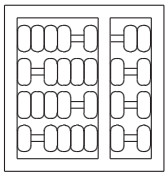


INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

MR-Image Segmentation of Brain Tissues based on Bias Correction and Optimum-Path Forest Clustering

F.A.M. Cappabianco *A.X. Falcão*
C.L. Yasuda *J.K. Udupa*

Technical Report - IC-10-07 - Relatório Técnico

March - 2010 - Março

MR-Image Segmentation of Brain Tissues based on Bias Correction and Optimum-Path Forest Clustering

Fábio Augusto Menocci Cappabianco* Alexandre Xavier Falcão†
Clarissa Lin Yasuda‡ Jayaram K. Udupa§

05/03/2010

Abstract

Magnetic resonance (MR) image segmentation of brain tissues has become crucial to advance research, diagnosis and treatment procedures in Neurology. Despite the large number of published papers, there is no standard method suitable for all cases. We present here a fast, accurate and robust to anatomical variations approach based on inhomogeneity/bias correction and clustering by optimum-path forest (OPF). The method assumes skull stripping and corrects voxel values in the brain based on local estimations of the white-matter intensities. We use two segmentation steps with different parameters (e.g., graph topologies and image features). The first separates the cerebrospinal fluid (CSF) from the white-matter (WM) and gray-matter (GM) tissues and the second separates WM from GM voxels. In each step, random samples are uniformly estimated to form a small unlabeled training set. Groups of training voxels with similar features are obtained by OPF clustering, a class label is assigned to each group and the labels are propagated to the remaining voxels. Each step is also repeated a few times (e.g., 3) to take the majority vote as the final label. The method was evaluated using several data sets from three protocols; control subjects, phantoms and patients; quantitative and qualitative evaluation methodologies; and two state-of-the-art approaches. The results show that it can solve brain tissue segmentation in less than 5 minutes on modern PCs with higher accuracy and robustness than the baseline approaches.

1 Introduction

Brain tissue segmentation (classification) from Magnetic Resonance (MR) images is considered in the study of a variety of pathologies, such as Alzheimer's disease [1, 2, 3], Multiple

*LIV - Institute of Computing - University of Campinas - Av. Albert Einstein, 1251, - Cidade Universitária "Zeferino Vaz" 13083-852 - Campinas - SP - Brazil. Financial support by Fapesp (05/59902-7)

†LIV - Institute of Computing - University of Campinas - Av. Albert Einstein, 1251, - Cidade Universitária "Zeferino Vaz" 13083-852 - Campinas - SP - Brazil

‡Department of Neurology/Neurosurgery - School of Medical Science - University of Campinas - Rua Tessália Vieira Camargo, 126 Cidade Universitária "Zeferino Vaz" 13083-887 - Campinas - SP - Brazil

§Medical Image Processing Group Department of Radiology - University of Pennsylvania - Blockley Hall - Fourth Floor 423 Guardian Drive - Philadelphia, PA 19104-6021, USA

Sclerosis [4] epilepsy [5, 6], schizophrenia [7], brain tumor [8], frontotemporal dementia [9], migraine [10], in addition to other applications [11, 12, 13, 14, 15]. The problem consists of isolating the brain immersed in cerebrospinal fluid (CSF) from a given MR-image of the head (*skull stripping*) and assigning a distinct label to gray-matter (GM), white-matter (WM) and CSF voxels (*voxel classification*). Some methods perform skull stripping before voxel classification [16, 17], but both operations can also be performed at the same time [18, 19]. Most methods also apply some kind of inhomogeneity correction before voxel classification. They may be based on frequency filtering [20, 21, 22], surface fitting [23, 16, 24, 25], stochastic models [17, 26, 27, 28] and energy minimization [29, 30]. Voxel classification can be obtained based on thresholding [31], region growing [32], clustering [33, 34], supervised learning [35, 36, 37, 17, 38, 39] and deformable models [40, 41]. Several methods also require image deformation into a standard space of a probabilistic template, whose predefined labels can be used as prior information [17, 36, 32, 42, 43, 44, 45].

Despite the large number of key contributions as reported in several surveys [46, 47, 48, 49], there is no standard approach which can outperform the other methods in all situations involving different subjects, imaging protocols. Moreover, there is no standard evaluation methodology for medical image segmentation [49] and most evaluation methods take into account ground-truths, which can not be perfect [50].

In this work, we propose a fast, accurate and robust to anatomical variations method for brain tissue segmentation based on *inhomogeneity correction* and *clustering by optimum-path forest* [51] (OPF). The method assumes that the brain (with CSF included) has been extracted from the image by skull stripping [52, 53, 54, 55, 56]. However, it requires stripping methods which can remove the brain stem [52, 53]. Our approach for inhomogeneity correction exploits the WM intensities as reference values, similarly to [23]. However, instead of fitting spline-based surfaces to the voxel intensities, we use a faster, simpler and yet effective correction function. A reference value for WM intensity in a neighborhood of each voxel is estimated and used to adjust that function locally.

