k$ (Modified-Probability, Shape-Coupled, and Fine-Tuned):

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{w}_i} = & \sum_{j=1}^m \int_{\Gamma_j} \frac{\partial \phi_j(\hat{\mathbf{x}})}{\partial \mathbf{w}_i} \times \left\{ \ln(\hat{p}_j(\hat{\mathbf{x}}) \times p_{fj}(\hat{\mathbf{x}}) \times p_{sj}(\hat{\mathbf{x}})) \right. \\ & - \ln(\hat{p}_{m+1}(\hat{\mathbf{x}}) \times p_{fm+1}(\hat{\mathbf{x}}) \times p_{sm+1}(\hat{\mathbf{x}})) + \frac{1}{|\Omega_j|} \\ & \times \int_{\Omega_j} (K(I(\mathbf{x}) - I(\hat{\mathbf{x}}))/\hat{p}_j(\mathbf{x})) d\mathbf{x} - \frac{1}{|\Omega_{m+1}|} \\ & \times \left. \int_{\Omega_{m+1}} (K(I(\mathbf{x}) - I(\hat{\mathbf{x}}))/\hat{p}_{m+1}(\mathbf{x})) d\mathbf{x} \right\} d\hat{\mathbf{x}}, \quad (14) \end{aligned}$$

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{p}_j^i} = & \int_{\Gamma_j} \frac{\partial \phi_j(\hat{\mathbf{x}})}{\partial \mathbf{p}_j^i} \times \left\{ \ln(\hat{p}_j(\hat{\mathbf{x}}) \times p_{fj}(\hat{\mathbf{x}}) \times p_{sj}(\hat{\mathbf{x}})) \right. \\ & - \ln(\hat{p}_{m+1}(\hat{\mathbf{x}}) \times p_{fm+1}(\hat{\mathbf{x}}) \times p_{sm+1}(\hat{\mathbf{x}})) + \frac{1}{|\Omega_j|} \\ & \times \int_{\Omega_j} (K(I(\mathbf{x}) - I(\hat{\mathbf{x}}))/\hat{p}_j(\mathbf{x})) d\mathbf{x} - \frac{1}{|\Omega_{m+1}|} \\ & \times \left. \int_{\Omega_{m+1}} (K(I(\mathbf{x}) - I(\hat{\mathbf{x}}))/\hat{p}_{m+1}(\mathbf{x})) d\mathbf{x} \right\} d\hat{\mathbf{x}}. \quad (15) \end{aligned}$$

Note that if for a point on the boundary, the probability of being in the outside region is larger than being in the inside region, then the boundary should expand. This happens because in this scenario, the terms inside the braces in the first lines of Eqs. (14) and (15) become negative, and assuming that the first terms on the first lines are positive, result in an increase in the parameters, as expected. The same holds true if the first terms are negative.

If we use the training datasets for the intensity pdf estimation of each structure and fix it during the optimization, we have simpler equations in the following form [which is used for the PCA-Only with $p_{fk}(\mathbf{x})$ and $p_{sk}(\mathbf{x})$ equal to one for all of the voxels]:

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{w}_i} = & \sum_{j=1}^m \int_{\Gamma_j} \frac{\partial \phi_j(\hat{\mathbf{x}})}{\partial \mathbf{w}_i} \times \{ \ln(\hat{p}_j(\hat{\mathbf{x}}) \times p_{fj}(\hat{\mathbf{x}}) \times p_{sj}(\hat{\mathbf{x}})) \\ & - \ln(\hat{p}_{m+1}(\hat{\mathbf{x}}) \times p_{fm+1}(\hat{\mathbf{x}}) \times p_{sm+1}(\hat{\mathbf{x}})) \} d\hat{\mathbf{x}}, \quad (16) \end{aligned}$$

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{p}_j^i} = & \int_{\Gamma_j} \frac{\partial \phi_j(\hat{\mathbf{x}})}{\partial \mathbf{p}_j^i} \times \{ \ln(\hat{p}_j(\hat{\mathbf{x}}) \times p_{fj}(\hat{\mathbf{x}}) \times p_{sj}(\hat{\mathbf{x}})) \\ & - \ln(\hat{p}_{m+1}(\hat{\mathbf{x}}) \times p_{fm+1}(\hat{\mathbf{x}}) \times p_{sm+1}(\hat{\mathbf{x}})) \} d\hat{\mathbf{x}}. \quad (17) \end{aligned}$$

FIG. 2. The probability map of the image intensity (a), the tissue type (b), and the location (c) of the thalamus in a sample slice along with their multiplication (d) and segmentation result generated by the Fine-Tuned (gray line) and the expert segmentation (white) (e) are shown. Brighter points indicate higher probabilities. It should be noted that each term captures a specific attribute of the structure, while their combination (multiplication) captures all attributes of the structure.

Thus, we do not need to update pdf's in each iteration. In this case, we lose quality due to the intensity variations between the training and testing datasets. Also, $\partial \phi_j(\mathbf{x}) / \partial \mathbf{p}_j^i$ and $\partial \phi_j(\mathbf{x}) / \partial \mathbf{w}_i$ are calculated as follows:

$$\frac{\partial \phi_k(\mathbf{x})}{\partial \mathbf{w}_i} = \phi_k^i(\mathbf{x}), \quad (18)$$

$$\frac{\partial \phi_k(\mathbf{x})}{\partial \mathbf{p}_k^i} \begin{bmatrix} \frac{\partial \phi_k(\tilde{x}_k, \tilde{y}_k, \tilde{z}_k)}{\partial \tilde{x}_k} \\ \frac{\partial \phi_k(\tilde{x}_k, \tilde{y}_k, \tilde{z}_k)}{\partial \tilde{y}_k} \\ \frac{\partial \phi_k(\tilde{x}_k, \tilde{y}_k, \tilde{z}_k)}{\partial \tilde{z}_k} \\ 0 \end{bmatrix}^T \times \frac{\partial M_k}{\partial \mathbf{p}_k^i} \times \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (19)$$

where $\partial M_k / \partial \mathbf{p}_k^i$ is computed using derivatives of each matrix element.

