

Constrained Gaussian Mixture Model Framework for Automatic Segmentation of MR Brain Images

Hayit Greenspan Amit Ruf and Jacob Goldberger

Abstract— An automated algorithm for tissue segmentation of noisy, low contrast magnetic resonance (MR) images of the brain is presented. A mixture model composed of a large number of Gaussians is used to represent the brain image. Each tissue is represented by a large number of Gaussian components to capture the complex tissue spatial layout. The intensity of a tissue is considered a global feature and is incorporated into the model through tying of all the related Gaussian parameters. The Expectation-Maximization (EM) algorithm is utilized to learn the parameter-tied, constrained Gaussian mixture model. An elaborate initialization scheme is suggested to link the set of Gaussians per tissue type, such that each Gaussian in the set has similar intensity characteristics with minimal overlapping spatial supports. Segmentation of the brain image is achieved by the affiliation of each voxel to the component of the model that maximized the a-posteriori probability. The presented algorithm is used to segment 3D, T1-weighted, simulated and real MR images of the brain into three different tissues, under varying noise conditions. Results are compared with state-of-the-art algorithms in the literature. The algorithm does not use an atlas for initialization or parameter learning. Registration processes are therefore not required and the applicability of the framework can be extended to diseased brains and neonatal brains.

Index Terms— Image segmentation, MRI brain segmentation, Mixture of Gaussians, GMM, EM, Constrained model

I. INTRODUCTION

Automatic segmentation of brain MR images to the three main tissue types: white matter (WM), gray matter (GM) and cerebro-spinal fluid (CSF), is a topic of great importance and much research. It is known that volumetric analysis of different parts of the brain is useful in assessing the progress or remission of various diseases, such as Alzheimer's disease, epilepsy, sclerosis and schizophrenia [21].

A variety of segmentation schemes exist in the literature. As it is very difficult to estimate automatic or semi-automatic segmentation results against an in-vivo brain, manual segmentation by experts is still considered to be the “gold standard” or “ground truth” for any automated algorithm. However, manual partitioning of large amounts of low-contrast/low-SNR brain data is strenuous work and is prone to large intra- and inter-observer variability. Fully automated intensity-based algorithms, on the other hand, exhibit high sensitivity to various noise artifacts, such as intra-tissue noise, inter-tissue intensity contrast reduction, partial-volume effects and

others [16]. Reviews on methods for brain image segmentation (e.g., [21]) present the degradation in the quality of segmentation algorithms due to such noise, and recent publications can be found addressing various aspects of these concerns (e.g. partial-volume effect quantification [7]). Due to the artifacts present, classical voxel-wise intensity-based classification methods, such as K-means modeling and Mixture of Gaussians modeling (e.g., [25], [11]), may give unrealistic results, with tissue class regions appearing granular, fragmented, or violating anatomical constraints.

Incorporating spatial information via a statistical atlas provides a means for improving the segmentation results (e.g. [22], [13], [18]). The statistical atlas provides the prior probability for each pixel to originate from a particular tissue class. Algorithms that are based on the maximum-a-posteriori (MAP) criterion utilize the atlas information in the algorithm iterations to augment the information in the presence of noisy data. Co-registration of the input image and the atlas, a computationally intensive procedure, is critical in this scenario [23]. It is important to note that the quality of the registration result is strongly dependent on the physiological variability of the may converge to an erroneous result in the case of a diseased or severely damaged brain. Moreover, the registration process is applicable only to complete volumes. A single slice cannot be registered to the atlas, thus, cannot be segmented using these state-of-the-art algorithms.

An additional conventional method to improve segmentation smoothness and immunity to noise is to model neighboring voxels interactions using a Markov Random Field (MRF) statistical spatial model [13], [27], [9]. Smoother structures are obtained in the presence of moderate noise as long as the MRF parameters controlling the strength of the spatial interactions are properly selected. Too high a setting can result in an excessively smooth segmentation and a loss of important structural details [15]. In addition, MRF-based algorithms are computationally intractable unless some approximation is used which still requires computationally intensive algorithms. Algorithms that use deformable models to incorporate tissue boundary information [19] imply inherent smoothness but require careful initialization and precisely calibrated model parameters in order to provide consistent results in the presence of a noisy environment. Several works can be found in the literature, such as fuzzy connectedness segmentation methods, that attempt to provide an alternative to the MRF modeling (e.g. [20], [26]). In [26], as in many other works, there still seems to be a need for a large number of parameters for the task.

The objectives of the current work are threefold: 1) incorpo-

H. Greenspan is with the Tel-Aviv University, Tel Aviv 69978, Israel. E-mail: hayit@eng.tau.ac.il

A. Ruf is with the Tel-Aviv University, Tel Aviv 69978, Israel. E-mail: amitruf@gmail.com

J. Goldberger is with the Bar-Ilan University, Ramat-Gan 52900, Israel. E-mail: goldbej@eng.biu.ac.il

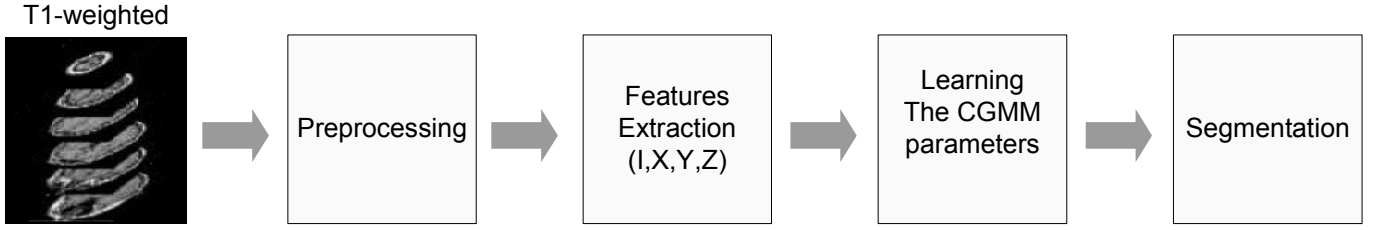


Fig. 1. Schematic description of the CGMM segmentation framework

rate spatial information within a statistically-based model in a simple and intuitive way, for both localization and smoothing; 2) achieve robust segmentation on real data along with strong resistance to noise, and 3) achieve accurate segmentation even in cases in which an atlas is not available or appropriate for the task. A robust, unsupervised, parametric method for segmenting 3D (or 2D) MR brain images with a high degree of noise and low contrast, is presented¹. We refer to the proposed framework as the *Constrained Gaussian Mixture Model* (CGMM) framework. In this framework, each tissue is modeled with multiple four-dimensional Gaussians, where each Gaussian represents a localized region (3 spatial features) and the intensity characteristic per region (T1 intensity feature). Incorporating the spatial information within the feature space is novel, as is using a large number of Gaussians per brain tissue to capture the complicated spatial layout of the individual tissues. Note that models to-date utilize the intensity feature only and use a single Gaussian per tissue type (e.g. [25], [18], [13]). The intensity of a tissue is considered a global feature and is incorporated into the model via parameter-tying: the intensity parameter of all the clusters that are related to the same tissue, is essentially the same, and is considered as a single parameter that appears several times in the model. The Expectation-Maximization (EM) algorithm is utilized to learn the parameter-tied, constrained Gaussian mixture model (GMM). An elaborate initialization scheme is suggested to link the set of Gaussians per tissue type, such that each Gaussian in the set has similar intensity characteristics with minimal overlapping spatial supports. Two key features of the proposed framework are: 1) combining global intensity modeling with localized spatial modeling, as an alternative scheme to MRF modeling, and 2) segmentation is entirely unsupervised; thus eliminating the need for atlas registration, or any intensity model standardization. The paper is organized as follows: a detailed description of the proposed framework is provided in Section II. Algorithm validation on both simulated and real brain volumes is presented in Section III. Discussion and conclusions are presented in Section IV.

II. THE CGMM SEGMENTATION FRAMEWORK

In this section a detailed description of the segmentation framework is presented. Figure 1 illustrates the four major stages of the framework. In a preprocessing stage the brain region is extracted and intensity inhomogeneities (bias) are corrected (based on [13] as described in section III). In a feature-extraction stage a four-dimensional feature vector is

extracted for each voxel v . The main feature is the voxel intensity, denoted by v^I . In order to include spatial information, the (X, Y, Z) position is appended to the feature vector. We use the notation $v^{XYZ} = (v^X, v^Y, v^Z)$ for the three spatial features.

The set of feature vectors extracted per volume is denoted by $\{v_t | t = 1, \dots, T\}$ where T is the number of voxels. The set of feature-vectors is input to the learning stage of the framework, in which a variation on the GMM is utilized to represent the intensity and spatial characteristics of the input image. Details of the probabilistic model and its parameter learning via the EM algorithm, are described in sections II-A and II-B. Segmentation of the brain image based on the learned model is described in section II-C. A model initialization scheme is described in section II-D.