After inhomogeneity correction, the brain voxels are classified in two segmentation steps with different parameters each (e.g., graph topologies and image features). The first separates CSF from WM and GM tissues and the second separates GM from WM. In each step, random samples are uniformly estimated to form a small unlabeled training set (e.g., a set with 500 voxels). A feature vector is assigned to each voxel and natural groups are found in the feature space by the OPF clustering approach. This approach interprets the samples as nodes of a graph, whose arcs connect the k-nearest neighbors (k-nn) in the feature space. The nodes are weighted by a probability density function (pdf) and the OPF algorithm computes an optimum-path forest in the k-nn graph, such that each tree (cluster) is rooted at one maximum of the pdf. The best value of k is found by optimization. Class labels are assigned to each group based on the knowledge of the order relation between intensities of different tissues. The class labels are then propagated to the remaining brain voxels. For robustness, each step is repeated a few times (e.g., 3) to take the majority vote as the final label.

The method presents considerable improvements with respect to our previous works [51, 57]. The processes to estimate training samples and assign class labels are more robust, inhomogeneity correction and label propagation are faster and more effective, and the method

has been extended to segment CSF, WM and GM in MR-images weighted on T1, T2, or PD, rather than only the two tissues in T1 images. We have also included extensive evaluation experiments using those three imaging protocols; images from control subjects, phantoms and patients; quantitative and qualitative evaluation methodologies; and two state-of-the-art approaches [16, 17]. Quantitative evaluation was based on *Delineation Operating Characteristic* [58], involving 38,400 experiments by varying 4 parameters of each method, and subjective evaluation used scores from two experts.

Note that, some methods take several minutes to separate CSF, GM and WM [59, 43] and others require multiple imaging protocols [42, 60]. Our approach completes segmentation in less than 5 minutes on modern PCs, using a single imaging protocol and with no need of registration with templates, which makes it more independent of anatomical variations and free of registration errors.

Inhomogeneity correction and brain voxel classification based on OPF are presented in Sections 2 and 3, respectively. Section 4 describes the evaluation methodologies and experiments are shown in Section 5. The results indicate that the proposed pipeline for brain tissue segmentation is fast and can be more accurate and robust than the baseline approaches [16, 17]. Finally, we state conclusions in Section 6.

2 Inhomogeneity correction

For a given MR-image of the brain, after skull stripping and without the brain stem, our approach for inhomogeneity correction assumes that there exists some WM voxels within a given neighborhood of each brain voxel to estimate a reference value [1]. This strategy seems reasonable, because WM voxels are the most sensitives to the inhomogeneity effects and they appear at least close to all parts in the brain. Then, the intensity correction function may change for each brain voxel depending on the reference value.

Let $\mathcal{N}$ be the set of voxels from CSF, GM and WM, and $I(s)$ be the original intensity of voxel $s \in \mathcal{N}$. We define a spherical neighborhood by a set $\mathcal{A}_r(s)$ containing image voxels t whose Euclidean distance to s is less or equal to r. WM voxels are usually included in $\mathcal{A}_r(s)$ for $r = 18.30$mm. Due to noise, we estimate the reference value $M(s)$ as the median value among the $n \ll \|\mathcal{A}_r(s)\|$ brightest (in T1) or darkest (in T2 and PD) voxels in $\mathcal{A}_r(s)$. The value $n = 15$ is usually enough. Finally, each voxel in $\mathcal{N}$ has its intensity $I(s)$ corrected to $J(s)$ by

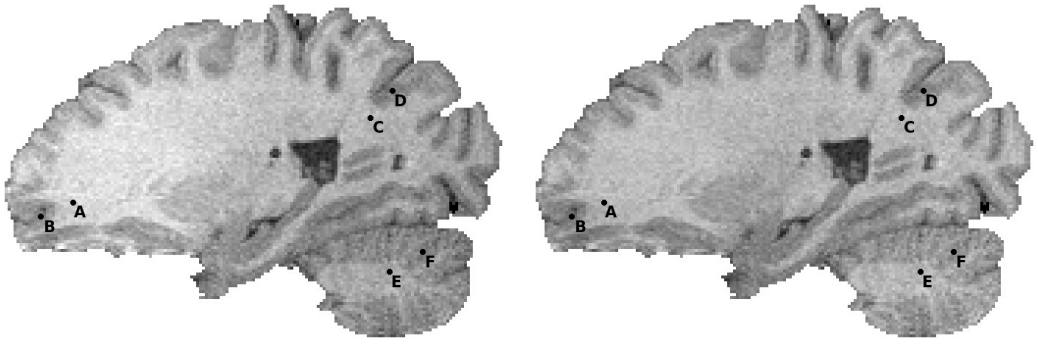
$$J(s) \quad = \quad I_{\max} - (M(s) - I(s))(2 - M(s)/I_{\max})^{\alpha}, \tag{1}$$

where $I_{\max}$ is the maximum voxel intensity in $\mathcal{N}$ and, similarly in T2 and PD images, each voxel in $\mathcal{N}$ has its intensity corrected to

$$J(s) \quad = \quad I_{\min} + (I(s) - M(s))(1 + I_{\min}/M(s))^{\alpha}, \tag{2}$$

where $I_{\min}$ is the minimum voxel intensity in $\mathcal{N}$. The differences between $I(s)$ and $M(s)$ in both equations tend to be higher in brighter regions than in darker regions of the brain, due to the multiplicative effect of the inhomogeneity [48]. Therefore, the idea is to increase the weight of those differences in the intensity correction of the darker regions than in

[1] This means that voxels brighter (in T1) or darker (in T2 and PD) than WM should also be removed during brain stripping.