Note that we do not use the intensity information (estimated pdf) in this step because the initial location is such that the estimated pdf may be inaccurate. Usually, parts of the initial model are outside the correct region of the structure, and thus the estimated pdf is inaccurate and may cause large segmentation errors. However, after fine-tuning the placement of the centers of mass of the models, the variances of the Gaussian kernels of the Parzen window for estimation of the intensity pdf's are estimated once using Eq. (6). Then, the iterative procedure explained in the previous subsection to optimize the energy function in Eq. (11) is executed.

This approach is very effective for the segmentation but it is limited to the shape model. We propose a strategy to overcome this limitation by adding an extra step, where the quasi-Newton based algorithm introduced above is followed by a fine-tuning step using a level-set method. To this end, the same energy function defined in Eq. (11) is minimized using a level-set formulation. This is the approach that we utilize only in the Fine-Tuned. With some mathematical operations, it can be shown that the level-set evolution of the region j can be described with the following partial differential equation:

$$\begin{aligned} \frac{\partial \phi_j(\mathbf{x})}{\partial t} = & |\nabla \phi_j(\mathbf{x})| \left(-\ln(p_j(\mathbf{x}) \times p_{fj}(\mathbf{x}) \times p_{sj}(\mathbf{x})) \right. \\ & + \ln(p_{m+1}(\mathbf{x}) \times p_{fm+1}(\mathbf{x}) \times p_{sm+1}(\mathbf{x})) - \frac{1}{|\Omega_j|} \\ & \times \int_{\Omega_j} (K(I(\hat{\mathbf{x}}) - I(\mathbf{x})) / p_j(\hat{\mathbf{x}})) d\hat{\mathbf{x}} + \frac{1}{|\Omega_{m+1}|} \\ & \left. \times \int_{\Omega_{m+1}} (K(I(\hat{\mathbf{x}}) - I(\mathbf{x})) / p_{m+1}(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \right). \quad (20) \end{aligned}$$

The initial sign distance functions of this step are the final shapes of the previous step (segmentation based on quasi-Newton algorithm). The proposed method overcomes the two problems explained in Sec. I (initialization and sensitivity to the weights) at the cost of the time used for this step, which is not significant because of the good initialization used as explained above.

IV. EXPERIMENTS AND RESULTS

IV.A. Data

To test and evaluate the proposed method, we have applied it on the MRI data from two different sources. We have

used datasets from the Internet Brain Segmentation Repository⁴⁰ as the primary data. These datasets are T1-weighted volumetric images with slightly different voxel sizes ($0.938 \times 0.938 \times 1.5$, $1.0 \times 1.0 \times 1.5$, and $0.837 \times 0.837 \times 1.5$ mm³). There are 18 datasets on which expert physicians have segmented 43 brain structures. These datasets consist of $256 \times 256 \times 128$ noncubical voxels. The smallest physical size that covers all of the datasets is $256 \times 256 \times 192$ mm³. As a preprocessing step, all datasets are resampled into $1.0 \times 1.0 \times 1.0$ mm³ cubical voxels after the registration phase. The resampling is a prerequisite for using the sign distance functions defined on isotropic spaces. The second datasets are the images and segmentation results of the hippocampi of 21 temporal lobe epilepsy (TLE) patients acquired using a General Electric 1.5 T Signa System (GE Medical Systems, Milwaukee, WI) and processed using the EIGENTool software.⁴¹ The TLE may cause changes in the size and shape of the hippocampus. Coronal T1-weighted MRI, acquired using a spoiled gradient-echo sequence with a pixel size of 0.781×0.781 mm² and a slice thickness of 2.0 mm is used.

IV.B. Procedure

Based on the structures of interest, a subvolume (ROI) is chosen to improve speed and robustness of the algorithm. This improves the execution speed and accuracy of the method since the pdf's of all regions inside and outside of each desired structure need to be estimated. We define the ROI as 1.05 times of the smallest cube that covers all of the desired structures in all of the training datasets. The additional 5% confidence interval has been sufficient for the test datasets (image segmentation phase) in our experiments.

To evaluate the results, we use the Dice coefficient (k) as a similarity index. In addition, as an alternative evaluation measure, we use the Hausdorff distance (H), which is the maximum of the minimum distances between the corresponding surface points of the two volumes. Due to the outliers, the 95 percentile of the Hausdorff distance (H_{95}) is used in our study as well as many previous studies. Moreover, the mean distance (M) is used as another evaluation measure. For 3D visualization of the structures, we use 3D SLICER (Ref. 42) and for registration of the training and testing datasets, we use the Insight Segmentation and Registration Toolkit (ITK) library.⁴³ For optimization and extraction of the principal shapes, we use MATLAB.⁴⁴ All programs are run on a 3.2 GHz Windows XP workstation with 4 Gbyte of RAM.

For the IBSR dataset, we segment the following seven structures on the two sides of the brain for a total of 14 structures: (1) Lateral ventricles, (2) caudate, (3) thalamus, (4) putamen, (5) pallidum, (6) hippocampus, and (7) amygdala. These structures are either used in many applications or hard to segment because of their unclear boundaries. They include a variety of shapes with different signal intensities, sizes, and other geometric features. We use the leave-one-out strategy to test the proposed segmentation algorithm. We first segment each of the left and right structures indi-

TABLE III. Dice coefficients (k) for the segmentation results of the proposed methods for seven brain structures in test datasets and the results of the other methods that reported their segmentation results for the IBSR datasets (Refs. 14, 15, 25, 33, and 46).

