Original article

Automated MRI brain tissue segmentation based on mean shift and fuzzy c-means using a priori tissue probability maps

Q. Mahmood [a,b], A. Chodorowski [a,b], M. Persson [a,b]

[a] *Signals and Systems, Chalmers University of Technology, 41296 Gothenburg, Sweden*
[b] *MedTech West, Sahlgrenska University Hospital, 41285, Gothenburg, Sweden*

Abstract

This paper presents a novel fully automated unsupervised framework for the brain tissue segmentation in magnetic resonance (MR) images. The framework is a combination of Bayesian-based adaptive mean shift, a priori spatial tissue probability maps and fuzzy c-means. Mean shift is employed to cluster the tissues in the joint spatial-intensity feature space and then a fuzzy c-means is applied with initialization by a priori spatial tissue probability maps to assign the clusters into three tissue types; white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF). The proposed framework is validated on a synthetic T1-weighted MR image with varying noise characteristics and spatial intensity inhomogeneity, obtained from the BrainWeb database as well as on 38 real T1-weighted MR images, obtained from the IBSR repository. The performance of the proposed framework is evaluated relative to the three widely used brain segmentation toolboxes: FAST, SPM and PVC, and the adaptive mean shift (AMS) and classical fuzzy c-means methods. The experimental results demonstrate the robustness of the proposed framework, and that it exhibits a higher degree of segmentation accuracy in segmenting both synthetic and real T1-weighted MR images compared to all competing methods.

1. Introduction

Accurate segmentation in magnetic resonance (MR) images plays a vital role in the quantitative analysis of normal and abnormal brain tissues [1–3]. It can also be useful in assigning individual tissues conductivity to construct realistic conductivity models for various neurological applications such as electroencephalography (EEG) source localization in epilepsy patients [4,5] and hyperthermia treatment planning for head and neck [6,7].

A wide range of automated segmentation methods has been proposed in the literature over the years. These can be broadly categorized into two major types, i.e. supervised and unsupervised segmentation methods.

Supervised segmentation methods [8–11] require labeled training datasets to extract the features and train a classifier. The classifier is then used to label unseen voxels. Artificial neural networks, k nearest neighbor (kNN) and support vector machines are examples of supervised segmentation methods. A major downside of the supervised methods is that they need a sufficiently large training dataset from a similar distribution as the data to be segmented. Therefore in practice these methods cannot be useful to the data obtained with a different scanning protocol and image modality or scanner [12].

Unsupervised segmentation methods don't require any labeled training datasets. The majority of unsupervised methods that have been proposed for automated segmentation of brain tissues are based on statistical parametric models [13–21]. These methods assume some distributional form for the underlying probability distribution of the data and seek to estimate its parameters. Some of these [13–15] are purely voxel wise intensity based clustering methods. A downside of these is that they may give poor tissue classifications in the presence of additive noise and multiplicative bias field [22]. To overcome these problems, some of the parametric methods [17–21] using a Markov random field (MRF) statistical spatial model. A ma-

jor drawback with these approaches is that the MRF algorithm is computationally expensive and requires critical parameters settings at higher dimensional feature space [22].

Mean shift (MS) [23,24] is one of the unsupervised clustering methods, which doesn't have this problem. It is an adaptive gradient approach to estimate the modes of the multivariate distribution underlying the feature space. The feature points that are associated with a mode form a cluster. The only parameter that influences the clustering is the bandwidth of the kernel. However, the use of a fixed bandwidth can cause over-clustering or under-clustering. Several approaches [25,26] have been proposed to solve this problem, wherein the bandwidth of each feature point is used to estimate the clusters. Therefore, the mean shift based on the adaptive bandwidth of the kernel is called adaptive mean shift (AMS).

AMS can also provide clustering by taking both the spatial and the intensity domain into account. Due to this characteristic, AMS can be more robust to noise and spatial intensity inhomogeneity artifacts in the MR brain images compared to intensity based clustering methods [22]. The output of the AMS is a set of clusters or modes. In order to get the desired number of clusters or tissue classes, merging is required.

In [22], the first adaptive mean shift framework is proposed for segmenting the brain tissues in the MR images. The framework is based on the mode merging and voxel-weighted k-means algorithm to categorize the clusters, obtained from the adaptive mean shift, into WM, GM and CSF tissue.

A downside of this procedure is that mode pruning in a range (intensity) domain can lose spatial information of modes (clusters), which may cause combining of the modes belonging to different tissue classes. Another downside of it is that the final merging of pruned modes into three tissue types, using the prior knowledge about the ordering of tissue intensity [22] in MR images to initialize the voxel-weighted k-means algorithm, may also lead to assigning the clusters to the wrong tissue class.

To overcome these problems, we here propose a new unsupervised segmentation framework, wherein we incorporate the spatial priors of the tissues to assign the clusters, obtained from the adaptive mean shift, into the three tissue types; WM, GM and CSF. The proposed framework is based on the Bayesian-based adaptive mean shift, a priori spatial tissue probability maps and the fuzzy c-means algorithm.

The organization of the paper is as follows. Section 2 describes the MRI data, proposed framework, overview of competing methods, segmentation performance measure, and statistical analysis. The experimental results for both synthetic and real MRI data are presented in Section 3. Section 4 discusses the results and finally, the conclusions are drawn in Section 5.

2. Materials and methods

2.1. MRI data

We used a synthetic as well as real MRI data to evaluate the performance of proposed framework relative to the competing methods. The real MRI data is composed by two sets of MRI scans of normal subjects, obtained from the Internet Brain Segmentation Repository (IBSR) [27]. One of the sets of MRI scans is known as IBSR18 while the other is known as IBSR20.

Synthetic dataset The synthetic dataset comprises a T1-weighted MRI scan of a normal subject for four (3%, 5%, 7% and 9%) different noise levels with two (20% and 40%) different spatial intensity inhomogeneity levels, obtained from the BrainWeb simulated brain database (SBD) [28]. The T1-weighted MRI scan is of size $181 \times 217 \times 181$ with cubic voxels of size 1 mm × 1 mm × 1 mm. A ground truth (labeled data) for the T1-wiegthed MRI scan was also obtained from the SBD.

IBSR18 dataset The IBSR18 dataset comprises 18 real T1-weighted MRI scans of normal subjects. Each MRI scan is of size $256 \times 256 \times 128$ with voxels of size 0.94 mm × 0.94 mm × 1.5 mm. A ground truth (manual segmentation) for each T1-weighted MRI scan for this dataset was obtained from the IBSR repository.

IBSR20 dataset The IBSR20 dataset comprises 20 real T1-weighted MRI scans of normal subjects. Each MRI scan is of size $256 \times 63 \times 256$ with voxels of size 1 mm × 3.1 mm × 1 mm. Each MRI scan in the dataset suffered from low contrast and bias field (spatial intensity inhomogeneity). A ground truth (manual segmentation) for each T1-weighted MRI scan for this dataset was also obtained from the IBSR repository.

2.2. Bayesian adaptive mean shift

Bayesian adaptive mean shift is a variation on the adaptive mean shift (AMS) [25,26] segmentation method originally proposed in [22] for brain tissue segmentation in MR images. In [22], the adaptive bandwidth of the kernel for the mean shift algorithm is defined in terms of the distance between the current feature point and its k-th nearest neighbor. However, the bandwidth value defined using this approach can be biased by outliers [29].

In [29], a global bandwidth estimation approach is proposed for the kernel that does not have this problem. In Bayesian adaptive mean shift, we used this approach locally for adaptive bandwidth estimation of the kernel for mean shift algorithm. The approach is based on a Bayesian method that involves fitting the Gamma distribution probability density function to the local variances of N sets of neighborhoods around the current feature point. See Appendix A for more details.