(a) (b)

Figure 1: (a) Brain image after skull stripping. Voxel intensities in some points of WM (a=1361, c=1155, e=1069) and GM (b=961, d=777, f=749). (b) After inhomogeneity correction, the intensities at the same points provide better separation between WM (a=3041, c=3056, e=3019) and GM (b=2222, d=2136; f=2228).

the correction of the brighter ones. The multiplicative factors $(2 - M(s)/I_{\max})$ and $(1 + I_{\min}/M(s))$ play this role in the above equations within the interval $[1, 2]^{\alpha}$, being $\alpha \geq 0$ a constant to control that effect (e.g., $\alpha = 0.33$). Afterwards, the values of $J(s)$ are linearly normalized within $[0, 4095]$.

Figure 1 presents an example of our method on a T1 image of the human brain. The WM intensity $I(c)$ is lower than the GM intensity $I(b)$, for example, but the method corrects all intensities to have WM voxels brighter than GM voxels in most parts of the brain.

3 Voxel classification

After inhomogeneity correction, brain voxels in $\mathcal{N}$ must be classified into CSF, WM and GM in two steps by following the pipeline shown in Figure 2. First, this pipeline is executed to separate CSF from GM and WM, and then it is executed to separate GM from WM. This was necessary because some parameters are different for each step, as we will describe in this section.

The pipeline consists of image feature extraction for $\mathcal{N}$ and random sampling to obtain a small unlabeled training set $\mathcal{T} \subset \mathcal{N}$. Groups of voxels with similar image features are then determined by the OPF clustering approach. Note that, each class may be represented by multiple groups. Some clustering approaches force the number of desired groups, but this usually increases the classification errors. Class labels are assigned to each group based on the knowledge of the order relation between intensities in different tissues, which differs in T1 with respect to T2 and PD. The class labels are then propagated to the remaining voxels in $\mathcal{N}$.

Given that $\mathcal{T}$ is a random set, we verify the need for re-sampling after group labeling and after segmentation (dashed lines in Figure 2). While the proportion between tissues is not within a given *acceptance interval* (in $\mathcal{T}$, after group labeling, and in $\mathcal{N}$, after segmentation), the preceding parts (except feature extraction) of the pipeline are repeated. However, in

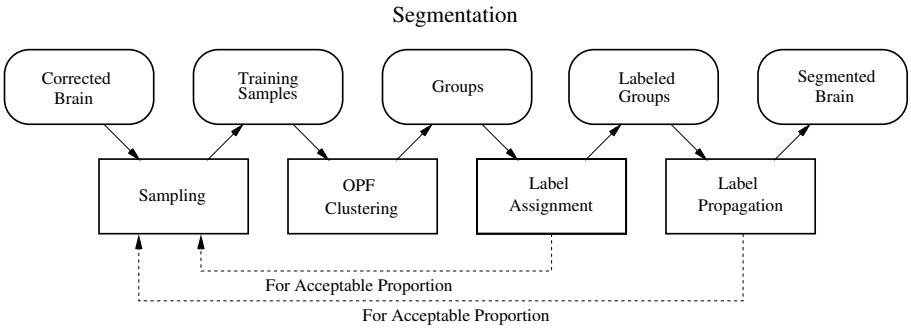


Figure 2: Voxel classification consists of feature extraction; random sampling to form a small training set; clustering by OPF to obtain natural groups; class label assignment to define a tissue label for each group; and label propagation to classify the remaining brain voxels. An acceptance criterion is used to repeat sampling and subsequent processes until the method achieves a reasonable result (dashed lines).

order to guarantee convergence, the acceptance interval increases at every 20 iterations of the loop. This makes the method more robust to abnormal brains. In practice, the convergence usually requires a few iterations of that loop. The whole pipeline is also repeated a few times (e.g., 3) to obtain the final label of each voxel by majority vote.

The details of the method are given in the next sections.

3.1 Feature extraction and voxel sampling

In order to separate CSF from GM and WM, the feature vector of each voxel $s \in \mathcal{N}$ is composed by its corrected brightness $J(s)$ and the two closest intensities to $J(s)$ within the 6-neighborhood of s. This feature vector is more sensitive to noise, being unappropriated to classify WM and GM, but excellent to classify subtle CSF voxels in the sulci of the brain. For segmentation of WM and GM, we use $J(s)$ and the six intensities around the median value of the voxels within the 26-neighborhood of s. This feature vector preserves the borders of the tissues and is robust to high-frequency noises.

An unlabeled training set $\mathcal{T}$ with 500 voxels is estimated by uniform-random sampling in the intensity interval of the voxels to be classified in $\mathcal{N}$.

3.2 Clustering by OPF

In the unlabeled training set $\mathcal{T}$, each sample s is represented by a feature vector $\vec{v}(s)$. Let $d(s,t)$ be the distance between s and t in the corresponding feature space (e.g., $d(s,t) = \|\vec{v}(t) - \vec{v}(s)\|$). The fundamental problem in data clustering is to identify natural groups in $\mathcal{T}$. These groups should have a minimum of overlapping between distinct tissues.