A. The Probabilistic Model

The complex spatial layout of an MRI brain image provides a challenge for conventional GMM modeling schemes. In order to accommodate the spatial complexity, we model an image as a mixture of *many* Gaussians:

$$f(v_t | \Theta) = \sum_{i=1}^n \alpha_i f_i(v_t | \mu_i, \Sigma_i) \quad (1)$$

such that v_t is the feature vector associated with the t -th voxel, n is the number of components in the mixture model, μ_i and Σ_i are the mean and the covariance of the i -th Gaussian component f_i , and α_i is the i -th mixture coefficient. The spatial shape of the tissues is highly non-convex. However, since we use a mixture of many components, each Gaussian component models a small local region. Hence, the implicit convexity assumption induced by the Gaussian distribution is reasonable (and is empirically justified in the next section).

The high complexity of the spatial structure is an inherent part of the brain image. The intra variability of the intensity feature within a tissue (bias) is mainly due to artifacts of the MRI imaging process and once eliminated (via bias-correction schemes) is significantly less than the inter-variability among different tissues. It is therefore sufficient to model the intensity variability within a tissue by a single Gaussian (in the intensity feature). To incorporate this insight into the model, we further assume that each Gaussian is linked to a single tissue and all the Gaussians related to the same tissue share the same intensity parameters.

Technically, this linkage is defined via a grouping function. In addition to the GMM parameter set Θ , we define a grouping function $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, k\}$ from the set of Gaussians

¹Preliminary results of the study presented in this paper appeared in [24].

to the set of tissues. We assume that the number of tissues is known and the grouping function is learned in the initialization step. The intensity feature should be roughly uniform in the support region of each Gaussian component, thus, each Gaussian spatial and intensity features are assumed uncorrelated. The above assumptions impose the following structure on the mean and variance of the Gaussian components:

$$\mu_i = \begin{pmatrix} \mu_i^{XYZ} \\ \mu_{\pi(i)}^I \end{pmatrix}, \quad \Sigma_i = \begin{pmatrix} \Sigma_i^{XYZ} & 0 \\ 0 & \Sigma_{\pi(i)}^I \end{pmatrix} \quad (2)$$

where $\pi(i)$ is the tissue linked to the i -th Gaussian component and μ_j^I and Σ_j^I are the mean and variance parameters of all the Gaussian components that are linked to the j -th tissue.

The main advantage of the CGMM framework is the ability to combine, in a tractable way, a local description of the spatial layout of a tissue with a global description of the tissue's intensity. The multi-Gaussian spatial model makes our approach much more robust to noise than intensity-based methods. Note that no prior atlas information is used in the modeling process. Figure 2 illustrates the concept of combining local spatial modeling with global intensity modeling. A (2σ) spatial projection of the region-of-support for each Gaussian in the model is shown. Different shades of gray represent the three distinct tissues present.

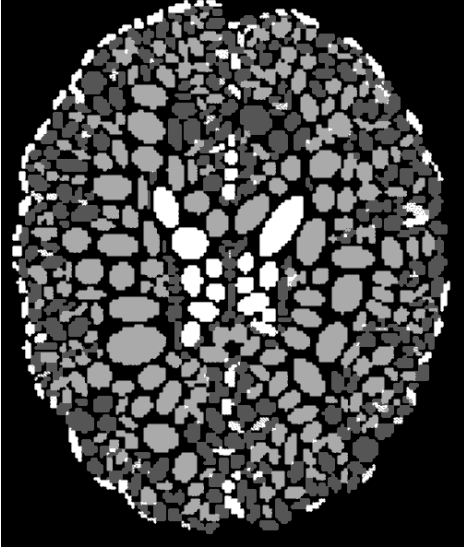


Fig. 2. Modeling each of the three tissues with multiple Gaussians. Gaussians with the same gray level share the same tissue label.

B. Learning the CGMM parameters

The EM algorithm [4] is utilized to learn the model parameters. In the proposed framework, Gaussians with the same tissue-label are constrained to have the same intensity parameters throughout. A modification of the standard EM algorithm for learning GMM is required, as shown in the following equations. The expectation step of the EM algorithm

for the CGMM model is (same as the unconstrained version):

$$w_{it} = p(i|v_t) = \frac{\alpha_i f_i(v_t|\mu_i, \Sigma_i)}{\sum_{l=1}^n \alpha_l f_l(v_t|\mu_l, \Sigma_l)} \quad (3)$$

$$i = 1, \dots, n \quad t = 1, \dots, T$$

We shall use the abbreviations:

$$n_i = \sum_{t=1}^T w_{it} \quad i = 1, \dots, n \quad (4)$$

$$k_j = \sum_{i \in \pi^{-1}(j)} n_i \quad j = 1, \dots, k$$

such that n_i is the expected number of voxels that are related to the i -th Gaussian component, $\pi^{-1}(j)$ is the set of all the Gaussian components related to the j -th tissue and k_j is the expected number of voxels in that tissue. The maximization in the M-step is done given the constraint on the intensity parameters:

$$\alpha_i = \frac{n_i}{n} \quad i = 1, \dots, n \quad (5)$$

$$\mu_i^{XYZ} = \frac{1}{n_i} \sum_{t=1}^T w_{it} v_t^{XYZ}$$

$$\Sigma_i^{XYZ} = \frac{1}{n_i} \sum_{t=1}^T w_{it} (v_t^{XYZ} - \mu_i^{XYZ}) (v_t^{XYZ} - \mu_i^{XYZ})^\top$$

$$\mu_j^I = \frac{1}{k_j} \sum_{i \in \pi^{-1}(j)} \sum_{t=1}^T w_{it} v_t^I \quad j = 1, \dots, k$$

$$\Sigma_j^I = \frac{1}{k_j} \sum_{i \in \pi^{-1}(j)} \sum_{t=1}^T w_{it} (v_t^I - \mu_j^I)^2$$

The grouping function π that links between the Gaussian components and the tissues is not altered by the EM iterations. Therefore, the affiliation of a Gaussian component to a tissue remains unchanged. However, since the learning is performed simultaneously on all the tissues, voxels can move between tissues during the iterations. Figure 3 shows the 2σ initial region support (a) versus the final region support (b) of the Gaussians, on the same synthetic image, after seven EM iterations.

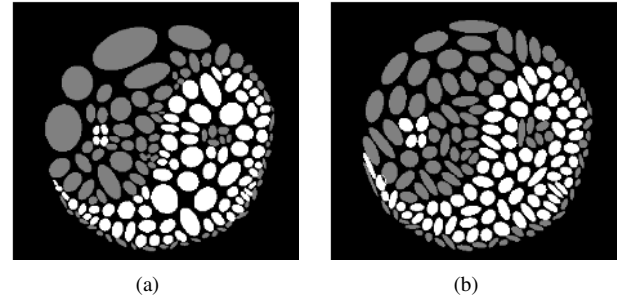


Fig. 3. Initial (a) and final (b) region support for all the Gaussians after seven EM iterations. Different shades of gray present different tissue labels.

C. Image Segmentation

A direct transition is possible between the estimated image representation and probabilistic segmentation. If the representation phase is a transition from voxels to clusters (Gaussians) in feature space, the segmentation process can be thought of as forming a linkage back from the feature space to the raw input domain. Each voxel is linked to the most probable Gaussian cluster, i.e. to the component of the model that maximizes the a-posteriori probability. In GMM that are based on intensity only and in which a single Gaussian models a particular class, the selection of the Gaussian with the highest posterior probability for each voxel immediately produces the required segmentation into one of the different tissues. The CGMM model uses multiple Gaussians for each tissue. Thus we need to sum over the posterior probabilities of all the identical tissue Gaussians to get the posterior probability of each voxel to originate from a specific tissue. Bayes rule implies that the posterior probability for a pixel to be linked to the j -th tissue is:

$$p(\text{tissue } j | v_t) = \frac{\sum_{i \in \pi^{-1}(j)} \alpha_i f_i(v_t | \mu_i, \Sigma_i)}{\sum_{i=1}^n \alpha_i f_i(v_t | \mu_i, \Sigma_i)} \quad (6)$$

$$j = 1, \dots, k, \quad t = 1, \dots, T$$

Hence the most likely tissue for the t -th voxel is:

$$\text{tissue-label}_t = \arg \max_{j \in \{1, \dots, k\}} p(\text{tissue } j | v_t) \quad t = 1, \dots, T \quad (7)$$

such that $\text{tissue-label}_t \in \{1, \dots, k\}$ is one of the tissues. The linkage of each voxel to a tissue label provides the final segmentation map.

D. Model Initialization

The EM algorithm requires appropriate initialization, as it is notoriously known to get stuck on local maxima of the likelihood function. In the multi-component CGMM model the problem is even more severe. A novel semi-supervised top-down initialization method is used in the current work, utilizing a-priori knowledge about the number of tissues of interest. The initialization is a dual step procedure, performed on a down-sampled version of the brain volume in order to decrease processing time. In the first step, intensity-based K-means clustering extracts a rough segmentation into six tissue classes: WM, GM, CSF, as well as three additional classes which take care of partial-volume effects as well as any other-than brain information that may be present in the image (see also [14]). It is assumed that the CSF class is of lowest intensity. In most cases, the brain image is cleaned sufficiently in the pre-processing stage and the WM class is the class with highest intensity. Cases exist in which fat residues of bright values, remain in the image. For this reason, the brightest class as well as its consecutive (less bright) classes are considered to belong to the WM category, until the number of WM voxels, in all the selected classes, is larger than a user defined threshold (MinWMfrac). All the voxels that were not labelled as WM or CSF, are labelled as GM voxels. This rough but important initial clustering determines an initial group of voxels that belong to each tissue. An example for the

segmentation achieved on a 2D image is shown in Figure 4(a) where the two gray levels represent the two classes present. The segmentation is used as a masking procedure (Figure 4(b)) that makes it possible to work on each tissue separately (region of interest is indicated in white).