Mean shift is itself an iterative algorithm for finding the modes of a multivariate probability density function given discrete data sampled from it. Let $\{\mathbf{x}_i \in \mathbb{R}^d \mid i = 1, \dots, n\}$ denote this set of points. Given a starting point $\mathbf{y}_1$, the following iterative rule defines successive locations of this point towards a denser region or mode (local maximum):

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^{n} \frac{1}{h_i^{d+2}} \mathbf{x}_i \, g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h_i}\right\|^2\right)}{\sum_{i=1}^{n} \frac{1}{h_i^{d+2}} g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h_i}\right\|^2\right)} \quad j = 1, 2, \dots \quad (1)$$

where $g(x)$ is the kernel profile of kernel G, and $h(\mathbf{x}_i) \equiv h_i$ is the adaptive bandwidth of the kernel for point $\mathbf{x}_i$. The points
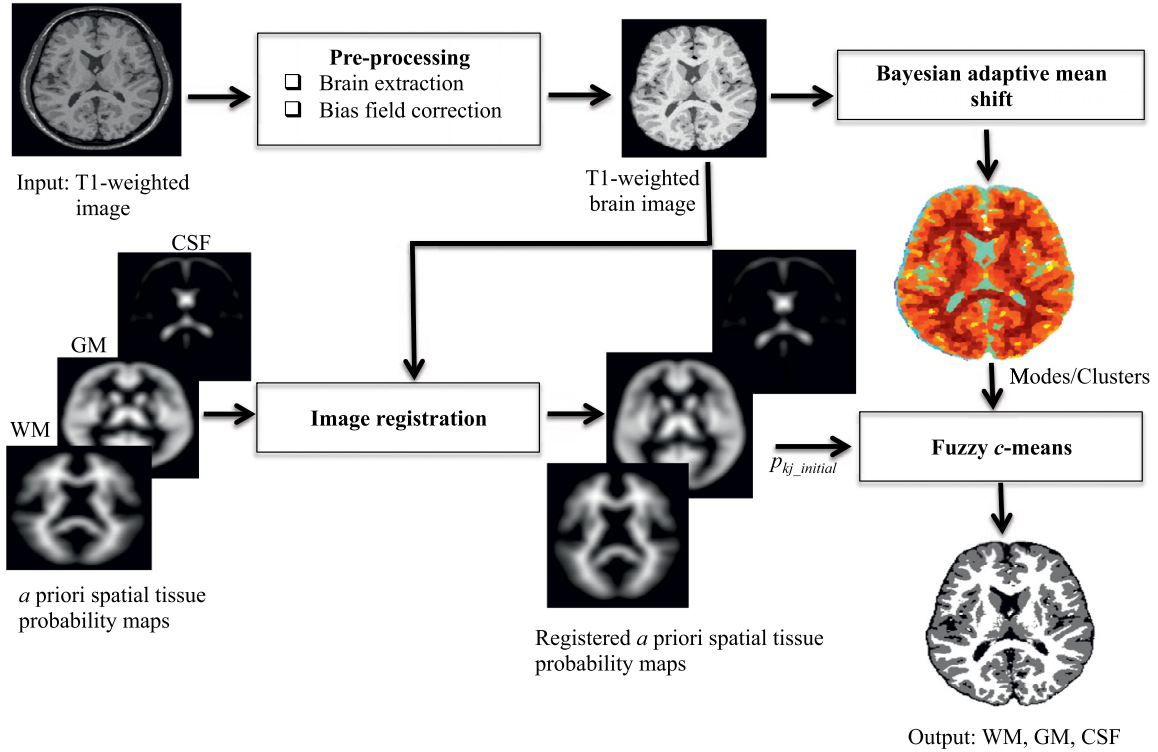
Fig. 1. Schematic procedure of the proposed framework.

that converge to the same mode constitute a cluster. To apply this to the problem of image segmentation one represents each pixel as a feature point $\mathbf{x}_i$ formed by concatenating its spatial coordinates and range values (e.g. T1-weighted) and employs the following joint spatial-intensity domain kernel G defined

$$G(\mathbf{x}) = \frac{c}{h_s^p h_r^d} g\left(\left\|\frac{\mathbf{x}^s}{h_s}\right\|^2\right) g\left(\left\|\frac{\mathbf{x}^r}{h_r}\right\|^2\right) \tag{2}$$

where $\mathbf{x}^s$ and $\mathbf{x}^r$ are the spatial and range components of $\mathbf{x}$, and h_s and h_r are the spatial and range bandwidths respectively. In this work, we used Gaussian kernel for both spatial and intensity domain. The output of Bayesian adaptive mean shift is a set of modes or clusters of brain tissues.

2.3. Fuzzy c-means algorithm

Let $\{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \ldots, \mathbf{z}_m\}$ denote a set of m clusters obtained from the Bayesian adaptive mean shift. The fuzzy c-means [30] is then applied to assign these clusters to the three tissue types by minimizing the cost function defined as

$$J = \sum_{j=1}^{m} \sum_{k=1}^{c} p_{kj}^g \|\mathbf{I}_j - \boldsymbol{\mu}_k\|^2 \quad j = 1, 2, \ldots, m \tag{3}$$

where $\mathbf{I}_j$ represents an intensity vector of cluster $\mathbf{z}_j$, $\boldsymbol{\mu}_k$ is the kth cluster (tissue) center and p_{kj} is known as membership function and it represents the probability that an intensity vector $\mathbf{I}_j$ of cluster $\mathbf{z}_j$ belongs to kth tissue. The constant g controls the fuzziness of the resulting partition.

To initialize the tissue centers, a priori spatial tissue probability maps are incorporated into the fuzzy c-means algorithm. The kth tissue center is then initialized as

$$\boldsymbol{\mu}_{k\_initial} = \frac{\sum_{j=1}^{m} p_{kj\_initial}^g \mathbf{I}_j}{\sum_{j=1}^{m} p_{kj\_initial}^g} \tag{4}$$

where $p_{kj\_initial}$ represents a priori spatial probability map of kth tissue, obtained from the International Consortium for Brain Mapping (ICBM) [31,32]. The membership functions and tissue centers are then updated as

$$p_{kj} = \frac{1}{\sum_{i=1}^{c} \left(\frac{\|\mathbf{I}_j - \boldsymbol{\mu}_k\|}{\|\mathbf{I}_j - \boldsymbol{\mu}_i\|}\right)^{2/(g-1)}} \tag{5}$$

$$\boldsymbol{\mu}_k = \frac{\sum_{j=1}^{m} p_{kj}^g \mathbf{I}_j}{\sum_{j=1}^{m} p_{kj}^g} \tag{6}$$

2.4. Summary of the proposed segmentation framework

Herein, we summarize the proposed segmentation framework. The schematic procedure of the proposed framework is shown in Fig. 1.

The proposed framework includes the following pre-processing steps: (1) Removal of non-brain tissue (such as skull, skin, fat, muscle, and spinal cord) from the MRI data (T1-weighted image) using a brain binary mask, generated from the provided ground truth. (2) Correction of Bias field using the N3 bias field correction algorithm [33], and (3) Co-registration of a priori spatial tissue probability maps, obtained from the ICBM [31,32], to the MRI brain data (T1-weighted brain image) by employing the Flirt registration tool in FSL [34]. Given

that these preliminary requirements are satisfied, the proposed framework segmenting the MRI brain into three tissue types; WM, GM and CSF is as follows.

1. The adaptive bandwidth h_i for each feature point $\mathbf{x}_i$ was estimated by employing the Bayesian-based estimator as described in Appendix A.
2. The modes or clusters $\{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \ldots, \mathbf{z}_m\}$ of the MRI brain tissue were then computed in Eq. (1) using the adaptive bandwidth h_i, obtained from the step 1. The clustering was done in the joint spatial-intensity domain using the joint kernel, defined in Eq. (2).
3. Finally, the fuzzy c-means algorithm was applied to categorize the clusters, obtained from the step 2, into the WM, GM, and CSF tissue by employing Eq. (3), wherein the center of the kth tissue $\boldsymbol{\mu}_{k\_initial}$ was initialized by incorporating the a priori spatial tissue probability maps $p_{kj\_initial}$ (obtained from the ICBM) using Eq. (4).

2.5. Overview of competing segmentation methods

The performance of the proposed framework was evaluated relative to the three widely used brain segmentation toolboxes: FAST (FMRIB's Automated Segmentation Tool) [18,35,36], SPM (Statistical Parametric Mapping toolbox) [32,35–37], and PVC (Partial Volume Classifier) [19,20,35] as well as to the existing adaptive mean shift (AMS) framework [22] and the classical fuzzy c-means method [30].

FAST   FAST is implemented in the FSL [38] software. In FAST, the underlying method for brain tissue segmentation is based on the hidden Markov random field model and associated Expectation-Maximization (HMRF-EM) algorithm [18].

SPM   The underlying method [37] for the brain tissue segmentation in SPM [32] is based on the parameter estimations of Gaussian mixer model (GMM), atlas registration and bias field correction at the same time iteratively. In this study, the latest SPM version we used for segmenting the brain tissues was SPM8.

PVC   PVC is implemented in the BrainSuite software [39]. The underlying method [19,20] for the brain tissue segmentation in PVC is based on the maximum-a-posteriori (MAP) classifier and spatial prior model of the brain.

Adaptive mean shift (AMS) framework   AMS framework is initially based on the adaptive mean shift to divide the brain tissue into a large number of clusters and then on the iterative mode pruning and voxel-weighted k-means to categorize these clusters into the three tissue types: WM, GM and CSF [22].