A graph $(\mathcal{T}, \mathcal{A})$ is defined such that the arcs $(s,t) \in \mathcal{A}$ connect k-nearest neighbors (k-nn) in the feature space. A(s) denotes the set of arcs that reaches the node s. The arcs are weighted by $d(s,t)$ and the nodes $s \in \mathcal{T}$ are weighted by a probability density value $\rho(s)$ [51].

$$\rho(s) = \frac{1}{\sqrt{2\pi\sigma^2}|\mathcal{A}(s)|} \sum_{\forall t \in \mathcal{A}(s)} \exp\left(\frac{-d^2(s,t)}{2\sigma^2}\right) \quad (3)$$

where $|\mathcal{A}(s)| = k$, $\sigma = \frac{h_f}{3}$, and h_f is the maximum arc weight in $(\mathcal{T}, \mathcal{A})$ [2]. This parameter choice considers all nodes for density computation, since the Gaussian function covers most samples within $d(s,t) \in [0, 3\sigma]$. The traditional method to estimate a probability density function (pdf) is by Parzen-window. Equation 3 can provide a Parzen-window estimation based on isotropic Gaussian kernel when we define the arcs by $(s,t) \in \mathcal{A}$ if $d(s,t) \leq h_f$. This choice, however, presents problems with the differences in scale and sample concentration. Solutions for this problem lead to adaptive choices of h_f depending on the region of the feature space [61]. By taking into account the k-nn, we are handling different concentrations and reducing the scale problem to the one of finding the best value of k within $[k_{\min}, k_{\max}]$, for $1 \leq k_{\min} < k_{\max} \leq |\mathcal{T}|$. For 500 nodes, we have found that $k_{\min} = 10$ and $k_{\max} = 30$ are suitable values for brain tissue separation. Our solution considers the minimum graph cut in $(\mathcal{T}, \mathcal{A})$ as provided by the clustering results for $k \in [k_{\min}, k_{\max}]$, according to the cut measure $C(k)$ [62].

$$C(k) = \sum_{i=1}^{c} \frac{W_i'}{W_i + W_i'}, \quad (4)$$

$$W_i = \sum_{\forall (s,t) \in \mathcal{A} | L(s)=L(t)=i} \frac{1}{d(s,t)}, \quad (5)$$

$$W_i' = \sum_{\forall (s,t) \in \mathcal{A} | L(s)=i, L(t)\neq i} \frac{1}{d(s,t)}, \quad (6)$$

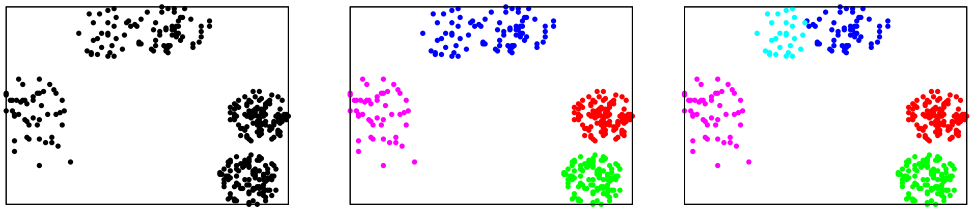
where $L(t)$ is the label of sample t, W_i' uses all arc weights between cluster i and other clusters, and W_i uses all arc weights within cluster $i = 1, 2, \ldots, c$.

Figure 3a shows an example with $|\mathcal{N}| = 340$ samples borrowed from [51], which forms a few clusters with different sample concentrations in a 2D feature space. Depending on the scale, there are one, three, four, or five natural groups. If $k_{\min} = 1$ and $k_{\max} \geq 150$, then the minimum cut will occur when all samples are grouped into a single cluster. The minimum cut for $k_{\min} = 1$ and $k_{\max} = 100$ identifies four clusters with a best $k = 37$ (Figure 3b), and by limiting the search within $[1, 30]$, the minimum cut identifies five clusters with a best $k = 29$ (Figure 3c).

The clustering results are obtained by extending the image foresting transform [63] (IFT) from the image domain to the feature space, as follows.

A path π_t is a sequence of adjacent samples starting from a root $r \in \mathcal{R}$, where $\mathcal{R}$ being a root set with one element for each local maximum of the pdf, and ending at a sample t, $\pi_t = \langle t \rangle$ being a trivial path and $\pi_s \cdot \langle s,t \rangle$ the concatenation of π_s and arc (s,t). Among all possible paths π_t with roots on the maxima of the pdf, we wish to find a path whose lowest

[2] After Equation 3 is computed, we include extra arcs in $\mathcal{A}$ to guaranty symmetry among adjacent nodes with the same density value [51].



(a) (b) (c)

Figure 3: (a) Feature space with distinct sample concentration per cluster. We can identify different cluster numbers depending on the scale. Reasonable solutions are (b) four and (c) five clusters, where two touch each other.

density value along it is maximum. Each maximum should then define one influence zone (cluster) by selecting the samples that are more strongly connected to it, according to this definition, than to any other maximum. The extra arcs in $\mathcal{A}$, which guarantee symmetry on density plateaus, and will also guarantee a single cluster per maximum of the pdf. More formally, we wish to find $V(t) = \max_{\forall \pi_t}\{f(\pi_t)\}$ where

$$\begin{aligned} f(\langle t \rangle) &= \begin{cases} \rho(t) & \text{if } t \in \mathcal{R} \\ \rho(t) - \delta & \text{otherwise} \end{cases} \\ f(\langle \pi_s \cdot \langle s,t \rangle \rangle) &= \min\{f(\pi_s), \rho(t)\} \end{aligned} \tag{7}$$

for $\delta = \min_{\forall (s,t) \in \mathcal{A} | \rho(t) \neq \rho(s)} |\rho(t) - \rho(s)|$. Higher values of delta reduce the number of maxima. We are setting $\delta = 1.0$ and scaling real numbers $\rho(t) \in [1, 1000]$ in this work. The OPF algorithm solves the maximization problem by computing an optimum-path forest — a predecessor map P with no cycles that assigns to each sample $t \notin \mathcal{R}$ its predecessor $P(t)$ in the optimum path from $\mathcal{R}$ or a marker *nil* when $t \in \mathcal{R}$.