The next step is a top-down procedure which propagates the tissue class information into the spatial space. Regions are split iteratively until convex regions are found that are suitable for Gaussian modeling. Each iteration of the spatial region splitting algorithm involves the following steps:

- A 3D connected-components (CC) algorithm is used to define distinct regions which should be modeled by one or more Gaussians. Figure 4(c) shows the CC regions in shades of gray.
- Each CC region is encircled with the smallest ellipsoid possible². If the volume inside the ellipsoid, that is *not* part of the region, is higher than a user defined threshold (“OutliersThresh”) and the ellipsoid volume supports more voxels than the defined threshold value of minimum voxels per region (“MinPixThresh”), the region is marked for further splitting. This process is illustrated in Figure 4(d), where the largest CC region is encircled with a 2D ellipsoid.
- A marked region is further split using K-means on the *spatial features* only, into 2-10 distinct (not necessarily connected) subregions, depending on the size of the ellipsoid and the user defined threshold for the minimum size of ellipsoid. Figure 4(f) illustrates splitting of two distinct regions.

The splitting algorithm iteratively proceeds as long as at least one region is marked for partitioning. Once the regions are determined, each region is modeled with a single Gaussian. The Gaussian’s spatial parameters (mean & variance) are estimated using the spatial features of the voxels supported by the region inscribed in the ellipsoid, while the intensity parameter is estimated using all the voxels supported by all the regions of the same tissue. Thus, Gaussians of the same tissue receive the same initial intensity parameter. Furthermore, each Gaussian is tagged with a label that indicates its tissue affiliation. Overall, the initialization process determines and fixes the grouping function π .

Figure 4(a)-(g) illustrates the steps of the initialization process as applied to a 2D synthetic image with two tissues. Figure 4(g) shows the final 2σ spatial projection of all the Gaussians at the end of the initialization process, where different shades of gray present different tissue labels.

III. EXPERIMENTAL VALIDATION

In the following we present a set of experiments to validate the proposed framework. Both simulated data as well as real brain data are used. The simulated data was taken from the BrainWEB simulator repository³. Real normal T1 MR brain data sets were taken from the Center for Morphometrics Analysis at the Massachusetts general Hospital repository⁴

²using Matlab’s® Geometric Bounding Toolbox (GBT®)

³<http://www.bic.mni.mcgill.ca/brainweb>

⁴<http://www.cma.mgh.harvard.edu/ibsr>

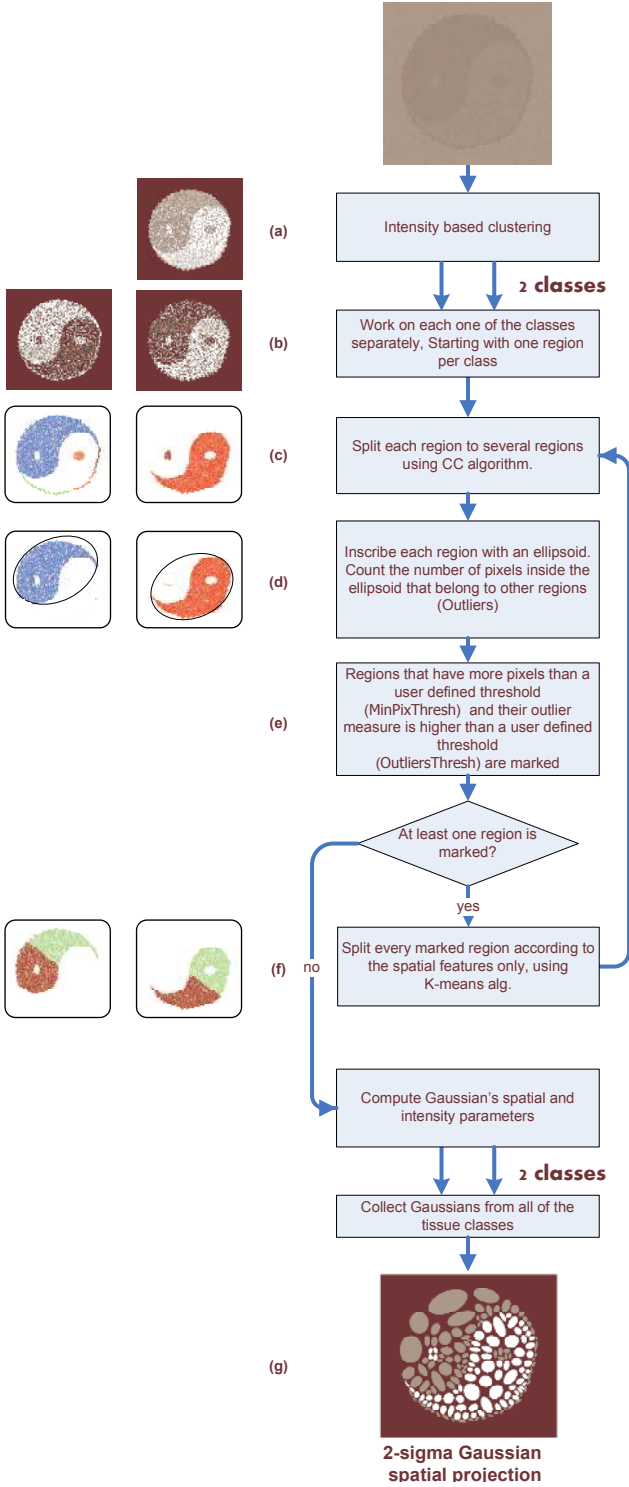


Fig. 4. Illustration of the initialization algorithm on 2D synthetic image with two tissues.

(hereon termed IBSR). The technical data pertaining to the presented experiments is shown in Table I.

In order to pre-process the data, as well as a means for comparison of the CGMM algorithm results, the Expectation-Maximization Segmentation (EMS) software⁵ package is used [10]. The EMS implements a fully automated model-based segmentation of MR images of the brain based on the methods described in [12], [13], [17]. The software package is an add-on to the Statistical Parametric Mapping (SPM99) package⁶ developed by the Wellcome Department of Cognitive Neurology, Institute of Neurology, University College London. All brain volumes were pre-processed to extract the brain region as well as to correct any interslice intensity inhomogeneity and within slice bias, where the bias correction is based on fourth-degree polynomial fitting as part of an EM algorithm [13]. Note that the presented algorithm differentiates between three tissues, with no intensity models used as prior information. In this unsupervised mode, no intensity scale normalization or standardization is needed.

In order to compare with state-of-the-art algorithms, in particular, the well-known EM-based segmentation algorithm of Van-Leemput (hereon termed KVL) [13], the volume data is pre-registered to the SPM statistical brain model using the `register` function of the EMS which implements an affine registration algorithm based on the maximum mutual information criterion described in [17].

A. Simulated Brain

The first experiment was performed on T1-weighted simulated data from the BrainWEB with 1%-9% noise levels and no spatial inhomogeneity. Two widely-used volumetric overlap metrics are used. The first, described by Dice [5] and recently used by Zijdenbos [28], is used to quantify the overlap between the automatic segmentation and the ground truth for each tissue. Denote by V_{ae}^k the number of voxels that are assigned to tissue k by both the ground truth and the automated algorithm. Similarly, let V_a^k and V_e^k denote the number of voxels assigned to tissue k by the algorithm and the ground truth, respectively. The overlap between the algorithm and the ground truth for tissue k is measured as:

$$M_1^k = \frac{2V_{ae}^k}{(V_a^k + V_e^k)} \quad (8)$$

The Dice metric attains the value of one if both segmentations are in full agreement and zero if there is no overlap at all. The second metric used is known as the Tanimoto coefficient [6] which is defined as:

$$M_2^k = \frac{V_{a \cap e}^k}{V_{a \cup e}^k} \quad (9)$$

where $V_{a \cap e}^k$ denotes the number of voxels classified as class k by both the proposed algorithm and the manual segmentation (our ground truth), and $V_{a \cup e}^k$ denotes the number of voxels classified as class k by either the proposed algorithm or the

⁵freely available from <http://bilbo.esat.kuleuven.ac.be/mic-pages/EMS/>

⁶freely available from <http://www.fil.ion.ucl.ac.uk/spm/>

TABLE I
TECHNICAL DATA FOR THE MRI VOLUMES USED IN THE EXPERIMENTS.

Experiment	Synthetic Data (BrainWEB simulator)	Real Data (IBSR web site)		
	Section III-A	Section III-C.1		Section III-C.2
Pulse Sequence	T1			
Number Of Volumes	1	9	9	18
Scan Technique	Spoiled Flash	Spoiled Gradient		Spoiled Flash
Rep. Time TR (ms)	18	40	50	unknown
Echo Time TE (ms)	10	8	9	unknown
Flip Angle (deg.)	30	50	50	unknown
Slice Thickness (mm)	1	3.1	3	1.5

the expert. It should be noted that although both M_1 and M_2 range from zero to one, for a given automatic classification and ground truth, one always has that $M_2^k \leq M_1^k$.