Fuzzy c-means (FCM)   FCM [30] is frequently used in the pattern recognition field. It is based on the distance function to partition the input data into clusters and it doesn't take into account spatial information.

2.6. Quantification of segmentation performances

The segmentation performance for each segmentation method was evaluated quantitatively using the Dice index (DI) [40].

The Dice index (DI) was computed for each tissue type, dataset, and segmentation method. The Dice index (DI) measures the degree of overlap between the ground truth and the segmentation result. It is defined as

$$DI = \frac{2V_{ae}}{(V_a + V_e)} \tag{7}$$

where V_{ae} is the number of voxels the segmentation result and the ground truth have in common, and V_a and V_e denote the number of voxels in the segmentation result and the ground truth respectively. The DI has value one for perfect segmentation and zero when there is no overlap between the segmentation result and ground truth.

2.7. Optimal parameters settings for the segmentation methods

The T1-weighted MRI scan for 0% noise level with 0% bias field level, obtained from the BrainWeb database [28], was used to empirically determine the optimal parameters settings for all segmentation methods.

The optimal parameters for each segmentation method, which gave the highest mean Dice index over all the brain tissues for this particular MRI scan were selected.

For the proposed framework, we set h_s (*spatial bandwidth*) = 3, and $M_1 = 100$, $M_2 = 300$, and $N = 10$ (parameters involved in estimating the adaptive bandwidth, described in Appendix A), and $g = 1.5$ (parameter defined in fuzzy c-means).

For the AMS framework, the parameter 'k' (nearest neighborhoods) was set to 120 for the adaptive bandwidth estimation of the kernel for the mean shift as suggested in [22].

For the FCM, we set $g = 1.5$ and it was initialized using the prior knowledge of tissue intensity ordering in T1-weighted MRI scan. For example, in the T1-weighted MRI scan the highest intensity value was used to initialize the cluster for WM tissue, the lowest intensity value was used to initialize the cluster for CSF tissue, and the mean of the highest and lowest intensity values was used to initialize the cluster for GM tissue.

In this study, the brain segmentation toolboxes: FAST, SPM and PVC were applied with the same parameter settings as described in [35].

2.8. Statistical analysis

To determine whether there exists a statistically significant difference in the voxel-wise classification performance between the proposed framework and each of the other methods for each tissue type and each dataset, several multiple comparison tests were performed. Each multiple comparison test involved performing five McNemar tests [41], each comparing proposed

framework to one of the other methods. Each McNemar test involved computing a 2×2 contingency table $\begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{bmatrix}$ where n_{11} is the number of voxels correctly classified by both methods, n_{12} is the number of voxels correctly classified by proposed framework but not the other method, n_{21} is the number of voxels incorrectly classified by proposed framework but correctly classified by the other method, and n_{22} is the number of voxels incorrectly classified by both methods.

For each McNemar test the null hypothesis was that the two methods have the same performance or error rate, i.e. $n_{12} = n_{21}$, versus the alternative hypothesis that they do not. The level of significance for each multiple comparison test was taken to be $\alpha = 0.05$ and so, using Bonferroni correction, the level of significance for each McNemar test was $\alpha = 0.05 \div 5 = 0.01$. Each McNemar test in essence tests whether the two methods classify in the same way; i.e. make the same misclassification errors.

3. Experimental results

3.1. Quantitative results

Synthetic dataset For the synthetic dataset, the Dice index for each noise level with 20% and 40% spatial intensity inhomogeneity levels for each tissue and segmentation method are presented in Figs. 2 and 3 respectively. The mean Dice index over all the noise levels for 20% and 40% spatial intensity inhomogeneity levels for each tissue and segmentation method are presented in Figs. 4 and 5 respectively.

IBSR18 dataset The Dice index for each subject for each tissue and segmentation method is presented in Fig. 6 and the mean Dice index over all the subjects for each tissue and segmentation method is presented in Fig. 7.

IBSR20 dataset The Dice index for each subject for each tissue and segmentation method is presented in Fig. 8 and the mean Dice index over all the subjects for each tissue and segmentation method is presented in Fig. 9.

3.2. Qualitative results

An example of the segmentation results for each method for the 9% noise level with 40% spatial intensity inhomogeneity level for axial slice 100 of the synthetic dataset is shown in Fig. 10.

An example of the segmentation results for each method for coronal slice 31 of the subject 15_3 from the IBSR20 dataset is shown in Fig. 11.

4. Discussion

We presented a novel fully automatic unsupervised segmentation framework for the segmentation of three tissue types from the MRI brain images. The framework is based on the Bayesian-based adaptive mean shift to initially divide the brain into a number of clusters and the fuzzy *c*-means algorithm,
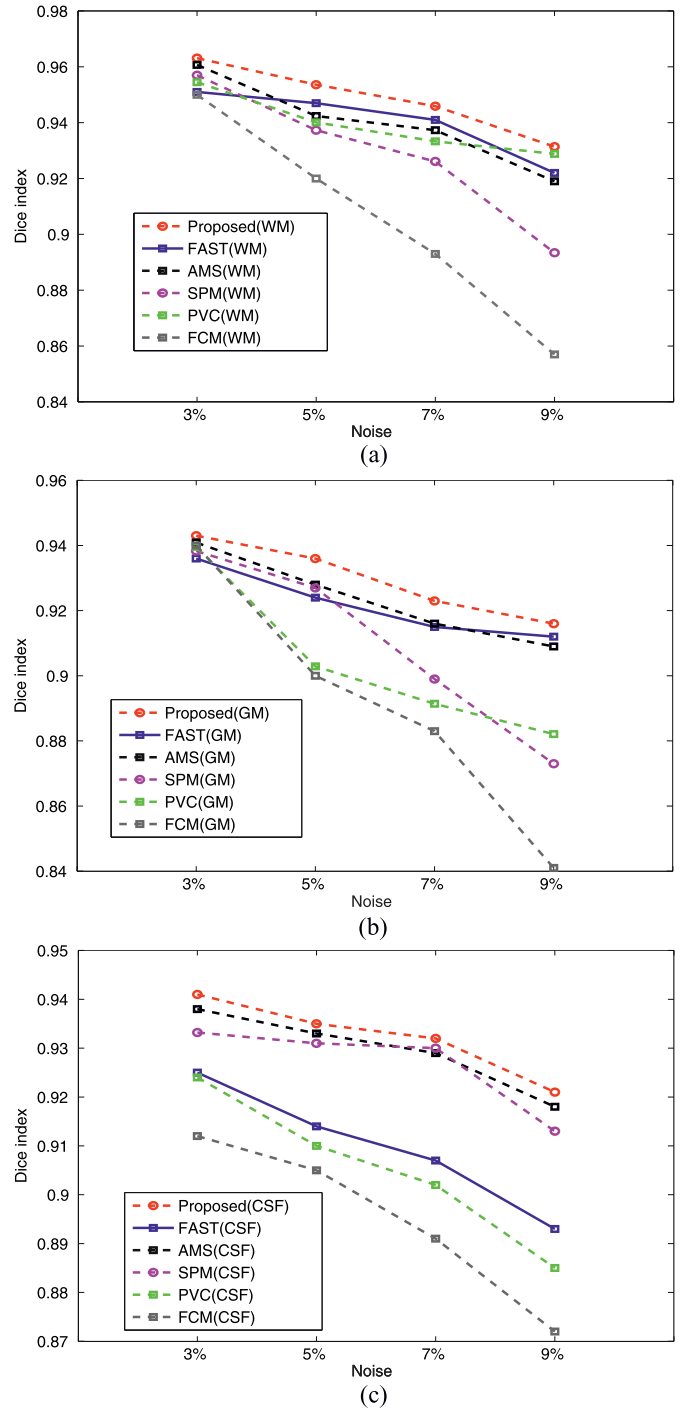

Fig. 2. Dice index for each segmentation method for the synthetic data for four different noise levels (3%, 5%, 7% and 9%) with 20% spatial intensity inhomogeneity level for (a) WM (b) GM and (c) CSF.

wherein the a priori spatial tissue probability maps are incorporated, to categorize the resulting clusters into WM, GM and CSF tissue.