	Thalamus	Putamen	Caudate	Pallidum	Hippocampus	Amygdala	Ventricle
Fine-Tuned	0.86 ± 0.04	0.82 ± 0.05	0.81 ± 0.04	0.76 ± 0.04	0.73 ± 0.05	0.70 ± 0.08	0.80 ± 0.06
Shape-Coupled	0.85 ± 0.04	0.80 ± 0.06	0.78 ± 0.04	0.75 ± 0.05	0.71 ± 0.06	0.68 ± 0.09	0.73 ± 0.10
Modified-Probability	0.86 ± 0.02	0.80 ± 0.07	0.78 ± 0.03	0.72 ± 0.10	0.68 ± 0.08	0.67 ± 0.10	0.75 ± 0.07
PCA-Only	0.85 ± 0.04	0.75 ± 0.14	0.72 ± 0.12	0.72 ± 0.10	0.66 ± 0.07	0.64 ± 0.10	0.67 ± 0.11
Akselrod <i>et al.</i>	0.84	0.79	0.80	0.74	0.69	0.63	–
Wu <i>et al.</i>	0.84	0.81	0.80	–	–	–	–
ISCA	0.80	0.78	0.74	0.70	0.64	0.58	–
Naive Prior	0.83	0.77	0.65	0.72	0.62	0.65	–
Gouttard <i>et al.</i>	–	0.78	0.76	0.71	0.67	0.64	0.85
Coifolo <i>et al.</i>	0.77	0.70	0.65	0.62	–	–	–
Bazin <i>et al.</i>	0.78 ± 0.05	0.82 ± 0.05	0.78 ± 0.04	–	–	–	0.85 ± 0.04
Hammer	0.77 ± 0.05	0.60 ± 0.07	0.74 ± 0.04	–	–	–	0.66 ± 0.08

vidually using the principal shapes extracted from the training datasets. We also segment the left and right structures or multiple structures together by the proposed coupling method. For the TLE datasets, we use the leave-one-out strategy and apply the same procedure used for the IBSR datasets.

IV.C. Results

The quantitative measures (k, M, H_{95}) used to evaluate the segmentation results for the selected structures of the 18 IBSR datasets are compared in Tables III–V, respectively. It should be noted that the measurements are obtained after the final sign distance functions are thresholded at zero. The results of the four implementations of our method are compared to eight other methods from the literature that have

used the IBSR datasets and reported their results when using all of the datasets. Note that the comparison does not include the methods that although used the IBSR datasets but did not report the above measures, e.g., Ref. 45. In the following, we first compare the performance measures of our proposed methods with those of the other methods in literature. Then, we compare the four implementations of our approach with each other. First, we compare the methods based on the k values in Table III.

A comparison of the PCA-Only (simplest case) and the eight methods from the literature shows that our method is generally superior to the methods of ISCA,¹⁴ Naive Prior,¹⁴ Coifolo *et al.*,³³ and Hammer.¹⁵ It is, however, inferior to the methods of Refs. 14, 46, and 15. The standard deviation values suggest that the PCA-Only is not as robust (stable) as

TABLE IV. Mean (M) distances for the segmentation results of the proposed methods for seven brain structures in test datasets and the results of the other methods that reported their segmentation results for the IBSR datasets (Refs. 14, 15, 25, 33, and 46).

	Thalamus	Putamen	Caudate	Pallidum	Hippocampus	Amygdala	Ventricle
Fine-Tuned	0.65 ± 0.18	0.58 ± 0.19	0.44 ± 0.13	0.70 ± 0.16	0.80 ± 0.17	0.76 ± 0.27	0.68 ± 0.23
Shape-Coupled	0.75 ± 0.22	0.72 ± 0.23	0.61 ± 0.14	0.74 ± 0.19	0.84 ± 0.18	0.78 ± 0.29	0.70 ± 0.28
Modified-Probability	0.74 ± 0.13	0.77 ± 0.34	0.64 ± 0.14	0.83 ± 0.35	0.98 ± 0.32	0.83 ± 0.32	0.66 ± 0.24
PCA-Only	0.78 ± 0.24	1.00 ± 0.70	0.87 ± 0.51	0.82 ± 0.32	1.01 ± 0.23	0.98 ± 0.27	0.87 ± 0.41
Akselrod <i>et al.</i>	1.44	1.6	1.44	1.43	1.88	1.67	–
ISCA	1.55	1.72	1.84	1.55	1.91	1.78	–
Coifolo <i>et al.</i>	1.70	1.46	1.71	1.51	–	–	–
Bazin <i>et al.</i>	1.45 ± 0.22	0.97 ± 0.26	0.94 ± 0.19	–	–	–	0.78 ± 0.17

TABLE V. Hausdorff (H_{95}) distances for the segmentation results of the proposed methods for seven brain structures in test datasets and the results of the other methods that reported their segmentation results for the IBSR datasets (Refs. 14, 15, 25, 33, and 46).

	Thalamus	Putamen	Caudate	Pallidum	Hippocampus	Amygdala	Ventricle
Fine-Tuned	2.31 ± 0.58	2.11 ± 0.60	2.17 ± 0.45	2.32 ± 0.61	2.51 ± 0.64	2.88 ± 0.85	3.55 ± 3.35
Shape-Coupled	2.38 ± 0.65	2.32 ± 0.63	2.17 ± 0.51	2.38 ± 0.69	2.62 ± 0.63	2.99 ± 0.94	4.90 ± 4.34
Modified-Probability	2.28 ± 0.42	2.29 ± 0.87	2.28 ± 0.66	2.58 ± 0.89	3.34 ± 1.09	3.06 ± 1.04	4.55 ± 3.59
PCA-Only	2.32 ± 0.62	2.88 ± 1.82	2.56 ± 1.30	2.53 ± 0.76	3.30 ± 0.91	3.60 ± 1.11	5.72 ± 3.76
Akselrod <i>et al.</i>	2.9	3.36	3.07	2.75	4.57	3.38	–
ISCA	3.2	3.89	4.46	3.2	4.44	3.89	–

some of the other methods. For example, for the putamen, it has a standard deviation of 0.14, which is relatively large. Many of the methods did not report the standard deviation but the method of Bazin and Pham¹⁵ reported a smaller standard deviation (0.05).

Next, we compare the Modified-Probability, Shape-Coupled, and Fine-Tuned with the other methods. The Modified-Probability is generally superior to all of the methods except for the methods of Refs. 14, 15, 46. The Shape-Coupled is superior to the other methods and similar to the method of Ref. 15. The Fine-Tuned is similar to the method of Ref. 15 for the putamen while being superior to the rest. It is also superior to the other methods for the amygdala and hippocampus. However, it is inferior to the methods of Refs. 15 and 46 but superior to the rest for the ventricles. In general, the Fine-Tuned is more robust than the other methods.