Figure 5 shows the segmentation results of the CGMM algorithm as compared to the KVL algorithm, for the two measures. The KVL algorithm was activated with: (1) No estimation of bias inhomogeneity (2) Markovian Random Field (MRF) model enabled. The CGMM algorithm was implemented with the following set of parameters: MinPixThresh=0.002%, OutliersThresh=0.007% and MinWMfrac=5%. These parameters were empirically found to give a good compromise between execution time and performance. They hereon serve as the default set of parameters.

It can be observed that the CGMM method outperforms the KVL algorithm in the presence of high noise levels and provides comparable results in the presence of very low noise levels. The performance of both algorithms is worse with 1% noise than with 9% noise. This fact was observed and explained in the original work by Leemput [13]. Note that the CGMM uses no prior spatial information, whereas the KVL algorithm depends on precise prior information for its initialization and parameter learning.

The strength of the CGMM framework is more clearly evident in its robustness to noise and smooth segmentation results in increased noise levels. Figure 6 demonstrates the segmentation on a single slice (95) from the BrainWEB simulator, across varying noise levels. We note that performance of the CGMM framework stays more constant with the increase in noise, as compared to the KVL algorithm. In the segmented image it is evident that smoother regions are obtained by the CGMM algorithm. Additional segmentation results of two very noisy brain simulator slices are presented in Figure 7.

B. Further Investigation of CGMM Performance

In this section further analysis of the CGMM framework is presented: The initialization process is illustrated and a sensitivity analysis on the initialization parameters is conducted. Finally, the effect of the intensity constraint is demonstrated. Figure 8 presents a schematic representation of the splitting process which is part of the CGMM initialization. A 2D projection of the 3D ellipsoids is shown at different stages of the initialization process. Figure 8(a) shows the results after the first step of the initialization procedure. GM and WM are

connected regions and each get a single Gaussian, while the CSF has several distinct regions in the image and thus several Gaussians are allocated instead of one. The splitting process is evident already in this first stage. It can be observed that the spatial overlap between ellipsoids from different tissues is consistently decreased following each iteration.

1) *Parameter Sensitivity Analysis:* Three parameters take part in the CGMM initialization process as follows (see section II-D):

- **MinWMfrac:** Part of the intensity-based K-means clustering. Defines the minimal fraction of the White Matter in the total brain volume. A higher value means that more clusters (out of the six initial clusters provided by the K-means) will be labelled as WM. This parameter is useful for cases in which spots of bright, non brain tissues, are apparent as a result of imperfect skull removal. When no bright residues are apparent the brightest class should contain most of the WM tissue, thus any positive threshold value suffices. Therefore a low threshold value of 5% is used for the regular cases. In cases with residue, a larger value of 20% is used (this number is based on empirical studies on ground-truth data of several healthy brain volumes that show the WM tissue takes at least 35% of the brain volume). In this case the algorithm defines all classes as WM until the class voxels contain at least 20% of the voxels.
- **MinPixThresh:** Part of the top-down splitting process. Defines the minimum fraction of the total brain volume inscribed in an ellipsoid. The parameter takes the values in $[0, 1]$. A higher value means that each Gaussian's support region will contain more voxels out of the total brain volume, thus, the splitting process will be shorter and the model will contain less Gaussians. This leads to a coarser segmentation but guarantees there are enough voxels in each ellipsoid to provide sufficient statistics to calculate the spatial parameters of the Gaussians. A low value ensures a higher-resolution in the segmentation and the proper modeling of thin structures, such as the CSF. Note that regions that are smaller than the defined parameter value will be omitted and defined as noise.
- **OutliersThresh:** Part of the top-down splitting process. Defines the maximum fraction of the total brain volume which does not belong to the region yet is inscribed

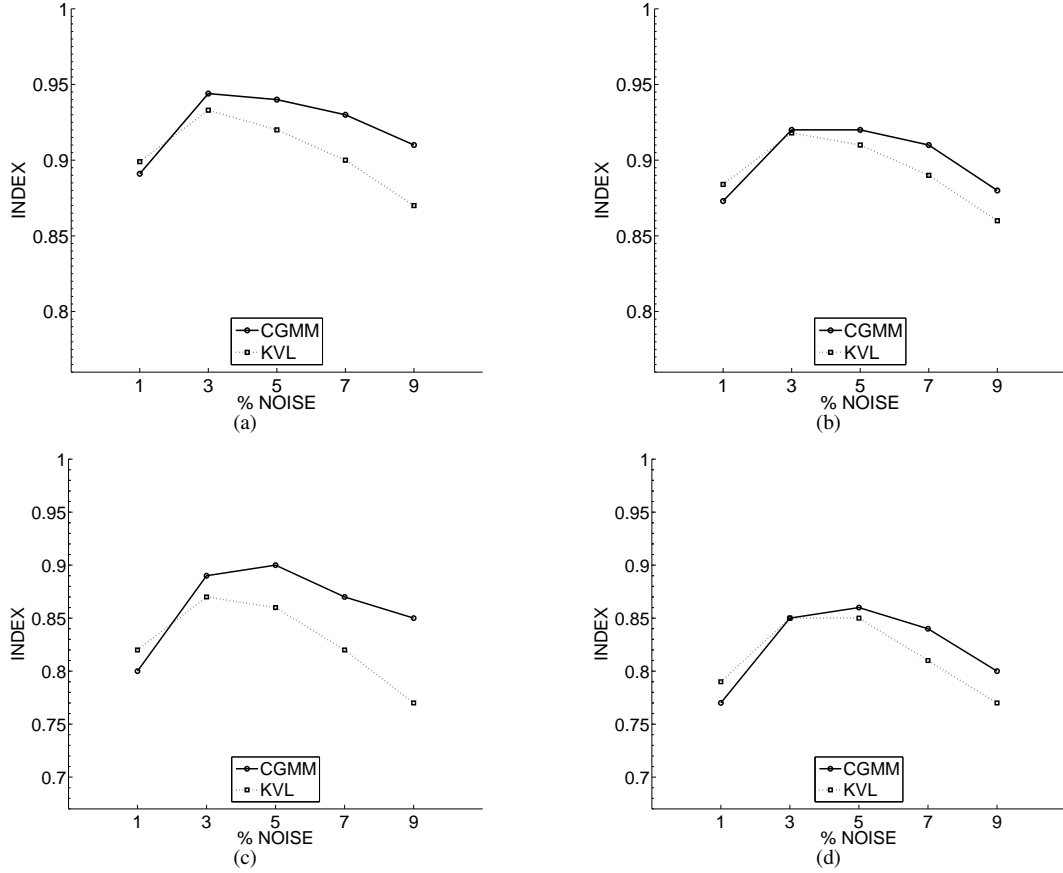


Fig. 5. Segmentation of BrainWEB MRI simulated data [2] with 0% spatial inhomogeneity and 9% noise, using the CGMM algorithm and the KVL algorithm. Top row- Dice performance index: (a) WM; (b) GM. Bottom row- Tanimoto performance index: (c) WM; (d) GM.

in the ellipsoid. The parameter takes the values in $[0, 1]$. A higher value means that each Gaussian would contain more outlier voxels in its support region. Therefore, the splitting process will be shorter and the model will contain less Gaussians. This leads to coarser segmentation results but prevents over-segmentation in presence of high thermal noise levels.

The sensitivity of the segmentation framework with-respect-to the parameter setting is explored next. A set of default values has been experimentally selected, as defined in Section III-A and listed in Table II. Segmentation performance is measured across several cases, in each scenario one of the parameters is changed in value while the others are kept fixed. A similar approach can be found in [18]. Performance is measured in a similar experimental setting to the one described in Section III-A, with a “worst case” simulated volume (9% noise) and the Dice overlap metric.

The effect of parameter variation is shown in Table II. The first row presents a default set of initialization parameters. These were empirically found to provide good results in a reasonable processing time on most of the brain volumes tested. The next rows exhibit gradual modification of a single parameter at a time. Minor monotonic performance deterioration of about 1% for the GM and 2% for the WM tissue is noticeable when increasing the `OutliersThresh` threshold from 0.005% to 0.030%. At the same time, the number

of Gaussians in the model decreases. A similar phenomena is present with regard to the `MinPixThresh` parameter. Increasing the threshold level from 0.004% to 0.030% presents a deterioration of about 3% for the GM and 2% for the WM tissue. The `MinWMfrac` parameter seems to have almost no influence on the results. This is due to the relatively clean data set of the BrainWEB simulator, which does not contain bright voxels residuals. The number of Gaussians in the model is noticeably correlated with the measured performance. As we can see from Table II there is not much sensitivity of the performance to the variation in the initialization parameters.