We also presented the evaluation of the segmentation accuracy of the proposed framework relative to three widely used brain segmentation toolboxes: FAST, SPM, and PVC and the adaptive mean shift (AMS) and classical fuzzy *c*-means (FCM) methods. The evaluation was performed on a synthetic
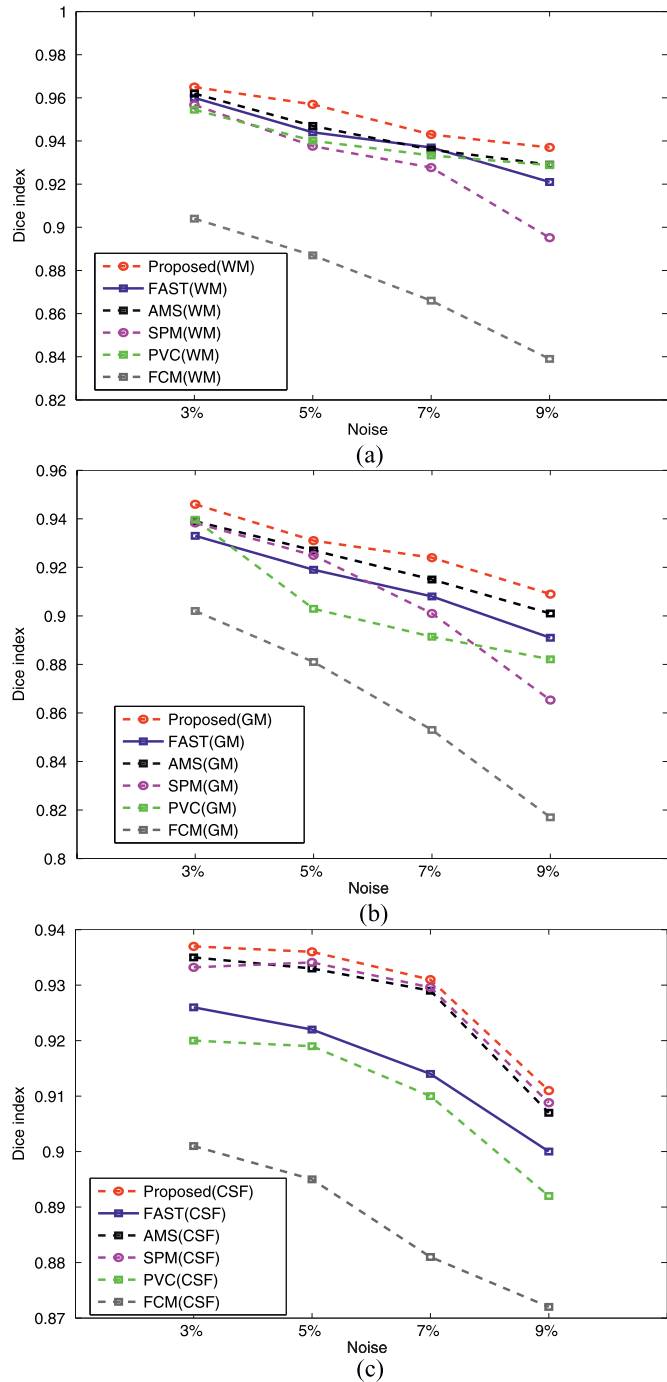
Fig. 3. Dice index for each segmentation method for the synthetic data for four different noise levels (3%, 5%, 7% and 9%) with 40% spatial intensity inhomogeneity level for (a) WM (b) GM and (c) CSF.
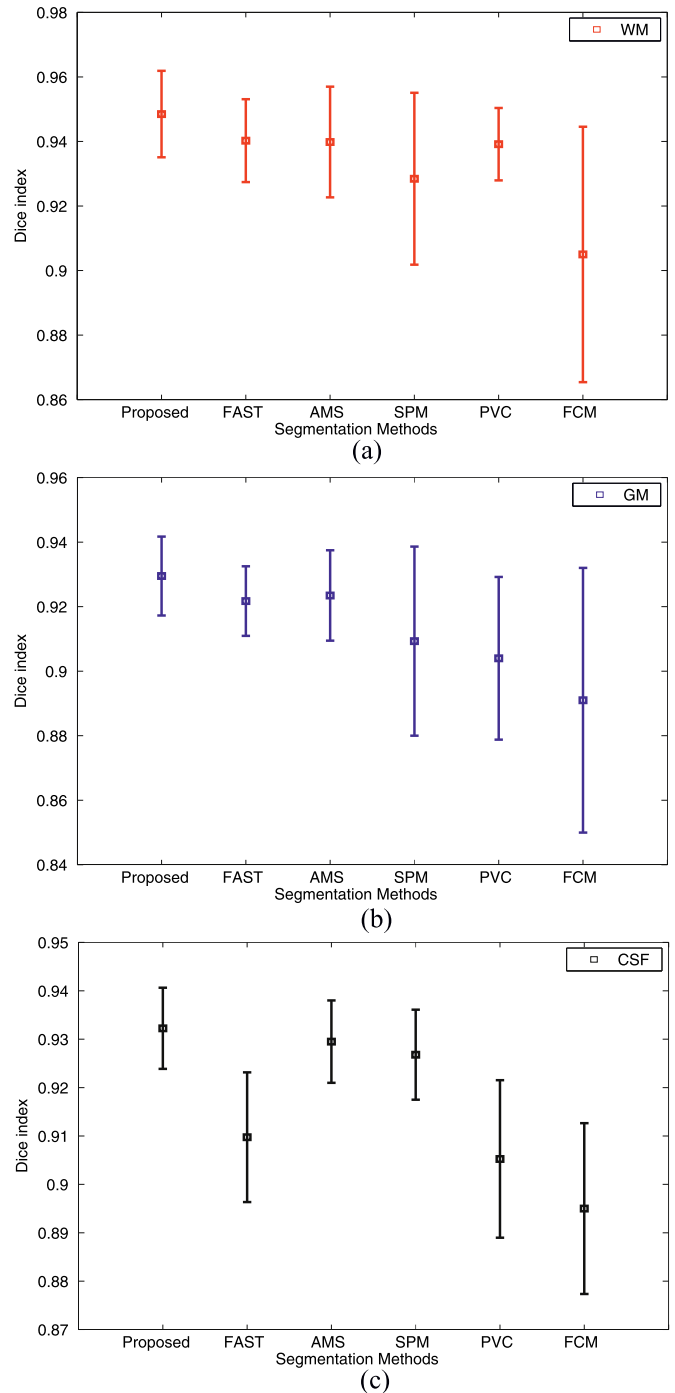
Fig. 4. Mean Dice index for each segmentation method for the synthetic data over all the noise levels with 20% spatial intensity inhomogeneity level for (a) WM (b) GM and (c) CSF. The whiskers show plus/minus one standard deviation.

T1-weighted MR image for four different levels of noise with two different levels of spatial intensity inhomogeneity, obtained from the BrainWeb database as well as on 38 real T1-weighted MR images, obtained from the IBSR repository.

The quantitative results for the synthetic dataset for each noise level with 20% spatial intensity inhomogeneity level (shown in Fig. 2) show that the proposed framework has better segmentation (higher Dice index) for each tissue compared to all competing methods except for the noise level 3% for which

the proposed framework is comparable to the AMS method for the WM and GM classification. Fig. 2 also shows that the proposed framework is comparable to the PVC for the noise level 9% for the WM classification.

The quantitative results for the synthetic dataset for each noise level with 40% spatial intensity inhomogeneity level (shown in Fig. 3) show that the proposed framework has better segmentation (higher Dice index) for each tissue compared to
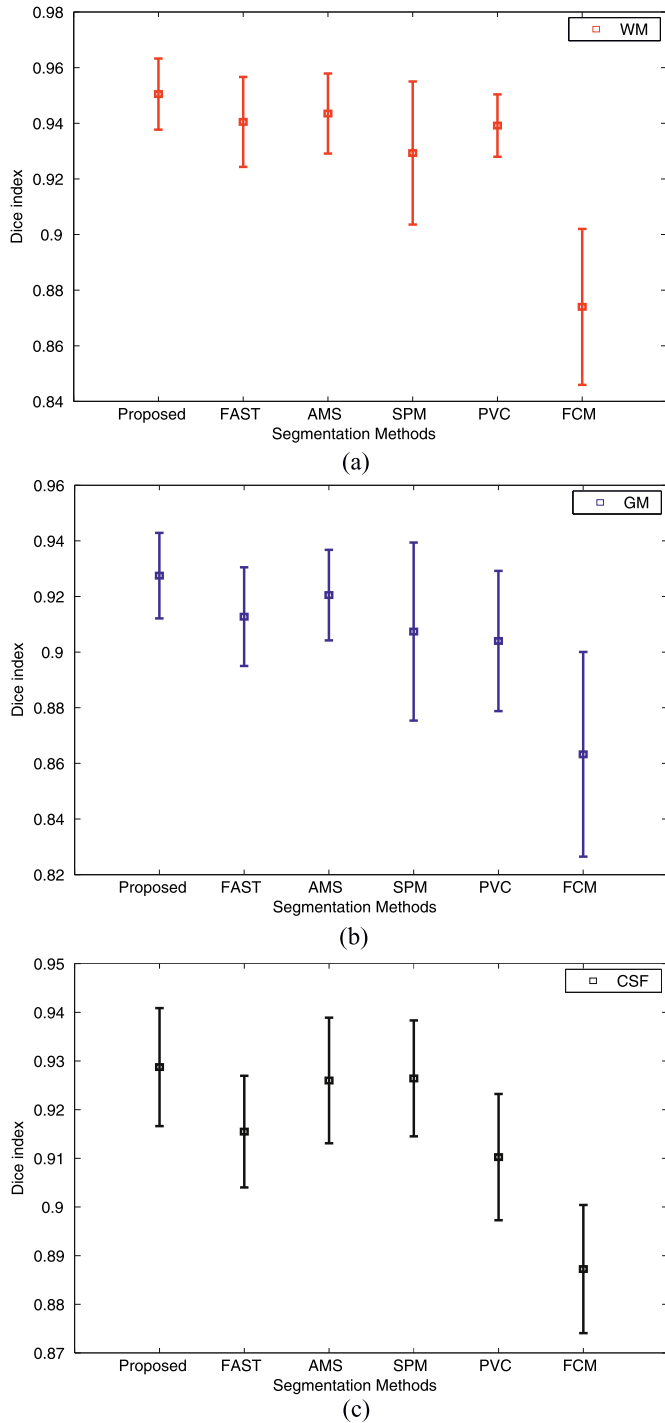
Fig. 5. Mean Dice index for each segmentation method for the synthetic data over all the noise levels with 40% spatial intensity inhomogeneity level for (a) WM (b) GM and (c) CSF. The whiskers show plus/minus one standard deviation.