A comparison based on the mean distance values (M) in Table IV shows that our methods are superior to the other methods for all structures. The only exception is the ventricles where the PCA-Only is inferior to that of Ref. 15. However, the standard deviations of the Fine-Tuned are about 15% lower than those of the method of Ref. 15, on average. Moreover, a comparison based on the H_{95} values in Table V shows that our method is superior to the others for all structures. The only exception is the amygdala where the PCA-Only is inferior to that of Ref. 14.

Overall, the method with most similar values of the three quantitative measures to those of our methods is the method of Ref. 14. A comparison of the Fine-Tuned with the method of Ref. 14 shows that our method produces superior results based on M and H_{95} . For example, for the caudate, it results in the M value of 0.44 mm, which is about 30% of that of the method of Ref. 14. In addition, our method is superior to theirs for the amygdala based on k .

Next, we compare the Fine-Tuned with our other methods (PCA-Only, Modified-Probability, and Shape-Coupled). A comparison of the Fine-Tuned with the Shape-Coupled shows superiority of the segmentations of the lateral ventricles, hippocampus, and caudate based on the k ($p < 0.05$). In addition, it shows significant improvement for the segmentation of the caudate and putamen based on M ($p < 0.05$). The overall superiority (considering all structures) of the Fine-Tuned relative to the Shape-Coupled is about 3% based on the k and about 8% ($p < 0.05$) based on the M . Moreover, the standard deviations of the k and M values of the Fine-Tuned are 18% and 13% less than those of the Shape-Coupled, on average.

A comparison of the Fine-Tuned and Modified-Probability illustrates that the first method is superior to the latter for the segmentation of the caudate and hippocampus based on the k ($p < 0.05$). In addition, the segmentation results of the Fine-Tuned are superior to those of the Modified-Probability for the putamen and caudate based on the M ($p < 0.05$). A comparison of the Fine-Tuned and PCA-Only shows that the Fine-Tuned generates significantly superior results for the caudate, pallidum, hippocampus, amygdala, and ventricles based on k ($p < 0.05$). Also, it is superior for

the segmentation of caudate, putamen, and ventricles based on M ($p < 0.05$). In overall, the superiority of the Fine-Tuned relative to the PCA-Only is 9% and 37% based on k and M , respectively ($p < 0.05$). In other cases, the superiority of the Fine-Tuned is not statistically significant.

Now, we compare the three methods with the coupling information and without fine-tuning (PCA-Only, Modified-Probability, and Shape-Coupled). A comparison of PCA-Only and Shape-Coupled illustrates that in overall (mean of the performance measures for all structures) the k and M of the Shape-Coupled are about 7% and 19% superior to the PCA-Only ($p < 0.05$). The difference in the performance may be attributed to the approach by which the two methods use the coupling information. In the Shape-Coupled, the coupling information is just based on the shape variation among the structures but in the PCA-Only, there is additional variation due to the pose of the structures. This additional variation might have made their eigenshapes different from those of the brain structures.

A comparison of Modified-Probability and Shape-Coupled shows similar findings. The superiority of the Shape-Coupled based on k is about 1% but based on M and H_{95} is about 6% and 3%, respectively. Additionally, the superiority of the Shape-Coupled is more pronounced for small structures with large pose variations (pallidum, hippocampus, and amygdala). In Fig. 3, a sample segmentation of the selected structures by the PCA-Only, Modified-Probability, Shape-Coupled, and Fine-Tuned are shown along with the expert segmentation. An important observation here is the marginal quality of the manual segmentation (with unsmooth surfaces) which is partly due to the slice-by-slice approach used in the manual segmentation. This leads to imperfect (nonrepresentative) values for the performance measures especially for H_{95} . In addition, Fig. 4 shows a combination of the pdf of the image intensities ($p_k(\mathbf{x})$), the pmf for the tissue type ($p_{fk}(\mathbf{x})$), and the pdf of the location ($p_{sk}(\mathbf{x})$) of the lateral ventricles, caudate, putamen, pallidum, hippocampus, and amygdala in a sample slice along with the segmentation results generated by the Fine-Tuned (gray line) and the expert segmentations (white).

For further comparison, we use the TLE datasets. Table VI compares the results of the methods and illustrates the superiority of the Fine-Tuned to the rest. Note that the fine-tuning step has improved the segmentation results of the patients' images more than the IBSR datasets (normal brains). This can be attributed to the fact that the hippocampi of the TLE patients are not symmetric (one hippocampus is usually smaller than the other). In this case, the superiority of the Fine-Tuned relative to the Shape-Coupled is 7% and 15% based on k and M , respectively, where the improvement based on the k is statistically significant ($p < 0.05$). Consequently, alleviating the constraint imposed by the principal shapes allows accurate segmentation of the abnormal parts of the structures that are not covered by the principal shapes. In addition, because of the complex relation between shape and pose of the hippocampi in the TLE patients, the Modified-Probability cannot extract the relation between the pose and

FIG. 3. Comparison of 3D views of the structures segmented manually (a), by PCA-Only (b), by Modified-Probability (c), by Shape-Coupled (d), and by Fine-Tuned (e).

shape appropriately. The superiority of the Shape-Coupled relative to the Modified-Probability based on k and M are 11% and 12%, respectively. Again, the improvement based on k is statistically significant ($p < 0.05$). Figure 5 compares the automatic and expert segmentation of the hippocampus on 6 coronal slices.

Finally, a critical issue for an automated segmentation method in clinical applications is its execution time. The mean execution time of our method is about 9 min for the algorithm with coupling information. This time is for the registrations (2 steps), the FANTASM (3- and 10-class clustering), the initialization, the quasi-Newton step, and the level-set step.