2) Intensity Constraint Effectiveness: A key characteristic of the CGMM framework is the global intensity constraint for Gaussians with the same tissue label. To demonstrate the effect of incorporating constraints within the modeling process, we compare the CGMM performance with regular EM-based modeling (with no intensity constraints). Figure 9 shows a comparison between the two approaches. The CGMM algorithm was initialized as described in Section II-D. The Gaussian parameters were once learned using EM-based modeling (without intensity constraints) and once learned using the CGMM framework (with intensity constraints). Segmentation using regular EM-based modeling, without posing any constraints on the Gaussian parameters, is shown in Figure 9(b). Corresponding segmentation results of the CGMM framework are shown in Figure 9(c). With no intensity constraints, the

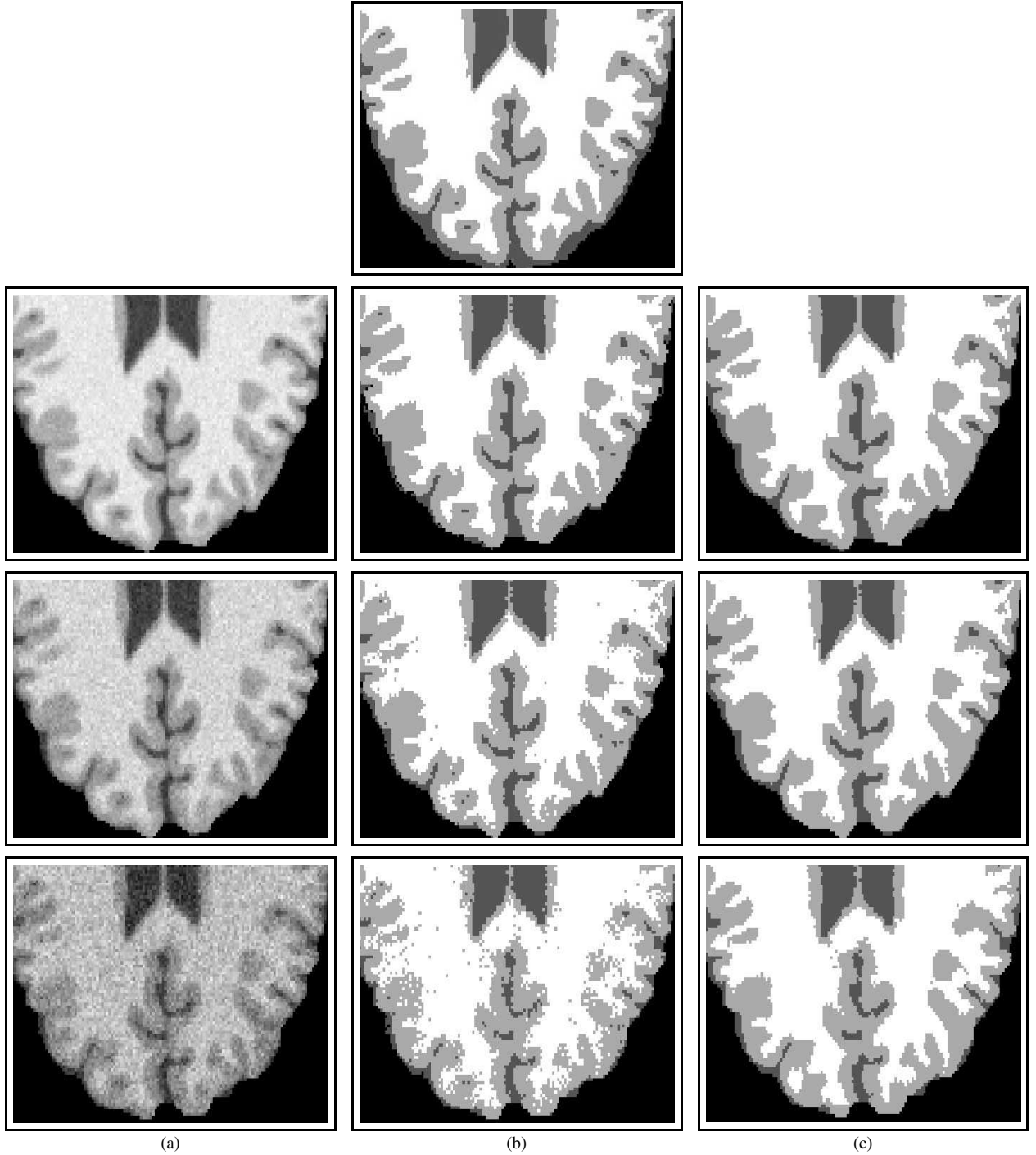


Fig. 6. Comparison of CGMM vs. KVL algorithm for segmentation of slice 95 from BrainWEB simulator with different noise levels. Upper image: Ground-Truth. Upper row: 3% noise, Middle row: 5% noise, Lower row: 9% noise. Columns: (a) Original image (b) KVL algorithm (c) CGMM algorithm.

segmentation result is noticeably distorted due to the high noise present. A considerable improvement is apparent when constraining the intensity parameters of the Gaussians from the same tissue.

C. Real Brain Volumes

The CGMM framework performance on real data sets is demonstrated next. A set of thirty-six normal T1-weighted real brain data sets was downloaded from the IBSR repository [3].

Two different data sets, each including 18 full-brain volumes, are used. The parameters of the sets, as copied from the IBSR web site, are summarized in Table I. Each of the brain volumes in the IBSR site is provided with manual segmentation by an expert technician.

1) *Real brain - Experiment I:* The first experiment compares the CGMM algorithm with six different segmentation algorithms that are provided as part of the IBSR website ⁷, as

⁷Detailed information on these algorithms is not available in the IBSR site

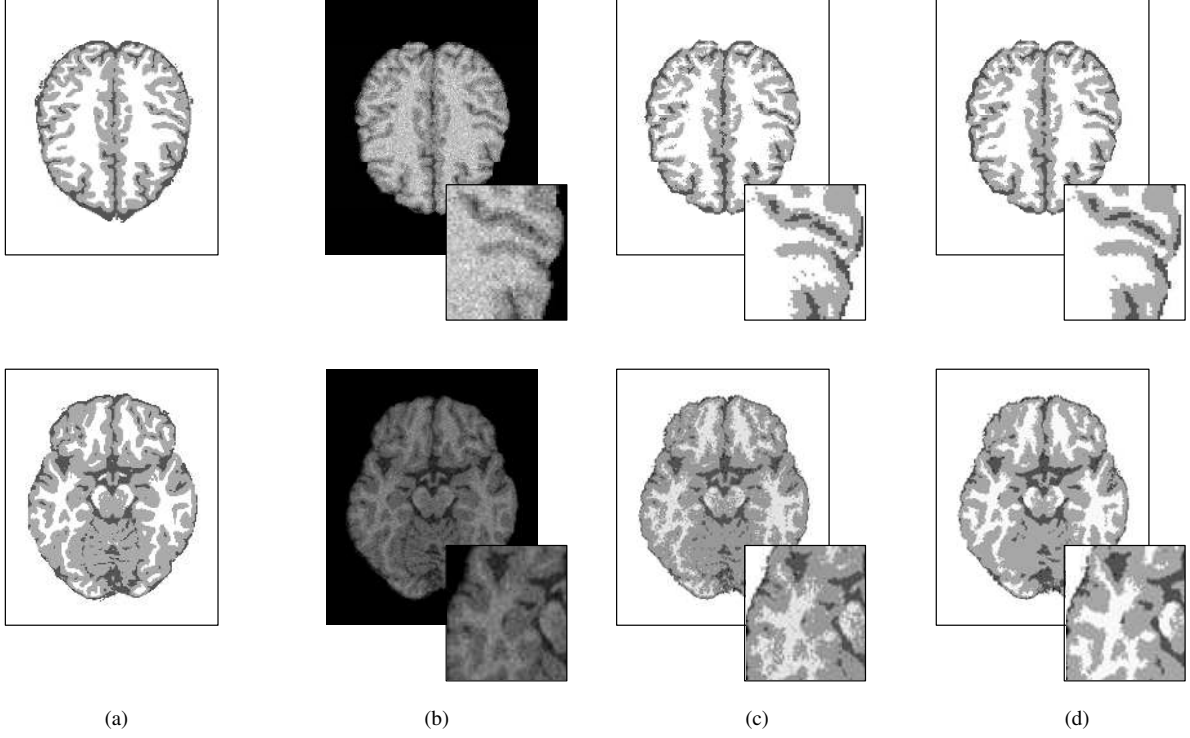


Fig. 7. Segmentation results of two sample slices from the BrainWeb volume with 9% thermal noise and 0% intensity bias. Slice number 115 is shown in first row and slice number 57 is shown in second row. Columns: (a) Ground-Truth (b) Original slice (c) KVL's algorithm (d) CGMM algorithm.

TABLE II

SENSITIVITY OF THE CGMM ALGORITHM PERFORMANCE TO VARIATIONS IN INITIALIZATION PARAMETERS MEASURED ON SIMULATED MR BRAIN SCAN WITH 9% NOISE, ACCORDING TO THE DICE PERFORMANCE METRIC.