Fig. 6. Dice index for each method for each subject from the IBSR18 dataset for (a) WM (b) GM and (c) CSF.

all competing methods except for the noise level 3% for which the proposed framework is comparable to the AMS method for the WM classification.

Figs. 4 and 5 show that on average over all the noise levels, the proposed framework is robust for each tissue type compared to all competing methods. Moreover, the results for the mean Dice index show that the proposed framework exhibits a de-
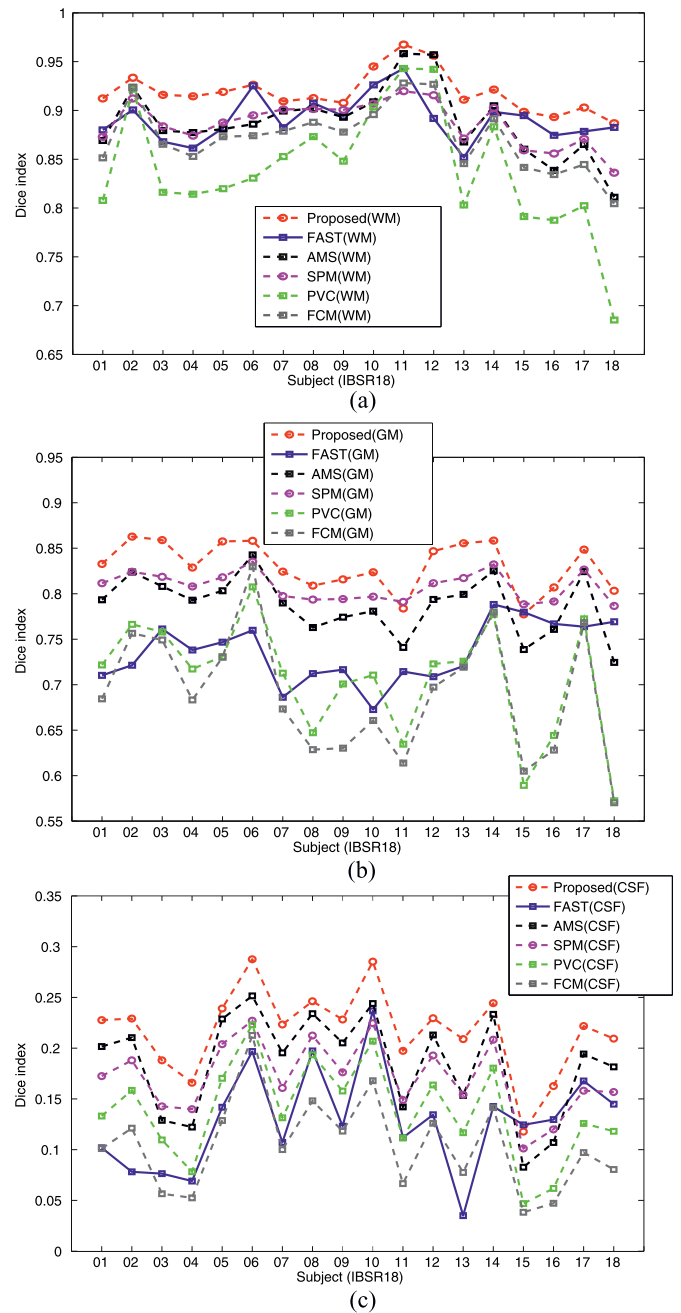
creased in performance for the CSF for 40% spatial intensity inhomogeneity level, FAST has decreased in performance for the CSF for both 20% and 40% spatial intensity inhomogeneity levels, AMS exhibits a decreased in performance for the WM for 20% spatial intensity inhomogeneity level and also has decreased in performance for the CSF for 40% spatial intensity inhomogeneity level, SPM exhibits a decreased in performance for the WM and GM for both 20% and 40% spatial intensity inhomogeneity levels, PVC has decreased in performance for the GM and CSF for both 20% and 40% spatial intensity inhomogeneity levels, and FCM exhibits a decreased in performance for each tissue type for both 20% and 40% spatial intensity in-
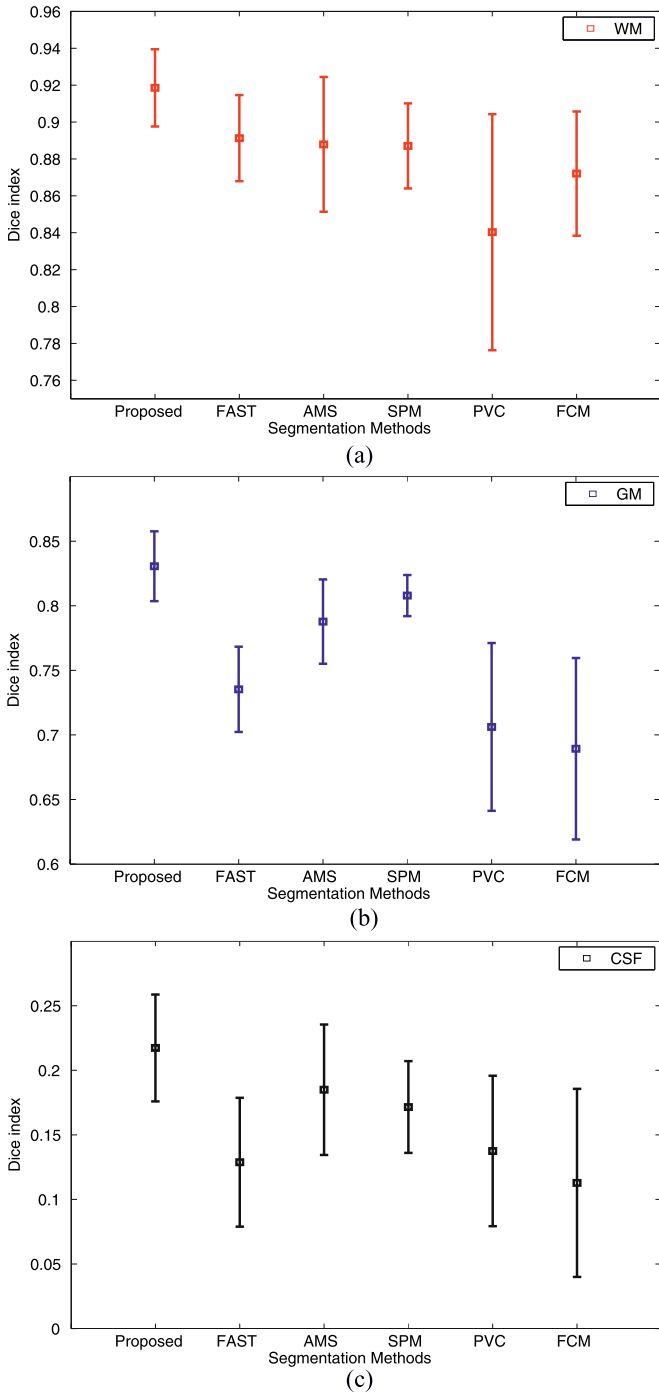
Fig. 7. Mean Dice index for each method over all the subjects of the IBSR18 dataset for (a) WM (b) GM and (c) CSF. The whiskers show plus/minus one standard deviation.