Algorithm 1 – Clustering by Optimum Path Forest

Input: Graph $(\mathcal{T}, \mathcal{A})$ and function ρ.
Output: Group label map L, connectivity map V, forest P, and a list $\mathcal{L}$ of ordered nodes.
Auxiliary: Priority queue Q, variables *tmp* and $l \leftarrow 1$.

1. *For all* $s \in \mathcal{T}$, *set* $P(s) \leftarrow nil$, $V(s) \leftarrow \rho(s) - \delta$, *insert* s *in* Q.
2. *While* Q *is not empty, do*
3. *Remove from* Q *a sample* s *such that* $V(s)$ *is maximum and insert* s *in* $\mathcal{L}$.
4. *If* $P(s) = nil$, *then*
5. *Set* $L(s) \leftarrow l$, $l \leftarrow l + 1$, *and* $V(s) \leftarrow \rho(s)$.
6. *For each* $t \in \mathcal{A}(s)$ *and* $V(t) < V(s)$, *do*
7. *Compute* $tmp \leftarrow \min\{V(s), \rho(t)\}$.
8. *If* $tmp > V(t)$ *then*
9. *Set* $L(t) \leftarrow L(s)$, $P(t) \leftarrow s$, $V(t) \leftarrow tmp$.
10. *Update position of* t *in* Q.

Algorithm 1 identifies one root in each maximum of the pdf ($P(s) = nil$ in Line 4 implies $s \in \mathcal{R}$), assigns to each root a distinct group label in Line 5, and computes the influence zone

(cluster) of each root as an optimum-path tree in P, such that the nodes of the tree receive the same label as its root in a map L (Line 9). It also outputs the optimum connectivity map V, forest P and a list $\mathcal{L}$ of voxels of $\mathcal{T}$ in their decreasing order of optimum value in V. This list is a novelty with respect to the algorithm used in [51, 57], which aims at fast class labeling propagation for the remaining samples in $\mathcal{N}$ (Section 3.4).

Algorithm 1 takes $O(k|\mathcal{T}| + |\mathcal{T}|\log\mathcal{T})$ computations, when Q is a binary heap, and the estimation of the best k requires its computation several times. This can take several minutes on modern PCs for $|\mathcal{T}| > 1000$. The problem becomes unmanageable for typical sets $\mathcal{N}$ with brain tissues. This explains the need for sampling a subset $\mathcal{T}$ from $\mathcal{N}$.

3.3 Class label assignment and verification for re-sampling

The clustering results are natural groups in the feature space, which may not correspond to the number of classes. Class label assignment in $\mathcal{T}$ is then necessary.

First consider the separation of CSF from GM and WM in T1 images. We expect that the obtained proportion P'_{csf} of CSF voxels in $\mathcal{T}$ be within an *acceptance interval* around the proportion P_{csf} obtained by Otsu's thresholding in $\mathcal{N}$. This interval is defined by $[(1-\delta)P_{csf}, (1+\delta)P_{csf}]$, for $0.0 \leq \delta \leq 1.0$ (e.g., $\delta = 0.2$). Therefore, the OPF clusters are relabeled from the darkest to the brightest one as CSF until P'_{csf} be the closest to P_{csf}. If $P'_{csf} \notin [(1-\delta)P_{csf}, (1+\delta)P_{csf}]$, then re-sampling and OPF clustering are repeated. In order to guarantee convergence, the value of δ increases at every 20 iterations without satisfying the acceptance criterion.

The same strategy is used to separate GM from WM in T1 images. In this case, however, Otsu's threshold is substituted by the mean intensity in the brain without CSF, in order to find a reasonable GM proportion. The value of δ and the rest of the process are the same. This method is more accurate and robust than the one used in [51, 57] for GM and WM separation in T1 images.

In T2 and PD images, the darkest tissue is WM and the brightest substance is CSF. So, the method is the same but the labeling process starts from the brightest to the darkest clusters.

3.4 Fast label propagation and verification for re-sampling

According to [51, 57], the class labels of the obtained clusters can be incrementally propagated to the remaining voxels $t \in \mathcal{N}\backslash\mathcal{T}$ by considering the following equation.

$$V(t) = \max_{\forall(s,t)\in\mathcal{A}} \{\min\{V(s), \rho(t)\}\}. \tag{8}$$

If $s^* \in \mathcal{T}$ is the node that satisfies this equation, then its class label can be simply assigned to t.

The condition $(s,t) \in \mathcal{A}$ in the above equation implies that t could be a k-nearest neighbor of s, if t belonged to the graph. This condition can be easily determined by storing the maximum distance $\omega(s)$ between s and its k-nearest neighbors in the graph for every $s \in \mathcal{T}$. Therefore, the candidate nodes for s^* must satisfy $d(s,t) \leq \omega(s)$.