V. DISCUSSION

Tsai *et al.*²¹ introduced a 2D method using PCA for coupling using the nonparametric pdf in a mutual information

FIG. 4. Combination of the probability maps of the image intensities, the tissue type, and the location are shown along with the segmentation results generated by the Fine-Tuned (gray line) and the expert segmentations (white) for the lateral ventricles (a), caudate (b), putamen (c), pallidum (d), hippocampus (e), and amygdala (f) in sample slices. Brighter points indicate higher probabilities.

framework. They used training datasets to estimate the pdf's and fixed them during the segmentation process. Because of the intensity variations among different structures, the use of the training datasets to estimate the pdf's of the image inten-

TABLE VI. Dice coefficients (k) and mean (M) and Hausdorff (H_{95}) distances for the segmentation results of the proposed methods for the hippocampus in the test datasets for the temporal lobe epilepsy datasets.

	k	M	H_{95}
Fine-Tuned	0.73 ± 0.06	0.90 ± 0.29	3.34 ± 1.12
Shape-Coupled	0.68 ± 0.08	1.06 ± 0.33	3.74 ± 1.21
Modified-Probability	0.61 ± 0.11	1.21 ± 0.40	5.29 ± 1.75
PCA-Only	0.43 ± 0.15	1.93 ± 0.48	7.45 ± 2.56

FIG. 5. Comparison of the segmentation results for the hippocampus generated by the Fine-Tuned (gray line) and the expert segmentation (white) in a series of coronal images from a temporal lobe epilepsy dataset.

sities in different structures fails in many cases.^{17,21} Thus, we estimate these pdf's in each iteration of the algorithm and have introduced an extended method in Ref. 13 to update the estimated pdf's during the iterations of the optimization algorithm.

A common problem of the aforementioned methods is that they are not stable in some cases and do not use any specific prior knowledge in the segmentation process. In addition, they have initialization problems and heavily depend on the training datasets. Akhondi-Asl and Soltanian-Zadeh²⁶ introduced a constrained optimization strategy to develop a robust segmentation method. Albeit its superiority in performance, it takes considerable time to execute and is somewhat sensitive to the parameters of the constraints. Another problem in many of the segmentation methods is the sensitivity of the results to the initialization of the algorithm.^{13,21,23}

An important observation is that the fine tuning step improves the segmentation results by overcoming the limitations of the prior shape model (principal shapes extracted from the training datasets). In other words, each dataset has parts similar to the training datasets that are represented by the prior knowledge. In addition, there are dissimilarities between the training datasets and the testing datasets that are not represented by the prior knowledge. The fine-tuning step removes the constraints imposed by the prior shape model and thus improves the segmentation results. Based on this analysis, the improvements are expected to be more pronounced when segmenting the patient data (abnormal structures). In addition, the fine-tuning step allows closer adjustment of the segmentation results to the image edges and thus the improvements generated by this approach are quantified the best by the Hausdorff distance.

Note that performances of different algorithms on different structures are different; none of the methods generate perfect segmentation results for all of the structures. Moreover, due to the small size of the amygdala relative to the other brain structures, the performance measure k shows a lower performance for the automatic segmentation of this structure. However, it is interesting to note that the improvements obtained by our proposed methods are more pronounced in this case.

A critical issue in comparing different segmentation methods is that unless the same datasets are used, the performance

measures are incomparable. This is based on the fact that depending on the quality and variation of the datasets, an algorithm can produce significantly different results. For example, Powell *et al.*²⁰ reported highly accurate segmentation results for the caudate, hippocampus, and putamen for a specific high-resolution dataset that is not publicly available. Yes, it is not clear how good that method will perform on the publicly available datasets like the IBSR dataset.

VI. CONCLUSIONS

We have presented a new method for the segmentation of the brain subcortical structures using their shape relationships and tissue type and location information. In the model development phase, the proposed method registers each structure individually before estimating their shape variations. The registration is done in two steps using similarity and affine transformations. In the segmentation phase, the proposed method applies an independent transformation with 12 parameters (rotation, scaling, translation, and shearing) to each structure, improving flexibility of the segmentation model. The energy function used for the segmentation is based on the entropy of different structures. Tissue type information is used to improve robustness and accuracy of the segmentation process. In addition, location information is used to further improve the segmentation process. With a powerful automatic initialization of the structures which considers tissue type and location information and the use of the quasi-Newton algorithm, a local minimum of the energy function is found. In the final step, a level-set based optimization is applied. Throughout the algorithm, the tissue type and location information are used to generate robust and accurate segmentations. To achieve accurate results, the intensity pdf's are calculated in each iteration of the algorithm and gradients are computed analytically. Experimental results have illustrated superiority of the proposed framework to the other methods in the literature. The execution time of our algorithm is a few minutes, which is appropriate for a variety of applications.

ACKNOWLEDGMENTS

This work was supported in part by grants from the NIH (Contract No. R01-EB002450) and the University of Tehran.

APPENDIX: ANALYTIC DERIVATION OF GRADIENTS

In this section, mathematical derivation of the derivatives of the energy function with respect to the optimization parameters is presented. The energy function is based on the functions defined on $m+1$ regions. However, the region $m+1$ may be constructed from the other m regions. These regions have no intersections in the beginning due to the proposed initialization method and it is assumed that they do not overlap during the segmentation process either. Also, a

common notation, ω , is used to represent the parameters of the energy function that are called w_i 's and p_k^i 's in the body of the paper. Then, the following equation is derived for the derivative of the function for the region $m+1$ with respect to the parameters:

$$\frac{\partial \left(\prod_{k=1}^m H(\phi^k(\mathbf{x})) \right)}{\partial \omega} = \sum_{k=1}^m \delta(\phi^k(\mathbf{x})) \frac{\partial \phi^k(\mathbf{x})}{\partial \omega}. \quad (\text{A1})$$

Note that, as expected, the derivatives for the region $m+1$ depend on those of the other m regions.