Fig. 8. Dice index for each method for each subject from the IBSR20 dataset for (a) WM (b) GM and (c) CSF.

homogeneity levels. The statistical tests (two-tailed t-tests) for the results for 20% spatial intensity inhomogeneity (shown in Fig. 4) reveal that no significant differences (*p-values* > 0.05) exist between the proposed framework and FAST, SPM, and AMS method for each tissue type. Moreover, they show that no significant differences (*p-values* > 0.05) exist between the proposed framework, and PVC and FCM method for the WM. However, significant differences (*p-values* < 0.05) exist be-
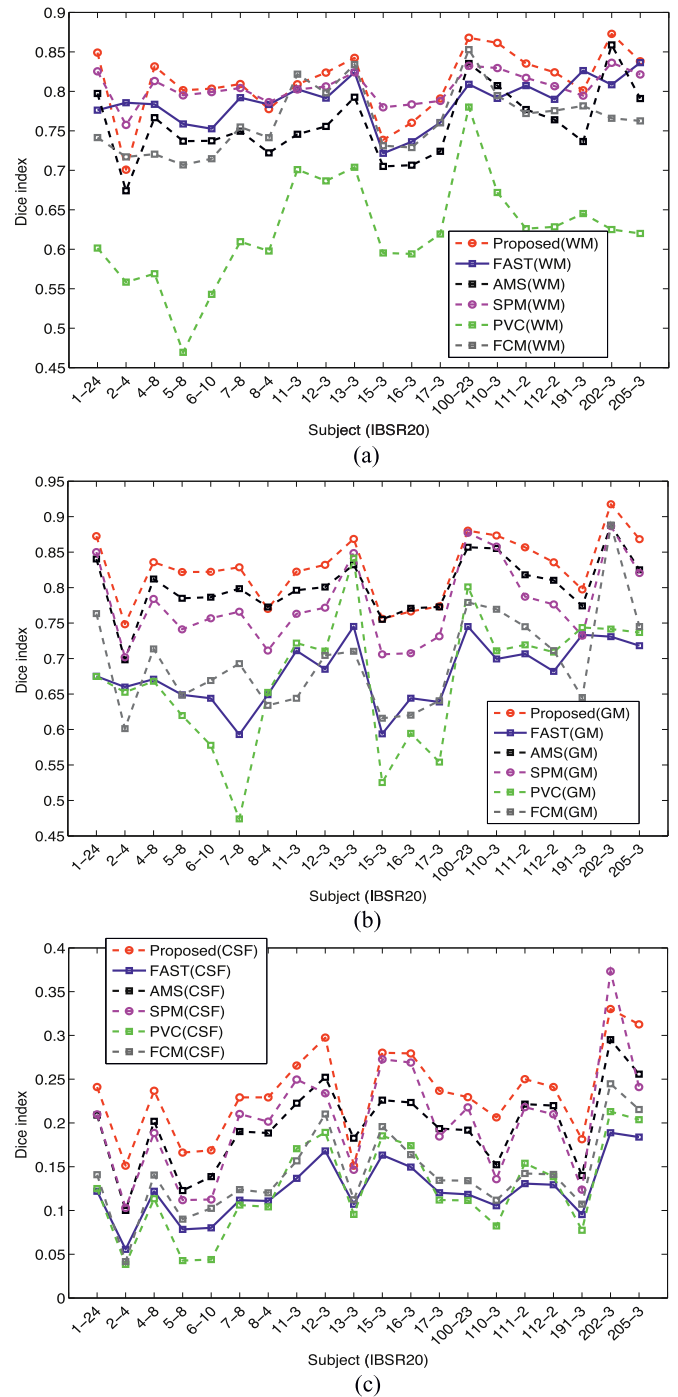
tween the proposed framework and PVC, and FCM method for the GM and CSF. The statistical tests (two-tailed t-tests) for the results for 40% spatial intensity inhomogeneity (shown in Fig. 5) show that no significant differences (*p-values* > 0.05) exist between the proposed framework and FAST, SPM, PVC, and AMS method for each tissue type. However, significant differences (*p-values* < 0.05) exist between the proposed framework and FCM method for each tissue type.
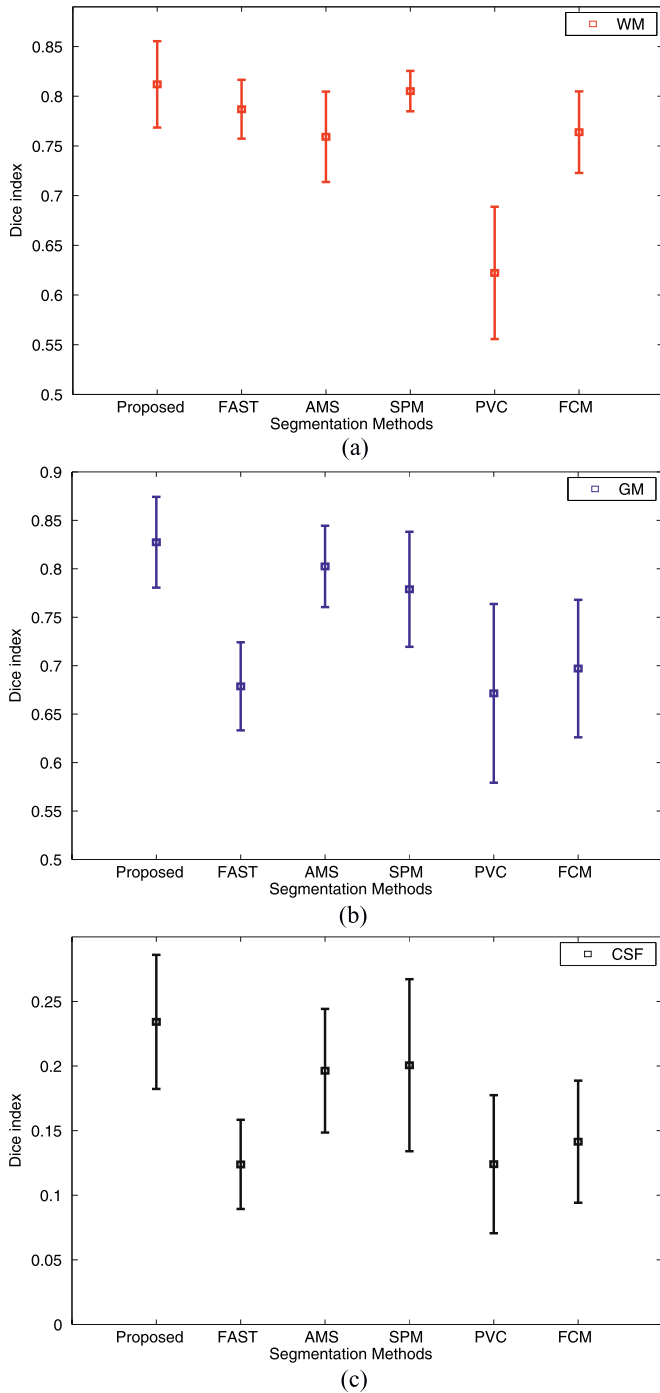
Fig. 9. Mean Dice index for each method over all the subjects of the IBSR20 dataset for (a) WM (b) GM and (c) CSF. The whiskers show plus/minus one standard deviation.

The quantitative results for the IBSR18 dataset (shown in Fig. 6) show that for the WM, the proposed framework has better segmentation (higher Dice index) for each subject compared to all competing methods except for the subjects 06, 08, 15, and 18 for which the proposed frame work is comparable (similar Dice index) to the FAST and the subject 12 for which the proposed frame work is comparable to the AMS method. For the GM, the proposed framework has better seg-

mentation (higher Dice index) for each subject compared to all competing methods except for the subjects 11, and 15 for which the SPM is slightly better than the proposed framework. For the CSF, the proposed framework has better segmentation (higher Dice index) for each subject compared to all competing methods except for the subject 15 for which the FAST is slightly better than the proposed framework. Fig. 7 shows that on average over all the subjects, the proposed framework is robust for each tissue type compared to all competing methods. Moreover, the results for the mean Dice index show that both proposed framework and FAST exhibit a decreased in performance for the CSF, AMS has decreased in performance for the WM and CSF, SPM exhibits a decreased in performance for the WM and GM, PVC has decreased in performance for each tissue type, and FCM exhibits a decreased in performance for the GM and CSF. The statistical tests (two-tailed t-tests) for the results (shown in Fig. 7) show that significant differences ($p\text{-}values < 0.05$) exist between the proposed framework and each competing methods for each tissue type.