The expensive part in the above equation is the computation of $\rho(t)$ for all $t \in \mathcal{N} \backslash \mathcal{T}$, which also requires the computation of the k-nearest neighbors of t in $\mathcal{T}$. We have considerably speeded up label propagation by avoiding the computation of $\rho(t)$ and choosing the node s^* which satisfies

$$V(s^*) = \max_{\forall s \in \mathcal{L}, d(s,t) \leq \omega(s)} \{V(s)\}, \tag{9}$$

where $\mathcal{L}$ is the ordered list resulting from Algorithm 1. Note that s^* is the first node in $\mathcal{L}$ which satisfies this equation, also avoiding the computation of $d(s,t)$ for the remaining nodes in $\mathcal{L}$. This variant is faster and yet effective, producing accurate results.

After label propagation, the same process discussed in Section 3.3 is applied to verify the need for re-sampling. In this case, however, the obtained tissue proportions are directly computed in $\mathcal{N}$ rather than in $\mathcal{T}$.

4 Methodology of evaluation

The evaluation of image segmentation methods is usually as complicated as the segmentation process itself, due to the absence of perfect ground truths. Even when the same specialist segments the same image several times, the results may suffer considerable variations [50, 64, 39]. Synthetic brain models (phantoms) may be an alternative [64] for evaluation using more reliable ground truths, but they do not provide the same authenticity of challenges as real images do [65]. Despite these drawbacks, it is important to consider both phantoms and real images with their available ground truths for evaluation. Besides, one should never expect 100% accuracy for any ground truth and segmentation method.

The choice of suitable effectiveness measures is also a very important factor in segmentation evaluation. These measures may be either objective (quantitative) or subjective (qualitative). Quantitative measures may take into account the distance between object contours in the ground truth and segmented images (e.g. Hausdorff distance [66]) or the differences between the overlaid segmentation and ground-truth masks. We use the second option.

Four basic measures are derived from the overlay between the segmented object and ground-truth masks [65]: the size of their intersection (*True Positive*, TP); the size of the intersection between object complement (background) and ground-truth masks (*False Negative*, FN); the size of the intersection between object and ground-truth complement masks (*False Positive*, FP); and the intersection between the background and ground-truth complement masks (*True Negative*, TN). FP and FN are empty sets in the case of perfect segmentation. These elementary measures can then be used to compose more complex ones:

$$\begin{aligned}
m_1(r, \hat{I}) &= \frac{|FP|}{|FP| + |TN|} \\
m_2(r, \hat{I}) &= \frac{|FN|}{|FN| + |TP|} \\
m_3(r, \hat{I}) &= \frac{2 * |TP|}{2 * |TP| + |FP| + |FN|} \\
m_4(r, \hat{I}) &= \frac{|TP|}{|TP| + |FP| + |FN|}
\end{aligned}$$

where r is the instance of parameters from a given segmentation method. Measure m_1 is

called FP volume Fraction, m_2 is FN volume Fraction, m_3 is Dice metric, and m_4 is Jaccard metric. Most published papers use m_3 and/or m_4 for brain tissue segmentation evaluation, because they are invariant with the background size.

We propose to use the *Delineation Operating Characteristic*(DOC) curve for evaluation [58], because it has the same advantage, and also takes into account an operational range of parameter values for each segmentation method Γ, set $\mathbf{r}$ of the instances of its parameters vectors, and image $\hat{I}$. By performing segmentation and plotting m_1 and $1-m_2$, we get a point on the DOC curve which indicates the performance of that method for each given choice of parameter values and image. The curve is fitted on several points by using the logarithmic model below (see plots in Figure 6).

$$y = \log_a((a-1)x+1).$$

The coordinates $(0,0)$ and $(1,1)$ represent the extreme cases of worst behavior where the method outputs a perfect background image, and the entire image domain as its segmentation, respectively.

The DOC curve can also give measures m_5 and m_6 defined below.

$$\begin{aligned} m_5(\hat{I}) &= \min \sqrt{m_2^2(r,\hat{I}) + m_1^2(r,\hat{I})}, \forall r \in \mathbf{r} \\ m_6(\hat{I}) &= \int_{x=0}^{1} \log_a((a-1)x+1). \end{aligned}$$

Measure m_5 represents the smallest segmentation error (in terms of simultaneously achieving minimum FP and FN) among those obtained from the tested parameters, using method Γ and image $\hat{I}$. Measure m_6 indicates robustness and accuracy of method Γ to segment image $\hat{I}$. Note that, the distribution of the operational points in Figure 6 indicates how much the parameter vectors in the set $\mathbf{r}$ must be refined to achieve a good result.

Those above measures m_5 and m_6 lead to an evaluation free from adjustment of parameters. Moreover, the DOC curve indicates the best operational point for a given method and image. The DOC curve also shows when the errors of the method are by excess (over segmentation) or by missing true object parts. This information is not clear in both Dice and Jaccard metrics. In fact, segmentations where $|FP| > |FN|$ receive a better score by m_3 and m_4 than segmentations where $|FN| < |FP|$.