Next, the following equations are derived for the derivatives of the estimated intensity pdf's with respect to the parameters. Note that the pdf of the region $m+1$ is defined based on the pdf's of the other regions. As such, their analytical expressions are different and their derivatives are presented separately in the following equations:

$$\begin{aligned} \frac{\partial p_k(\mathbf{x})}{\partial \omega} &= \frac{\partial}{\partial \omega} \left(\frac{\int_{\Omega} H(-\phi_k(\hat{\mathbf{x}})) K(I(\mathbf{x}) - I(\hat{\mathbf{x}})) d\hat{\mathbf{x}}}{\int_{\Omega} H(-\phi_k(\hat{\mathbf{x}})) d\hat{\mathbf{x}}} \right) = \frac{- \left(\int_{\Omega} \delta(\phi_k(\hat{\mathbf{x}})) \frac{\partial \phi_k(\hat{\mathbf{x}})}{\partial \omega} K(I(\mathbf{x}) - I(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \right) \times \left(\int_{\Omega} H(-\phi_k(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \right)}{\left(\int_{\Omega} H(-\phi_k(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \right)^2} \\ &\quad + \frac{\left(\int_{\Omega} \delta(\phi_k(\hat{\mathbf{x}})) \frac{\partial \phi_k(\hat{\mathbf{x}})}{\partial \omega} d\hat{\mathbf{x}} \right) \times \left(\int_{\Omega} H(-\phi_k(\hat{\mathbf{x}})) K(I(\mathbf{x}) - I(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \right)}{\left(\int_{\Omega} H(-\phi_k(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \right)^2} \\ &= \frac{p_k(\mathbf{x}) \times \left(\int_{\Omega} \delta(\phi_k(\hat{\mathbf{x}})) \frac{\partial \phi_k(\hat{\mathbf{x}})}{\partial \omega} d\hat{\mathbf{x}} \right) - \left(\int_{\Omega} \delta(\phi_k(\hat{\mathbf{x}})) \frac{\partial \phi_k(\hat{\mathbf{x}})}{\partial \omega} K(I(\mathbf{x}) - I(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \right)}{|\Omega_k|}, \end{aligned} \quad (\text{A2})$$

$$\begin{aligned} \frac{\partial p_{m+1}(\mathbf{x})}{\partial \omega} &= \frac{\partial}{\partial \omega} \left(\frac{\int_{\Omega} \prod_{k=1}^m H(\phi_k(\hat{\mathbf{x}})) K(I(\mathbf{x}) - I(\hat{\mathbf{x}})) d\hat{\mathbf{x}}}{\int_{\Omega} \prod_{k=1}^m H(\phi_k(\hat{\mathbf{x}})) d\hat{\mathbf{x}}} \right) \\ &= \frac{\left(\int_{\Omega} \sum_{k=1}^m \left(\delta(\phi_k(\hat{\mathbf{x}})) \frac{\partial \phi_k(\hat{\mathbf{x}})}{\partial \omega} \right) K(I(\mathbf{x}) - I(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \right) \times \left(\int_{\Omega} \prod_{k=1}^m H(\phi_k(\mathbf{x})) d\hat{\mathbf{x}} \right)}{\left(\int_{\Omega} \prod_{k=1}^m H(\phi_k(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \right)^2} \\ &\quad - \frac{\left(\int_{\Omega} \sum_{k=1}^m \left(\delta(\phi_k(\hat{\mathbf{x}})) \frac{\partial \phi_k(\hat{\mathbf{x}})}{\partial \omega} \right) d\hat{\mathbf{x}} \right) \times \left(\int_{\Omega} \prod_{k=1}^m H(\phi_k(\hat{\mathbf{x}})) K(I(\mathbf{x}) - I(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \right)}{\left(\int_{\Omega} \prod_{k=1}^m H(\phi_k(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \right)^2} \\ &= - \frac{p_{m+1}(\mathbf{x}) \times \left(\int_{\Omega} \sum_{k=1}^m \left(\delta(\phi_k(\hat{\mathbf{x}})) \frac{\partial \phi_k(\hat{\mathbf{x}})}{\partial \omega} \right) d\hat{\mathbf{x}} \right) - \left(\int_{\Omega} \sum_{k=1}^m \left(\delta(\phi_k(\hat{\mathbf{x}})) \frac{\partial \phi_k(\hat{\mathbf{x}})}{\partial \omega} \right) K(I(\mathbf{x}) - I(\hat{\mathbf{x}})) d\hat{\mathbf{x}} \right)}{|\Omega_{m+1}|}. \end{aligned} \quad (\text{A3})$$

Now, we can move on to the derivation of the derivatives of the energy function with respect to the parameters. To simplify the equations, $p_{it}(\mathbf{x}) = p_{fi}(\mathbf{x}) \times p_{si}(\mathbf{x})$ is used since neither of $p_{fi}(\mathbf{x})$ and $p_{si}(\mathbf{x})$ depends on the parameters; the only parameter dependent part is $p_j(\mathbf{x})$. In addition, to visualize all of the terms, the $-\ln$ function in the energy function is replaced by a general function F , resulting the following general energy function: $J = \sum_{j=1}^{m+1} \int_{\Omega_j} F(p_j(\mathbf{x}) \times p_{tj}(\mathbf{x})) d\mathbf{x}$:

$$\begin{aligned}
\frac{\partial J}{\partial \omega} &= \sum_{j=1}^{m+1} \frac{\partial \int_{\Omega_j} F(p_j(\mathbf{x}) \times p_{ij}(\mathbf{x})) d\mathbf{x}}{\partial \omega} = \sum_{j=1}^m \frac{\partial \int_{\Omega} H(-\phi_j(\mathbf{x})) F(p_j(\mathbf{x}) \times p_{ij}(\mathbf{x})) d\mathbf{x}}{\partial \omega} + \frac{\partial \int_{\Omega} \prod_{k=1}^m H(\phi_k(\mathbf{x})) F(p_{m+1}(\mathbf{x}) \times p_{im+1}(\mathbf{x})) d\mathbf{x}}{\partial \omega} \\
&= - \sum_{j=1}^m \int_{\Omega} \delta(\phi_j(\mathbf{x})) \frac{\partial \phi_j(\mathbf{x})}{\partial \omega} F(p_j(\mathbf{x}) \times p_{ij}(\mathbf{x})) d\mathbf{x} + \sum_{j=1}^m \int_{\Omega} H(-\phi_j(\mathbf{x})) F'(p_j(\mathbf{x}) \times p_{ij}(\mathbf{x})) \times p_{ij}(\mathbf{x}) \frac{\partial p_j(\mathbf{x})}{\partial \omega} d\mathbf{x} \\
&\quad + \int_{\Omega} \sum_{k=1}^m \left(\delta(\phi_k(\mathbf{x})) \frac{\partial \phi_k(\mathbf{x})}{\partial \omega} \right) F(p_{m+1}(\mathbf{x}) \times p_{im+1}(\mathbf{x})) d\mathbf{x} + \int_{\Omega} \prod_{k=1}^m H(\phi_k(\mathbf{x})) F'(p_{m+1}(\mathbf{x}) \times p_{im+1}(\mathbf{x})) \times p_{im+1}(\mathbf{x}) \frac{\partial p_{m+1}(\mathbf{x})}{\partial \omega} d\mathbf{x}.
\end{aligned} \tag{A4}$$