The quantitative results for the IBSR20 dataset (shown in Fig. 8) show that for the WM, the proposed framework has better and similar segmentation (higher and similar Dice index) for each subject compared to all competing methods except for the subjects 2_4, 15_3, and 16_3 for which the SPM has better segmentation, the subjects 2_4, and 191_3 for which the FAST performs better and the subject 11_3 for which the FCM method is better. For the GM, the proposed framework has better segmentation (higher Dice index) for each subject compared to all competing methods except for the subjects 8_4, 15_3, 16_3 and 17_3 for which the AMS is comparable (similar Dice index) and the subject 100_23 for which the SPM is comparable. For the CSF, the proposed framework has better segmentation (higher Dice index) for each subject compared to all competing methods except for the subject 13_3 for which the AMS performs better and the subject 202_3 for which the SPM is better. Fig. 9 shows that on average over all the subjects, the proposed framework is robust for each tissue type compared to all competing methods. Moreover, the results for the mean Dice index show that both proposed framework and AMS have decreased in performance for the CSF, FAST exhibits a decreased in performance for the GM and CSF, both SPM and FCM have decreased in performance for the GM and CSF and PVC exhibits a decreased in performance for each tissue type. The statistical tests (two-tailed t-tests) for the results (shown in Fig. 9) reveal that no significant differences ($p\text{-}values > 0.05$) exist between the proposed framework and FAST, AMS, SPM, and PVC for the WM. Moreover, they show that no significant differences ($p\text{-}values > 0.05$) exist between the proposed framework and AMS, and SPM for the GM and CSF. However, significant differences ($p\text{-}values < 0.05$) exist between the proposed framework and PVC for the WM and also significant differences ($p\text{-}values < 0.05$) exist between the proposed framework and FAST, PVC, and FCM method for the GM and CSF.
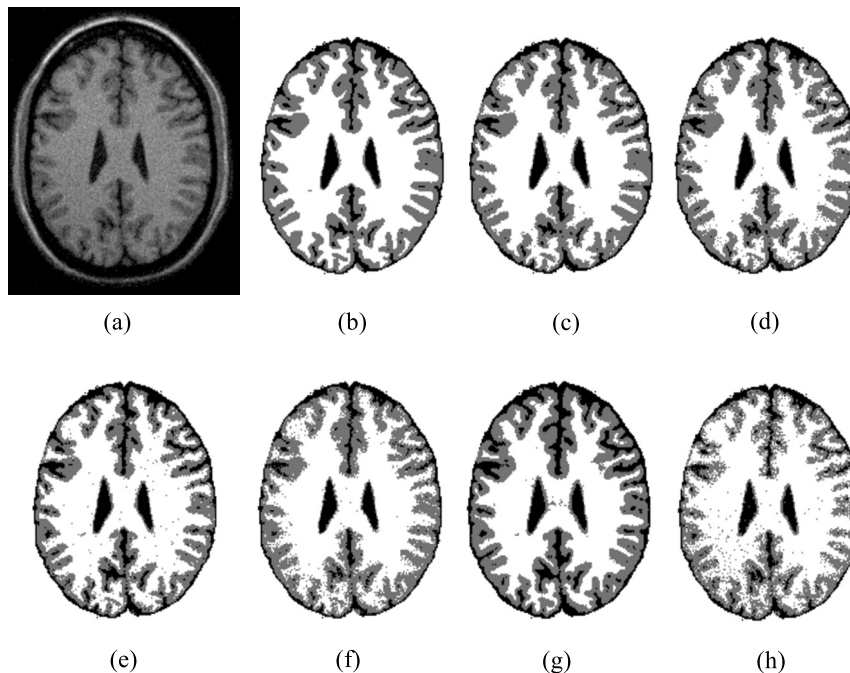
Fig. 10. Segmentation example for axial slice 100 of the synthetic dataset for 9% noise level with 40% spatial intensity inhomogeneity level: (a) T1-weighted (b) Ground truth (c) Proposed (d) AMS (e) FAST (f) SPM (g) PVC (h) FCM (WM in white, GM in gray and CSF in black).
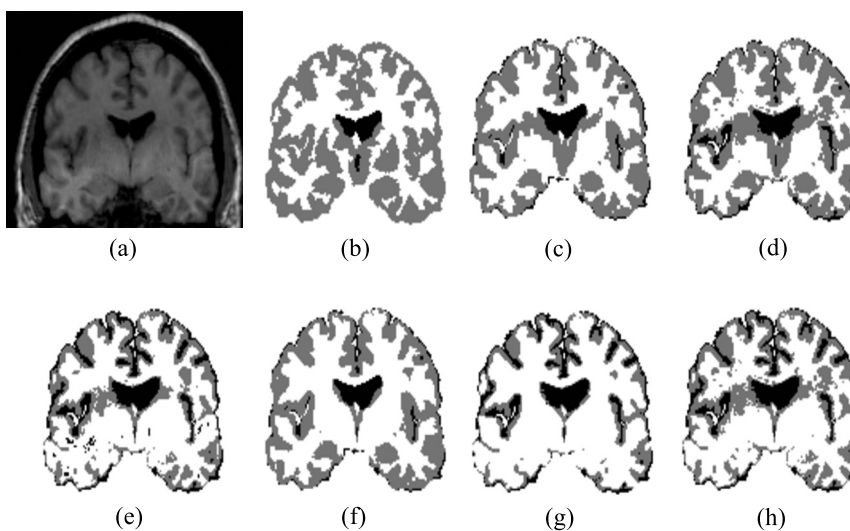

Fig. 11. Segmentation example for coronal slice 31 of the subject 15_3 from the IBSR20 dataset: (a) T1-weighted (b) Ground truth (c) Proposed (d) AMS (e) FAST (f) SPM (g) PVC (h) FCM (WM in white, GM in gray and CSF in black).

In this study, for the synthetic dataset, the segmentation results for the FAST toolbox for the WM and GM are comparable to those published in [21]. For the PVC, the results for the WM and GM are similar to those reported in [19] and for the SPM (SPM8), the results are comparable to those published in [37]. However, the results for the CSF for the FAST toolbox are higher than those published in [21]. The reason is that in [21] a different brain binary mask was used to remove the non-brain tissue that affects the accuracy of CSF classification.

For both the IBSR18 and the IBSR20 datasets, the segmentation results for the brain segmentation toolboxes (FAST, SPM and PVC) are comparable to those published in [35].

For each dataset and tissue, the McNemar tests provide evidence that the proposed framework is significantly different (*p-values* < 0.01) to all competing methods.

The qualitative results (shown in Fig. 10) show that compared to all segmentation methods, the proposed framework is less sensitive to noise and spatial intensity inhomogeneity; in particular for the segmentation of WM and GM tissue.

The qualitative results (shown in Fig. 11) show that relative to the ground truth, the proposed framework has a higher tendency to misclassify the WM as GM tissue compared to the SPM, which has more smooth WM segmentation especially in a region close to the ventricles.

5. Conclusions

We have proposed and evaluated a fully automatic unsupervised framework for segmenting the WM, GM and CSF tissue in MR brain images. We demonstrated the robustness of our proposed framework on the synthetic T1-weighted MR image with varying noise characteristics and spatial intensity inhomogeneity as well as on 38 real T1-weighted MR images. The experimental results show the accuracy and efficacy of the proposed framework, and that it consistently outperforms the three widely used brain segmentation toolboxes: FAST, SPM and PVC as well as the AMS and FCM methods for each tissue classification for all datasets. Moreover, incorporation of a priori spatial tissue probability maps in the proposed framework makes the tissue segmentation objective and reproducible.

For future work, we will extend the proposed framework for segmenting abnormal tissues such as multiple sclerosis lesions and tumors using multi-modal MR images.