	GRAY	WHITE	CSF	Number of Gaussians in the model
Default values: OutliersThresh=0.007% MinPixThresh=0.002% MinWMfrac=5%	0.886	0.918	0.735	5067
OutliersThresh=0.005%	0.889	0.920	0.742	6191
OutliersThresh=0.009%	0.885	0.916	0.735	4305
OutliersThresh=0.011%	0.884	0.914	0.734	3821
OutliersThresh=0.015%	0.882	0.912	0.732	3157
OutliersThresh=0.030%	0.876	0.905	0.723	2236
MinPixThresh=0.001%	0.892	0.918	0.757	4922
MinPixThresh=0.004%	0.880	0.918	0.708	4922
MinPixThresh=0.006%	0.878	0.916	0.701	4649
MinPixThresh=0.008%	0.874	0.914	0.690	4088
MinPixThresh=0.015%	0.860	0.903	0.655	2295
MinPixThresh=0.030%	0.848	0.889	0.634	1140
MinWMfrac=1%	0.886	0.918	0.735	5067
MinWMfrac=10%	0.886	0.918	0.735	5067
MinWMfrac=15%	0.870	0.904	0.741	5046
MinWMfrac=20%	0.870	0.904	0.741	5046

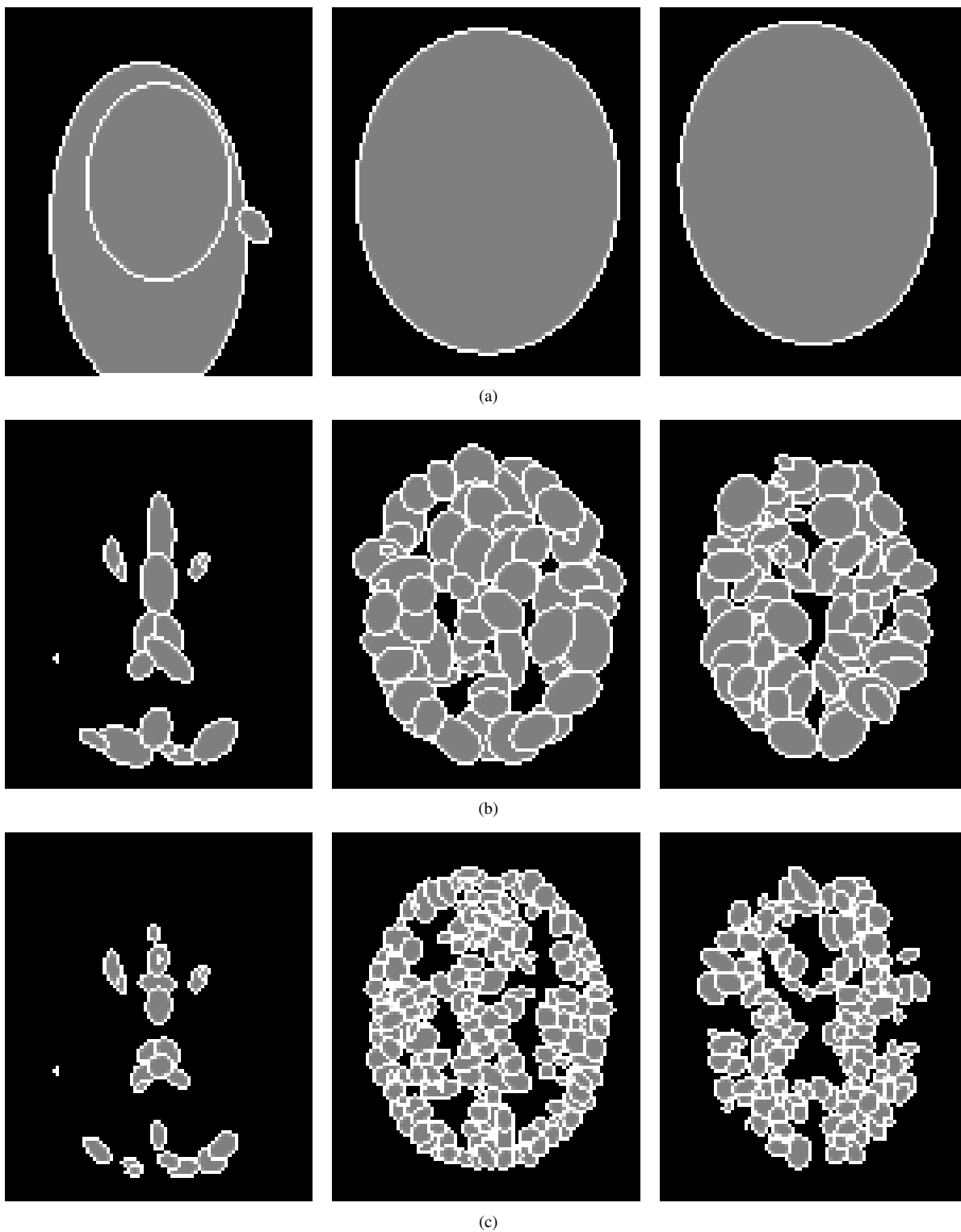


Fig. 8. Splitting each tissue during the initialization process of the CGMM algorithm on the BrainWEB simulated data set (slice number 85) with 9% noise and no intensity inhomogeneity. Left: CSF; Center: GM; Right: WM. (a) first iteration (b) fourth iteration (c) eighth iteration.

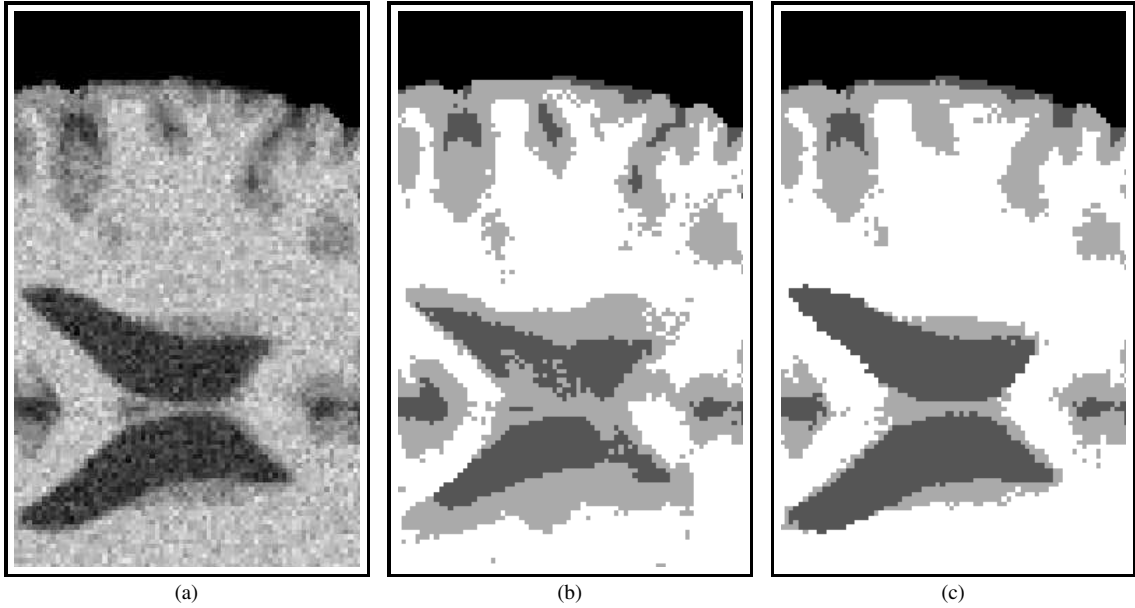


Fig. 9. (a) Slice number 95 from the BrainWEB normal brain simulator; (b) GMM (without constraints); (c) CGMM (with constraints).

well as with a state-of-the-art segmentation algorithm, which we term the *Marroquin algorithm* [18]. Marroquin’s algorithm presents a fully automatic Bayesian segmentation algorithm. The algorithm incorporates robust non-rigid registration of a brain atlas to the specimen to be segmented, as a means for better initialization and to set prior probabilities for each class. A variant of the EM algorithm is then used to find the optimal segmentation given the intensity models along with a spatial coherence assumption in the form of a MRF model.

The experiment is conducted on a data set of 18 normal T1-weighted real MR brain data (see Table I). This data set is becoming a standard for comparative algorithm testing. It has been used in a variety of volumetric studies in the literature as it contains varying levels of difficulty, with the worst scans consisting of low contrast and relatively large spatial inhomogeneities.

The overlap metric used by the IBSR repository is the Tanimoto coefficient [6]. Figure 10 shows the comparison results. Note that a higher value indicates a better correspondence to the ground-truth (expert’s segmentation). The results summarized in Table III indicate that the CGMM algorithm outperforms the set of algorithms provided in the IBSR site for both GM and WM, and performs comparably to the Marroquin algorithm. The CGMM algorithm has the lowest standard deviation for both GM and WM tissues.

2) *Real brain - Experiment II:* In a second experiment, a different set of 18 T1-weighted, normal real MR brain data from the IBSR repository was used (see Table I). This set of data was used to compare the CGMM algorithm with the KVL algorithm, on real brain data.

The CGMM algorithm was applied with the default set of parameters except for the MinWMfrac value which was increased to 20% in order to compensate for bright spots apparent in some of the MR brain scans. Figure 11 shows a comparison of the CGMM method with the KVL algorithm using the Dice metric (Equation 8). Results are summarized

in Table IV and indicate comparable performance between the KVL algorithm and the CGMM approach on this set of MR brain scans. A gain in performance seems to be evident in the CSF tissue. Poor segmentation results for both algorithms can be seen in brain volumes IBSR8 and IBSR12. The difficulty in these two cases was found to be the result of imperfect brain isolation in the preprocessing stage, which translates to an error in the performance measure due to differing volumes.