By substituting $\partial p_k(\mathbf{x})/\partial \omega$ and $\partial p_{m+1}(\mathbf{x})/\partial \omega$ from Eqs. (A2) and (A3) into Eq. (A4), we get

$$\begin{aligned}
\frac{\partial J}{\partial \omega} &= \int_{\Omega} \sum_{j=1}^m \left(\delta(\phi_j(\mathbf{x})) \frac{\partial \phi_j(\mathbf{x})}{\partial \omega} \right) \times (F(p_{m+1}(\mathbf{x}) \times p_{im+1}(\mathbf{x})) - F(p_j(\mathbf{x}) \times p_{ij}(\mathbf{x}))) d\mathbf{x} + \sum_{j=1}^m \int_{\Omega} H(-\phi_j(\mathbf{x})) F'(p_j(\mathbf{x}) \times p_{ij}(\mathbf{x})) \\
&\quad \times p_{ij}(\mathbf{x}) \frac{\partial p_j(\mathbf{x})}{\partial \omega} d\mathbf{x} + \int_{\Omega} \prod_{k=1}^m H(\phi_k(\mathbf{x})) F'(p_{m+1}(\mathbf{x}) \times p_{im+1}(\mathbf{x})) \times p_{im+1}(\mathbf{x}) \frac{\partial p_{m+1}(\mathbf{x})}{\partial \omega} d\mathbf{x} = \int_{\Omega} \sum_{j=1}^m \left(\delta(\phi_j(\mathbf{x})) \frac{\partial \phi_j(\mathbf{x})}{\partial \omega} \right) \\
&\quad \times (F(p_{m+1}(\mathbf{x}) \times p_{im+1}(\mathbf{x})) - F(p_j(\mathbf{x}) \times p_{ij}(\mathbf{x}))) d\mathbf{x} + \sum_{j=1}^m \int_{\Omega} H(-\phi_j(\mathbf{x})) F'(p_j(\mathbf{x}) \times p_{ij}(\mathbf{x})) \times p_{ij}(\mathbf{x}) \\
&\quad \times \left(\frac{\int_{\Omega} \delta(\phi_j(\hat{\mathbf{x}})) \frac{\partial \phi_j(\hat{\mathbf{x}})}{\partial \omega} (p_j(\mathbf{x}) - K(I(\mathbf{x}) - I(\hat{\mathbf{x}}))) d\hat{\mathbf{x}}}{|\Omega_j|} \right) d\mathbf{x} - \int_{\Omega} \prod_{k=1}^m H(\phi_k(\mathbf{x})) F'(p_{m+1}(\mathbf{x}) \times p_{im+1}(\mathbf{x})) \times p_{im+1}(\mathbf{x}) \\
&\quad \times \left(\frac{\left(\int_{\Omega} \sum_{k=1}^m \left(\delta(\phi_k(\hat{\mathbf{x}})) \frac{\partial \phi_k(\hat{\mathbf{x}})}{\partial \omega} \right) (p_{m+1}(\mathbf{x}) - K(I(\mathbf{x}) - I(\hat{\mathbf{x}}))) d\hat{\mathbf{x}} \right)}{|\Omega_{m+1}|} \right) d\mathbf{x}.
\end{aligned} \tag{A5}$$

Now, $F(p)$ is replaced by $-\ln(p)$ to get the derivatives of the proposed energy function. With further simplification, we obtain the final results in Eqs. (11) and (12).

^{a)}There is no conflict of interest. There were no restrictions or review from any funding sources.

^{b)}Author to whom correspondence should be addressed. Electronic mail: hamids@rad.hfh.edu; Telephone: (313) 874-4482; Fax: (313) 874-4494.

¹E. Angelini, Y. Jin, and A. Laine, "State of the art of level set methods in segmentation and registration of medical imaging modalities," *Handbook of Medical Image Analysis: Advanced Segmentation and Registration Models* (Springer, Berlin, 2005).

²V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.* **22**, 61–79 (1997).

³M. Jacob, T. Blu, and M. Unser, "Efficient energies and algorithms for parametric snakes," *IEEE Trans. Image Process.* **13**, 1231–1244 (2004).

⁴X. Bresson, "Image segmentation with variational active contours," Ph.D. thesis, Swiss Federal Institute of Technology (EPFL), 2005.

⁵A. Huang, G. Nielson, A. Razdan, G. Farin, D. Baluch, and D. Capco, "Thin structure segmentation and visualization in three-dimensional biomedical images: A shape-based approach," *IEEE Trans. Vis. Comput. Graph.* **12**, 93–102 (2006).

⁶M. Woolrich and T. Behrens, "Variational Bayes inference of spatial mixture models for segmentation," *IEEE Trans. Med. Imaging* **25**, 1380–1391 (2006).

⁷N. Paragios and R. Deriche, "Geodesic active regions and level set meth-

ods for supervised texture segmentation," *Int. J. Comput. Vis.* **46**, 223–247 (2002).

⁸S. Jehan-Besson, A. Herbulot, M. Barlaud, and G. Aubert, "Shape gradient for image and video segmentation," *Mathematical Models in Computer Vision: The Handbook* (Springer, Berlin, 2005).

⁹A. Khan, L. Wang, and M. Beg, "FreeSurfer—Initiated fully-automated subcortical brain segmentation in MRI using large deformation diffeomorphic metric mapping," *Neuroimage* **41**, 735–746 (2008).

¹⁰M. Leventon, W. Grimson, and O. Faugeras, *Statistical Shape Influence in Geodesic Active Contours* (IEEE Computer Society, Los Alamitos, CA, 2000), Vol. 1.

¹¹K. Pohl, J. Fisher, R. Kikinis, W. Grimson, and W. Wells, "Shape based segmentation of anatomical structures in magnetic resonance images," *Lect. Notes Comput. Sci.* **3765**, 489–498 (2005).