Acknowledgements

Appendix A

A.1. Bayesian-based adaptive bandwidth estimator

The bandwidth is modeled [29] by the a posteriori probability density function $p(s \mid \mathbf{x})$ of local data spread or variance s given the data (feature) point $\mathbf{x}$. Let $M < n$ be the number of nearest neighbors to a data sample $\mathbf{x}_i$. We can then define the pseudolikelihood

$$P(s \mid \mathbf{x}) = \prod_{j=1}^{N} P(s \mid \mathbf{x}_{M_j}) \tag{A.1}$$

where $P(s \mid \mathbf{x}_{M_j})$ is the probability of local data spread s depending on the M_j nearest neighborhood samples to $\mathbf{x}_{M_j}$ and $\{M_j \mid j = 1, \ldots, N\}$ is the set of N such neighborhoods of various sizes. The evaluation of these probabilities over the entire set of M_j is then given by

$$P(s \mid \mathbf{x}_{M_j}) = \int P(s \mid M_j, \mathbf{x}_{M_j}) P(M_j \mid \mathbf{x}_{M_j}) dM_j \tag{A.2}$$

Applying Bayes rule we get

$$P(M_j \mid \mathbf{x}_{M_j}) = \frac{P(\mathbf{x}_{M_j} \mid M_j) P(M_j)}{P(\mathbf{x}_{M_j})} \tag{A.3}$$

where $P(\mathbf{x}_{M_j} \mid M_j)$ is the probability of the data sample $\mathbf{x}_{M_j}$ given the M_j nearest neighborhood. Hereinafter $P(M_j)$ is considered to have uniform distribution on the interval $[M_1, M_2]$. Several values are selected for M_j in this interval according to

$$M_j = M_1 + j \frac{M_2 - M_1}{N} \tag{A.4}$$

For a given M_j the local variance s_j is computed as

$$s_j = \frac{\sum_{l=1}^{M_j} \|\mathbf{x}_{i,l} - \mathbf{x}_i\|^2}{M_j - 1} \quad i = 1, 2 \ldots, n, \ j = 1, 2, \ldots, N \tag{A.5}$$

where $\mathbf{x}_{i,l}$ is the l-th nearest neighbor to the data point $\mathbf{x}_i$. The distribution of variances is modeled as the Gamma distribution defined as

$$p(s \mid \alpha, \beta) = \frac{\beta^\alpha s^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta s} \quad s \geq 0, \ \alpha, \beta > 0 \tag{A.6}$$

where

$$\Gamma(t) = \int_0^\infty r^{t-1} e^{-r} dr \tag{A.7}$$

is the Gamma function, and α and β define the shape and the scale of the Gamma distribution respectively.

These parameters are estimated using the maximum likelihood approach [29]. The estimate of the adaptive bandwidth is identically the mean of this distribution, i.e.

$$\hat{h}(\mathbf{x}_i) = \hat{\alpha}\hat{\beta} \quad i = 1, 2, \ldots, n \tag{A.8}$$

References

[1] Courchesne E, Chisum HJ, Townsend J, Cowles A, Covington J, Egaas B, et al. Normal brain development and aging: quantitative analysis at in vivo MR imaging in healthy volunteers. Radiology 2000;3:672–82.

[2] Simmons A, Westman E, Muehlboeck S, Mecocci P, Vellas B, Tsolaki M, et al. MRI measures of Alzheimer's disease and the AddNeuroMed study. Ann NY Acad Sci 2009;1180:47–55.

[3] Daryoush M, Kouzani AZ, Soltanian-Zadeh H. Segmentation of multiple sclerosis lesions in MR images: a review. Neuroradiology 2012;54:299–332.

[4] Rullmann M, Anwander A, Dannhauer M, Warfield SK, Duffy FH, Wolters CH. EEG source analysis of epileptiform activity using a 1 mm anisotropic hexahedra finite element head model. NeuroImage 2009;44:399–410.

[5] Shirvany Y, Mahmood Q, Edelvik F, Persson M, Hedstrom A, Particle JS. Swarm optimization applied to EEG source localization of somatosensory evoked potentials. IEEE Trans Neural Syst Rehabil Eng 2013;22:11–20.

[6] Fortunati V, Verhaart RF, van der Lijn F, Niessen WJ, Veenland JF, Paulides MM, van Walsum T. Hyperthermia critical tissues automatic segmentation of head and neck CT images using atlas registration and graph cuts. ISBI 2012:1683–6.

[7] Fortunati V, Verhaart RF, van der Lijn F, Niessen WJ, Veenland JF, Paulides MM, van Walsum T. Tissue segmentation of head and neck CT images for treatment planning: a multiatlas approach combined with intensity modeling. Med Phys 2013;40:071905.

[8] Zijdenbos A, Forghani R, Evans A. Automatic quantification of MS lesions in 3D MRI brain data sets: validation of INSECT. MICCAI 1998;1496:439–48.

[9] Cocosco C, Zijdenbos A, Evans A. A fully automatic and robust brain MRI tissue classification method. Med Image Anal 2003;7:513–27.

[10] Cárdenes R, Warfield SK, Macías EM, Santana JA, Ruiz-Alzola J. An efficient algorithm for multiple sclerosis segmentation from brain MRI. EUROCAST 2003:542–51.

[11] Akselrod-Ballin A, Galun M, Gomori JM, Basri R, Brandt A. Atlas guided identification of brain structures by combining 3D segmentation and SVM classification. MICCAI 2006:209–16.

[12] van Opbroek A, Ikram MA, Vernooij M, de Bruijne M. Supervised image segmentation across scanner protocols: a transfer learning approach. In: MLMI. Lecture Notes in Computer Science, vol. 7588. 2012. p. 160–7.

[13] Van Leemput K, Maes F, Vandermeulen D, Suetens P. Automated model-based tissue classification of MR images of the brain. IEEE Trans Med Imaging 1999;18:897–908.

[14] Van Leemput K, Maes F, Vandermeulen D, Suetens P. A unifying framework for partial volume segmentation of brain MR images. IEEE Trans Med Imaging 2003;22:105–19.

[15] Dugas-Phocion G, Ballester MÁG, Malandain G, Lebrun C, Ayache N. Improved EM-based tissue segmentation and partial volume effect quantification in multi-sequence brain MRI. MICCAI 2004:26–33.

[16] Greenspan H, Ruf A, Goldberger J. Constrained Gaussian mixture model framework for automatic segmentation of MR brain images. IEEE Trans Med Imaging 2006;25:1233–45.

[17] Marroquín JL, Vemuri BC, Botello S, Calderon F. An accurate and efficient Bayesian method for automatic segmentation of brain MRI. IEEE Trans Med Imaging 2002;21:934–44.

[18] Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation–maximization algorithm. IEEE Trans Med Imaging 2001;20:45–57.

[19] Shattuck DW, Sandor-Leahy SR, Schaper KA, Rottenberg DA, Leahy RM. Magnetic resonance image tissue classification using a partial volume model. NeuroImage 2001;13:856–76.

[20] Tohka J, Dinov ID, Shattuck DW, Toga AW. Brain MRI tissue classification based on local Markov random fields. Magn Reson Imaging 2010;28:557–73.

[21] Scherrer B, Forbes F, Garbay C, Dojat M. Distributed local MRF models for tissue and structure brain segmentation. IEEE Trans Med Imaging 2009;28:1278–95.

[22] Mayer A, Greenspan H. An adaptive mean-shift framework for MRI brain segmentation. IEEE Trans Med Imaging 2009;28:1238–49.

[23] Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans Inf Theory 1975;21:32–40.

[24] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell 2002;24:603–19.

[25] Comaniciu D. An algorithm for data-driven bandwidth selection. IEEE Trans Pattern Anal Mach Intell 2003;25:281–8.

[26] Georgescu B, Shimshoni I, Meer P. Mean-shift based clustering in high dimensions: a texture classification example. ICCV 2003:456–63.

[27] Internet Brain Segmentation Repository Center for Morphometric Analysis (IBSR). http://www.cma.mgh.harvard.edu/ibsr.

[28] BrainWeb. http://brainweb.bic.mni.mcgill.ca/brainweb/.

[29] Bors AG, Nasios N. Kernel bandwidth estimation for nonparametric modeling. IEEE Trans SMC 2009;39:1543–55.

[30] Bezdek JC. Pattern recognition with fuzzy objective function algorithms. Norwell, MA, USA: Kluwer Academic Publishers; 1981.

[31] International Consortium for Brain Mapping (ICBM). http://www.loni.ucla.edu/ICBM/Downloads/Downloads_ICBMprobabilistic.shtml.

[32] Statistical Parametric Mapping (SPM). http://www.fil.ion.ucl.ac.uk/spm/.

[33] Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 1998;17:87–97.

[34] Jenkinson M, Bannister PR, Brady JM, Smith SM. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage 2002;17:825–41.

[35] Valverde S, Oliver A, Cabezas M, Roura E, Lladó X. Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations. J Magn Reson Imaging 2014:1522–2586.

[36] Xu H, Thirion B, Allassonnière S. Bayesian estimation of probabilistic atlas for tissue segmentation. IRBM 2014;35:27–32.

[37] Ashburner J, Friston KJ. Unified segmentation. NeuroImage 2005;26:839–51.

[38] Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. NeuroImage 2012;62:782–90.

[39] BrainSuite Software. http://neuroimage.usc.edu/neuro/BrainSuite/.

[40] Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:297–302.

[41] Thomas GD. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput 1998;10:1895–919.