Finally, m_5 is applied to a set of n images $\hat{\mathbf{I}} = \{\hat{I}_i\}, i \in \{1,2,...,n\}$. To facilitate this, we use the mean values of m_1 and m_2 given by:

$$\begin{aligned} \bar{m_1}(r,\hat{\mathbf{I}}) &= \frac{1}{n}\sum_{1}^{n} m_1(r,\hat{I}_i). \\ \bar{m_2}(r,\hat{\mathbf{I}}) &= \frac{1}{n}\sum_{1}^{n} m_2(r,\hat{I}_i). \end{aligned}$$

Then, for the set of images $\hat{I}_i$, $i \in 0,1,...,n$, we have

$$\bar{m_5}(\hat{\mathbf{I}}) = \min \sqrt{\bar{m_1}^2(r,\hat{\mathbf{I}}) + \bar{m_2}^2(r,\hat{\mathbf{I}})}, \forall r \in \mathbf{r} \quad (10)$$

Measure $\bar{m}_6$ is also defined in a similar way by interpolating the mean values of $\bar{m}_1(r, \hat{\mathbf{I}})$ and $1 - \bar{m}_2(r, \hat{\mathbf{I}})$, $\forall r \in \mathbf{r}$.

When ground truths are not available, we use the following subjective evaluation. Specialists perform visual analysis based on orthogonal slices of the data volume and assign grades to the segmentation results of each method from 1(bad) to 5(good) in a blinded manner. In order to avoid inconsistencies, the segmentation results from all methods are analysed at the same time for any chosen slice without a knowledge of the association between results and methods. These results are also shown by overlaying the original image with a semi-transparent object mask or with the object's borders. Partial volume is tolerated and grades from 2 to 4 require explanation from the specialist who describes where the segmentation errors are.

5 Experiments and results

We have evaluated the proposed pipeline — inhomogeneity correction, CSF/GMWM segmentation and GM/WM segmentation using the optimum-path forest (OPF) algorithm — in comparison with two popular methods: FAST (*FMRIB's Automated Segmentation Tool* available in the FSL framework) [17], which corrects inhomogeneity during tissue segmentation, and PVC (*Partial Volume Classifier* available in the BrainSuite framework), which uses the Bias Field Corrector (BFC) to handle inhomogeneity [16]. In all cases, the brain was first isolated from the image by using the same skull stripping procedure.

The experiments involved four databases of 3D images: 18 MRT1-images from the Internet Brain Segmentation Repository (IBSR) [3], 9 synthetic images (three from each MRI protocol T1, T2 and PD) from BrainWeb Phantoms (BWP)[4] with different levels of noise and inhomogneity, 10 MRT1-images from the University of Campinas (UNICAMP), and 10 MRT2-images from the Department of Radiology from University of Pennsylvania (UPENN). We also included 10 patient MRT1-images from the University of Campinas (PATIENTS) images, being 5 preoperative and 5 postoperative, to evaluate the rubustness of the methods to anatomy variations.

5.1 Quantitative experiments

Quantitative evaluation based on the DOC curve was performed for BWP and IBSR databases, because they have ground-truth images. The execution time to build a DOC curve may be considerably long, since it depends on the computational time of each method and its number of parameters. For the sake of simplicity, we first selected 4 instances of 4 parameters for each method in a given range to test all their combinations. After defining the best combination of parameters, we instantiated the parameters again around that point in a closer adjacency to refine the results. So, each method was executed 512 times for each 3D image for finding the best operating point via the DOC curve.

[3] http://www.cma.mgh.harvard.edu/ibsr/
[4] http://mouldy.bic.mni.mcgill.ca/brainweb/

For the OPF pipeline, the selected parameters and their instances were the expected GM proportion P_{gm} (from 0.4 to 0.90), the GM proportion interval δ (from 0.1 to 0.4), the inhomogeneity radius in A_r (from 13.0 to 26.0), and the parameter α (0.0 to 1.0) of the inhomogeneity correction technique. For FAST, we selected the bias field smoothing extent (from 2 to 47), the initial segmentation spatial smoothness (from 0.001 to 0.1), the segmentation spatial smoothness (from 0.01 to 0.9), and the number of iterations during bias-field removal (from 1 to 8). For PVC and BFC, we picked the spatial prior strength (from 0.01 to 0.8), the histogram radius (from 8 to 16), the bias sample spacing (from 8 to 24), and the control point spacing (from 32 to 96). For FAST and PVC we took the parameter intervals around the automatic values suggested by the software.

Figures 4 and 5 show m_5 and m_6, respectively, for each of the three methods over the IBSR database. Note that the OPF pipeline achieved the best performance in all cases. For m_5, we also plotted the mean metric value of each method using label "R1", and with label "R2", the metric value for the parameter combination which gave the minimum total error over the 18 images. Although FAST could obtain a better result than PVC for the images individually, their performances were similar for the best parameter combination. Similarly, for measure m_6 and its mean value (label "R1"), OPF was superior to both compared approaches.