The Dice overlap metric is widely used in the literature (e.g [18], [13]) but it is by no means claimed to be the most suitable metric for comparing performance of different segmentation algorithms. In particular, the Dice metric depends on the size and the shape complexity of a given object and is related to the image sampling. Assuming that most of the errors occur at an object boundaries, small objects are penalized and get a much lower score than large objects. No consensus currently exists regarding a necessary and sufficient set of measures for segmentation performance characterization. In a recently published work, the *Mean absolute surface distance* metric (which is based on the Hausdorff distance metric [1]) is suggested to compare different MR brain segmentation results [8]. Figure 12 shows a comparison between the CGMM method and KVL’s algorithm, based on the mean absolute surface distance metric as implemented by the VALMET software package⁸ [8]. The results are summarized in Table V. Using this metric, a lower value means better segmentation. As seen, the standard deviation of both algorithms is considerably higher than the differences in their average values. Thus, we can only say that the two algorithms show a comparable performance according to the mean absolute surface distance metric.

⁸freely available at <http://www.ia.unc.edu/dev/download/valmet/index.htm>

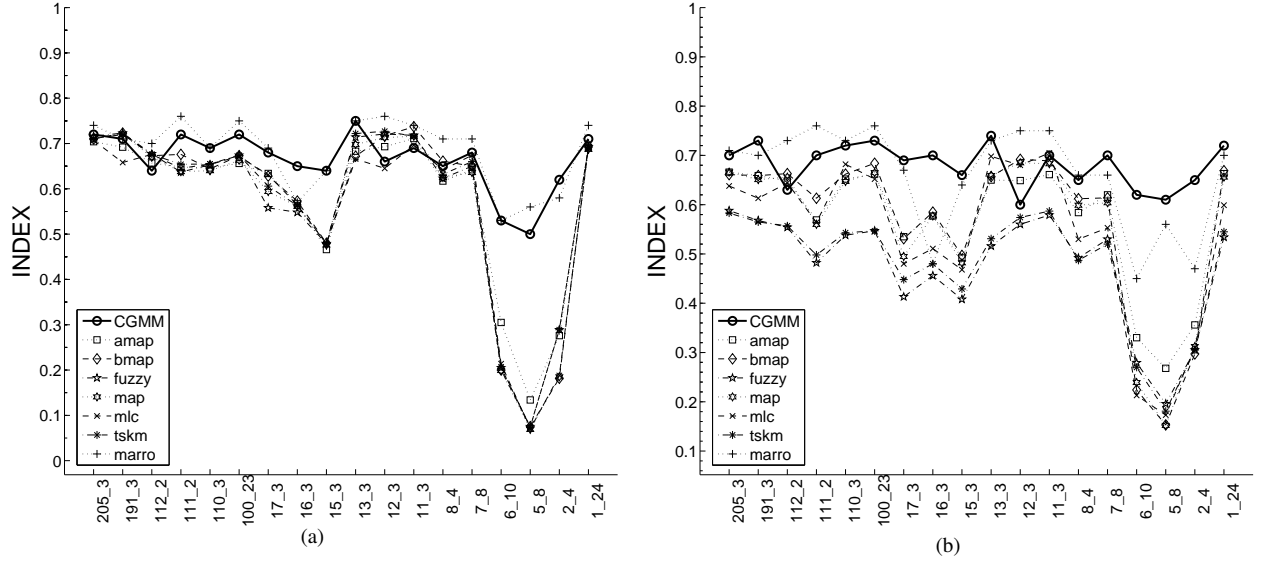


Fig. 10. Tanimoto's performance metric for different segmentation algorithms on different real brain MR scans from the IBSR repository. The bold line corresponds to our algorithm. (a) WM; (b) GM.

TABLE III

MEAN AND STANDARD DEVIATION OF TANIMOTO'S OVERLAP METRIC FOR VARIOUS SEGMENTATION METHODS, CALCULATED OVER A FIRST SET OF 18 BRAIN SCANS FROM THE IBSR REPOSITORY.

Method (abbr)	Source	GM		WM	
		Mean	Std. Dev.	Mean	Std. Dev.
adaptive MAP (amap)	IBSR	0.57	0.13	0.58	0.17
biased MAP (bmap)	"	0.56	0.17	0.58	0.21
fuzzy c-means (fuzzy)	"	0.47	0.11	0.58	0.19
Maximum a Posteriori Probability (map)	"	0.55	0.16	0.57	0.20
tree-structure k-means (tskm)	"	0.48	0.12	0.58	0.19
Maximum-Likelihood (mlc)	"	0.54	0.16	0.57	0.20
Marroquin (marro)	[18]	0.66	0.10	0.68	0.10
Constrained GMM (CGMM)		0.68	0.04	0.66	0.06

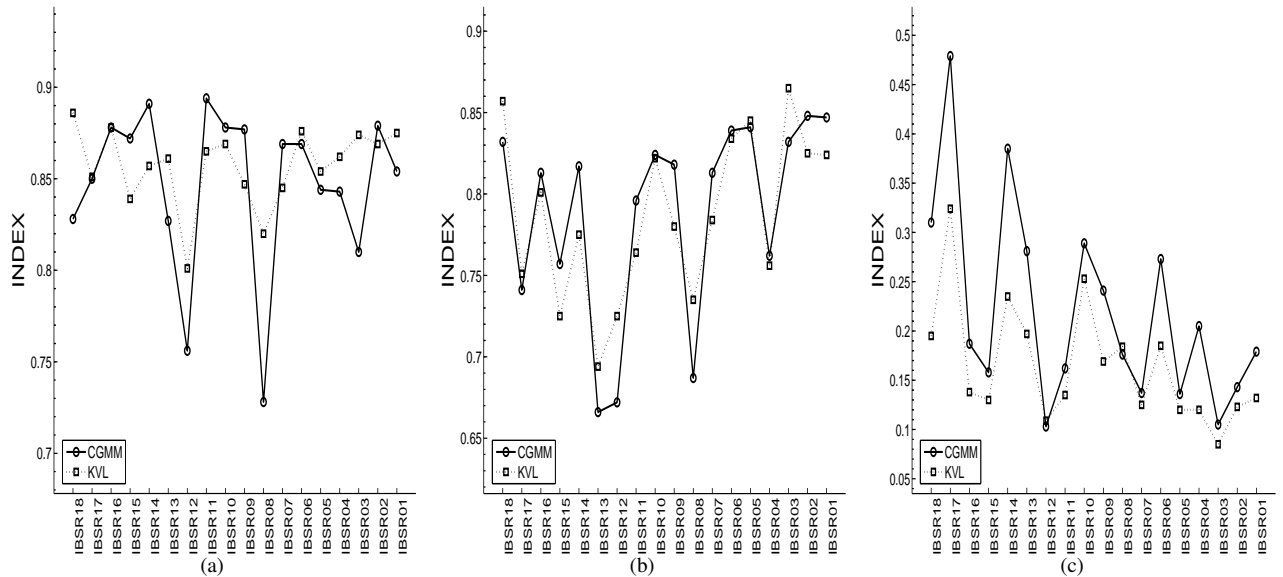


Fig. 11. Dice overlap metric of the CGMM (solid) and the KVL algorithm (dashed) per each of the real MR brain data sets from the IBSR repository. (a) WM; (b) GM; (c) CSF.

TABLE IV

MEAN AND STANDARD DEVIATION OF DICE OVERLAP METRIC RESULTS FOR CGMM AND KVL ALGORITHMS CALCULATED OVER A SECOND SET OF 18 REAL BRAIN SCANS FROM THE IBSR REPOSITORY.

Method	GM		WM		CSF	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
KVL	0.786	0.050	0.857	0.021	0.164	0.060
CGMM	0.789	0.060	0.847	0.044	0.219	0.100

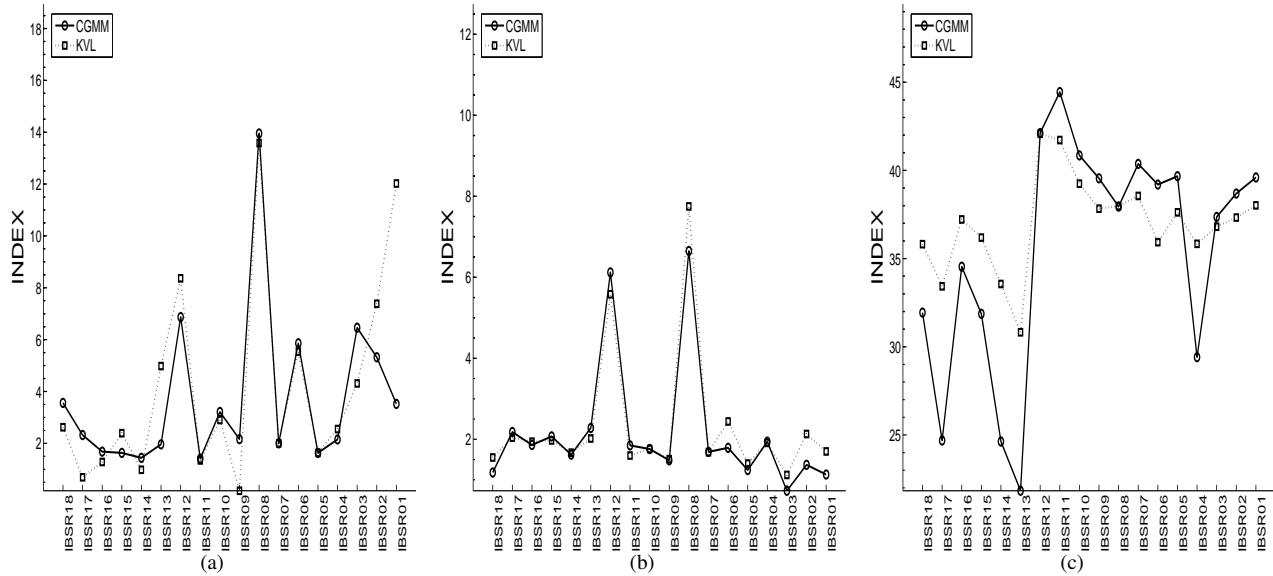


Fig. 12. Mean Absolute Surface Distance metric of the CGMM (solid) and the KVL algorithm (dashed) per each of the real MR brain data sets from the IBSR repository. (a) GW; (b) WM; (c) CSF.