¹²M. Chupin, A. Hammers, R. Liu, O. Colliot, J. Burdett, E. Bardin, J. Duncan, L. Garnero, and L. Lemieux, "Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: Method and validation," *Neuroimage* **46**, 749–761 (2009).

¹³A. Akhondi-Asl and H. Soltanian-Zadeh, "Effect of number of coupled structures on the segmentation of brain structures," *J. Signal Proc. Syst. Signal, Image, Video Technol.* **54**, 215–230 (2009).

¹⁴A. Akselrod-Ballin, M. Galun, J. Gomori, A. Brandt, and R. Basri, "Prior knowledge driven multiscale segmentation of brain MRI," *Lect. Notes Comput. Sci.* **4792**, 118–126 (2007).

¹⁵P. Bazin and D. Pham, "Homeomorphic brain image segmentation with topological and statistical atlases," *Med. Image Anal.* **12**, 616–625 (2008).

- ¹⁶J. Corso, Z. Tu, A. Yuille, and A. Toga, "Segmentation of sub-cortical structures by the graph-shifts algorithm," *Lect. Notes Comput. Sci.* **4584**, 183–197 (2007).
- ¹⁷B. Fischl, D. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, and S. Klaveness, "Whole brain segmentation automated labeling of neuroanatomical structures in the human brain," *Neuron* **33**, 341–355 (2002).
- ¹⁸A. Litvin and W. Karl, *Coupled Shape Distribution-Based Segmentation of Multiple Objects* (Springer, Berlin, 2005), pp. 345–356.
- ¹⁹K. Pohl, J. Fisher, J. Levitt, M. Shenton, R. Kikinis, W. Grimson, and W. Wells, "A unifying approach to registration, segmentation, and intensity correction," *Lect. Notes Comput. Sci.* **3749**, 310–318 (2005).
- ²⁰S. Powell, V. Magnotta, H. Johnson, V. Jammalamadaka, R. Pierson, and N. Andreasen, "Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures," *Neuroimage* **39**, 238–247 (2008).
- ²¹A. Tsai, W. Wells, C. Tempany, E. Grimson, and A. Willsky, "Mutual information in coupled multi-shape model for medical image segmentation," *Med. Image Anal.* **8**, 429–445 (2004).
- ²²Z. Tu and A. Toga, "Towards whole brain segmentation by a hybrid model," *Lect. Notes Comput. Sci.* **4792**, 169–177 (2007).
- ²³J. Yang, L. Staib, and J. Duncan, "Neighbor-constrained segmentation with level set based 3-D deformable models," *IEEE Trans. Med. Imaging* **23**, 940–948 (2004).
- ²⁴D. Pham, *Robust Fuzzy Segmentation of Magnetic Resonance Images* (IEEE, Washington, D.C., 2001), pp. 127–131.
- ²⁵J. Wu and A. Chung, "A novel framework for segmentation of deep brain structures based on Markov dependence tree," *Neuroimage* **46**, 1027–1036 (2009).
- ²⁶A. Akhondi-Asl and H. Soltanian-Zadeh, "Constrained optimization of nonparametric entropy-based segmentation of brain structures," *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, Paris, France, 2008, pp. 41–44 (unpublished).
- ²⁷T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models—Their training and application," *Comput. Vis. Image Underst.* **61**, 38–59 (1995).
- ²⁸B. Hong, E. Prados, S. Soatto, and L. Vese, *Shape Representation Based on Integral Kernels: Application to Image Matching and Segmentation* (IEEE, Washington, D.C., 2006), Vol. 6, p. 3.
- ²⁹S. Pizer, D. Fritsch, P. Yushkevich, V. Johnson, and E. Chaney, "Segmentation, registration, and measurement of shape variation via image object shape," *IEEE Trans. Med. Imaging* **18**, 851–865 (1999).
- ³⁰O. Colliot, O. Camara, and I. Bloch, "Integration of fuzzy spatial relations in deformable models—Application to brain MRI segmentation," *Pattern Recogn.* **39**, 1401–1414 (2006).
- ³¹P. Bahman-Bijari, A. Akhondi-Asl, and H. Soltanian-Zadeh, "Interactive coupled object segmentation using symmetry and distance constraint," *Proceedings of the Third Cairo International Biomedical Engineering Conference, Cairo, 2006* (unpublished).
- ³²B. Scherrer, M. Dojat, F. Forbes, and C. Garbay, "LOCUS: Local cooperative unified segmentation of MRI brain scans," *Lect. Notes Comput. Sci.* **4791**, 219–227 (2007).
- ³³C. Ciofolo and C. Barillot, *Brain Segmentation with Competitive Level Sets and Fuzzy Control* (Springer, Berlin, 2005), Vol. 3565, pp. 333–344.
- ³⁴J. Maintz and M. Viergever, "A survey of medical image registration," *Med. Image Anal.* **2**, 1–36 (1998).
- ³⁵J. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.* **7**, 308–313 (1965).
- ³⁶S. Smith, "Fast robust automated brain extraction," *Hum. Brain Mapp* **17**, 143–155 (2002).
- ³⁷E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.* **33**, 1065–1076 (1962).
- ³⁸B. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability (Chapman and Hall, New York, 1986).
- ³⁹K. Bathe and A. Cimento, "Some practical procedures for the solution of nonlinear finite element equations," *Comput. Methods Appl. Mech. Eng.* **22**, 59–85 (1980).
- ⁴⁰IBSR, <http://www.med.harvard.edu/AANLIB/home.html>.
- ⁴¹EIGENTOOL, <http://radiologyresearch.org/eigentool.htm>.
- ⁴²SLICER, <http://slicer.org/>.
- ⁴³ITK, <http://www.itk.org/>.
- ⁴⁴Mathworks, <http://www.mathworks.com/>.
- ⁴⁵J. Zhou and J. Rajapakse, "Segmentation of subcortical brain structures using fuzzy templates," *Neuroimage* **28**, 915–924 (2005).
- ⁴⁶S. Gouttard, M. Styner, S. Joshi, R. Smith, H. Hazlett, and G. Gerig, *Subcortical Structure Segmentation Using Probabilistic Atlas Priors* (SPIE, Bellingham, WA, 2007), Vol. 6512, p. 65122J1–12.