TABLE V

MEAN AND STANDARD DEVIATION OF THE ABSOLUTE SURFACE DISTANCE METRIC RESULTS FOR CGMM AND KVL ALGORITHMS CALCULATED OVER A SECOND SET OF 18 REAL BRAIN SCANS FROM THE IBSR REPOSITORY.

Method	GM		WM		CSF	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
KVL	2.32	1.65	4.15	3.88	37.00	2.72
CGMM	2.16	1.59	3.73	3.12	35.48	6.62

IV. CONCLUSION

We present a fully automated, parametric, unsupervised algorithm for tissue classification of MR images of the brain that is also very noise resistant. The CGMM framework provides an alternative to MRF modeling. Spatial information is incorporated into the learned model along with the intensity feature. The addition of the spatial features within the GMM model provides for a more localized analysis. An important characteristic of the proposed model is the global intensity parameter-tying across Gaussians, per tissue type. The global intensity constraint ensures more global intensity learning and smoother segmentation results. A key feature of the CGMM framework is the fact that the analysis is done adaptively on the data at-hand. No a-priori knowledge is used apart from general properties about the intensity order of the tissues in T1 MR. This can support cases in which the use of an atlas is not appropriate and can eliminate the need for registration of the data to an atlas as part of the overall analysis procedure.

The algorithm was tested on different slices from a very noisy simulated brain volume. Quantitative comparison with KVL's state-of-the-art algorithm in the presence of extremely noised images was performed and showed that the CGMM algorithm presents superior results in the presence of moderate to high noise, both visually and quantitatively. This result holds with increased noise of up to 9%, a level that presents a challenge for most existing segmentation algorithms. Further investigation was conducted on a set of 36 real brain volumes from the IBSR repository. Eighteen brain volumes were used to compare the CGMM algorithm with the KVL algorithm; A second set of eighteen volumes was used for a comparison with Marroquin's algorithm as well as with six additional classic algorithms whose performance information was taken from the IBSR web site. Three different coefficients for similarity were used: the Dice metric (equation 8), the Tanimoto metric (equation 9) and the mean absolute surface distance metric implemented by the Valmet [8] software package. The

real brain volumes contain moderate thermal noise effects. Results show a significantly better performance of the CGMM framework as compared with the six segmentation algorithms provided as results in the IBSR site (Figure 11). The CGMM framework performed comparably to the Marroquin algorithm (Figure 11) as well as to the KVL algorithm (Figure 12) on the real volumes.

The algorithm stability and consistency over a wide range of initial values was demonstrated. The processing time of the algorithm depends on the number of Gaussians which are used to model the brain data set as well as on the number of voxels in the image. The number of Gaussians is determined in the initialization stage, as a by-product of the initialization parameters, noise level and tissue's complexity. Using the default initialization parameters on the first set of 18 real brain volumes, the average processing time was 7 minutes for a single brain volume, executed on a single 3.0GHz pentium4 processor of a PC machine, with 1Gbyte memory.

The CGMM framework combines local spatial information with global intensity modeling. The grouping of Gaussians has a resemblance to using a Markov Random Field model. The main difference is that the intensity information is linked adaptively and globally within the image, in contrast to a MRF model that integrates information from the nearest neighbors only and in predetermined neighborhoods. These differences may be the reason for the improved segmentation results with decreased tissue region granularity in the presence of extreme noise.

Eliminating the need for an anatomical atlas provides the opportunity to apply the CGMM framework to diseased brains and neonatal brains as well as to additional organs (abdomen, chest, etc). Furthermore, the algorithm can be applied to a subset of the full volume or even to 2D images. Currently we are working on an extension of the model to incorporate intensity inhomogeneities as well as to support multi-channel data.

REFERENCES

- [1] N. Aspert, D. Santa-Cruz, and T. Ebrahimi. Mesh: measuring errors between surfaces using the Hausdorff distance. In *ICME 2208:1*, pages 705–708. IEEE, 2002.
- [2] C. A. Cocosco, V. Kollokian, R. K. Kwan, and A. C. Evans. Brainweb: online interface to a 3D MRI simulated brain database. *Neuroimage*, 5(4), 1997.
- [3] Ctr. Morphometric Anal. [Online]. IBSR Internet brain segmentation repository. tech. rep. <http://neuro-www.mgh.harvard.edu/cma/ibsr/>, 2000.
- [4] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, (39):1–38, 1977.
- [5] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [6] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley Interscience, 2001.
- [7] G. Dugas-Phocion, M. Ángel González Ballester, G. Malandain, C. Lebrun, and N. Ayache. Improved EM-based tissue segmentation and partial volume effect quantification in multi-sequence brain MRI. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 26–33, 2004.
- [8] G. Gerig, M. Jomier, and M. Chakos. Valmet: A new validation tool for assessing and improving 3D object segmentation. In *LNCS 2208*, pages 516–528. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2001.
- [9] K. Held, E. R. Kops, B. J. Krause, W. M. Wells III, R. Kikinis, and H. Müller-Gärtner. Markov random field segmentation of brain MR images. *IEEE Trans. Med. Imaging*, 16(6):878–886, 1997.
- [10] K. V. Leemput. Expectation-Maximization segmentation (EMS). <http://www.medicalimagecomputing.com/EMS/>, 2001.
- [11] T. Kapur, W. E. Grimson, W. M. Wells, and R. Kikinis. Segmentation of brain tissue from magnetic resonance images. *Med Image Anal.*, 1(2):109–27, 1996.
- [12] K. V. Leemput, F. Maes, D. Vandermeulen, A. C. F. Colchester, and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Trans. Med. Imaging*, 20(8):677–688, 2001.
- [13] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. imaging*, 18(10):897–908, 1999.
- [14] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens. A unifying framework for partial volume segmentation of brain MR images. *IEEE Trans. Med. Imaging*, 22(1):105–119, 2003.
- [15] S. Z. Li. *Markov random field modeling in computer vision*. Springer-Verlag, 1995.
- [16] A. Macovski. Noise in MRI. *Magn. Reson. Med.*, 36(3):494–7, 1996.
- [17] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging*, 16(2):187–198, 1997.
- [18] J. L. Marroquin, B. C. Vemuri, S. Botello, F. Calderon, and A. Fernandez-Bouzas. An accurate and efficient bayesian method for automatic segmentation of brain MRI. *IEEE Trans. Med. imaging*, 21(8):934–45, 2002.
- [19] T. McInerney and D. Terzopoulos. Medical image segmentation using topologically adaptable surfaces. In *Proceedings of the First Joint Conference on Computer Vision, Virtual Reality and Robotics in Medicine and Medical Robotics and Computer-Assisted Surgery*, pages 23–32. Springer-Verlag, 1997.
- [20] D. L. Pham. Spatial models for fuzzy clustering. *CVIU*, 84(2):285–297, 2001.
- [21] D. L. Pham, C. Xu, and J. L. Prince. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2:315–337, 2000.
- [22] M. Prastawa, J. Gilmore, W. Lin, and G. Gerig. Automatic segmentation of neonatal brain MRI. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 10–17, 2004.
- [23] T. Rohlfing and C. R. Maurer Jr. Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. *IEEE Transactions on Information Technology in Biomedicine*, 7(1):16–25, 2003.
- [24] A. Ruf, H. Greenspan, and J. Goldberger. Tissue classification of noisy MR brain images using constrained GMM. *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2005.
- [25] W. Wells, R. Kikinis, W. Grimson, and F. Jolesz. Adaptive segmentation of MRI data. *IEEE Trans. Med. imaging*, 15:429–42, 1996.
- [26] Z. Xue, D. Shen, and C. Davatzikos. Classic: Consistent longitudinal alignment and segmentation for serial image computing. In *IPMI*, volume 3565, pages 101–113, Jul 2005.
- [27] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden markov random field model and the Expectation-Maximization algorithm. *IEEE Trans. Med. imaging*, 20(1):45–57, 2001.
- [28] A. P. Zijdenbos and B. M. Dawant. Brain segmentation and white matter lesion detection in MR images. *Critical Reviews in Biomedical Engineering*, 22(5-6):401–65, 1994.