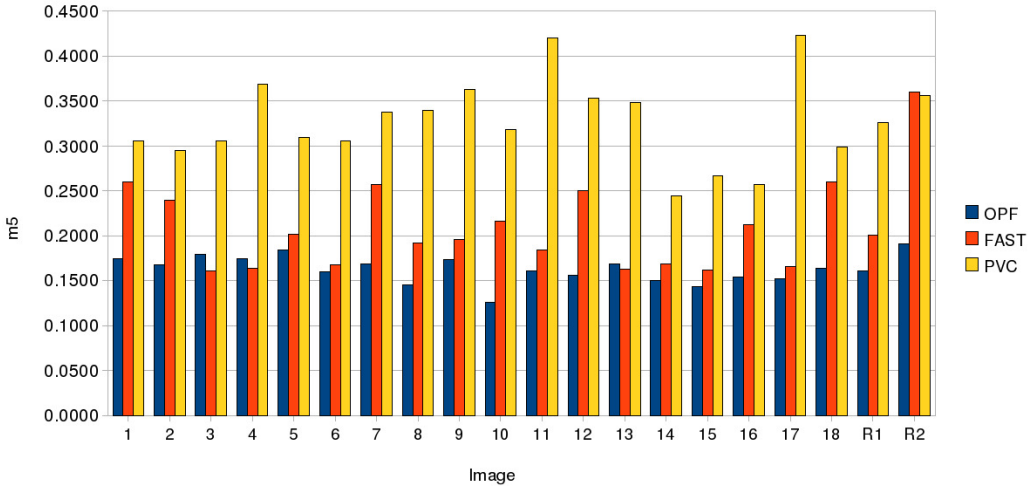


Figure 4: Measure m_5 for OPF, Fast and PVC method over the IBSR database. The mean value $\bar{m}_5$ (label "R1") and the best parameter combination (label "R2") over the 18 images are also shown. Lower values indicate better performance.

Figure 6 shows samples of the operational points for each method in image 15 of the IBSR database and their interpolated DOC curves. Their distribution suggest that the OPF pipeline can be tuned more easily for optimal setting.

The same analysis was conducted over the BWP data set. Figures 7 and 8 show m_5 and m_6, respectively, for each of the three methods over the BWP database. The image labels along the x axis show the MRI protocol and the induced noise and inhomogeneity

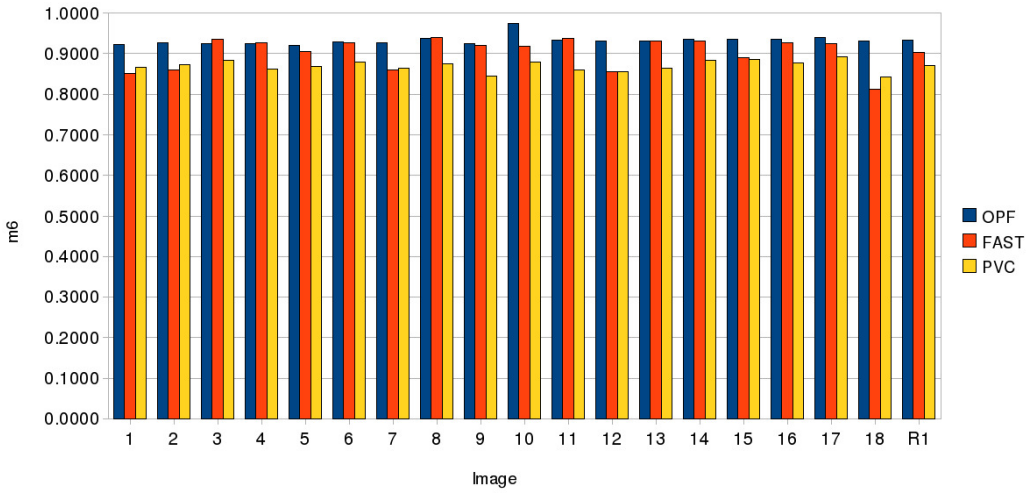


Figure 5: Measure m_6 for OPF, Fast, and PVC methods over the IBSR database. The mean value $\bar{m_6}$ (label "R1") is also shown. Higher values indicate better performance.

levels as follows: "protocol(noise, inhomogeneity)". As PVC does not support T2 and PD protocols, the figures present only OPF and FAST segmentation results over them. The columns "R1" and "R2" have the same meaning as in Figures 4 and 5. The results show that all methods did well OVER BWP data set, but OPF was superior, specially in the "R1" and "R2" cases. The accuracy over PD images was lower because they are pourer in contrast when compared to T1 and T2 images.

We can also see in Figures 7 and 8 that the inhomogeneity was efficiently removed in all protocols, by OPF. The performance of OPF was not compromised by high inhomogeneity levels.

A paired t-test was carried out for comparing methods pairwise on the IBSR and BWP data sets. The results are summarized on Table 1, where the P values are listed. We observe from the table that whenever OPF performed better (most of the cases, except for m_6 in T1-BWP and m_5 in T2 and PD-BWP data sets), its superior performance was statistically significant ($P \leq 0.05$). Any apparent inferior performance was not significant statistically.

5.2 Qulitative experiments

The databases from UPENN, UNICAMP and PATIENTS were chosen for qualitative evaluation. For these tests, the default parameters given by the softwares of each method were used. Two specialists assigned grades from 1 (Unacceptable) to 5 (Excellent) for the segmentation results by visual analysis without a knowledge of the methods associated with the results. Table 2 presents the subjective performance of each method in the three data sets. Given that PVC was not designed for T2 and PD images, we could include it in the comparison based on MRT2-images from UPENN.

Again, a paired t-test was carried out for comparing methods pairwise on the UPENN,

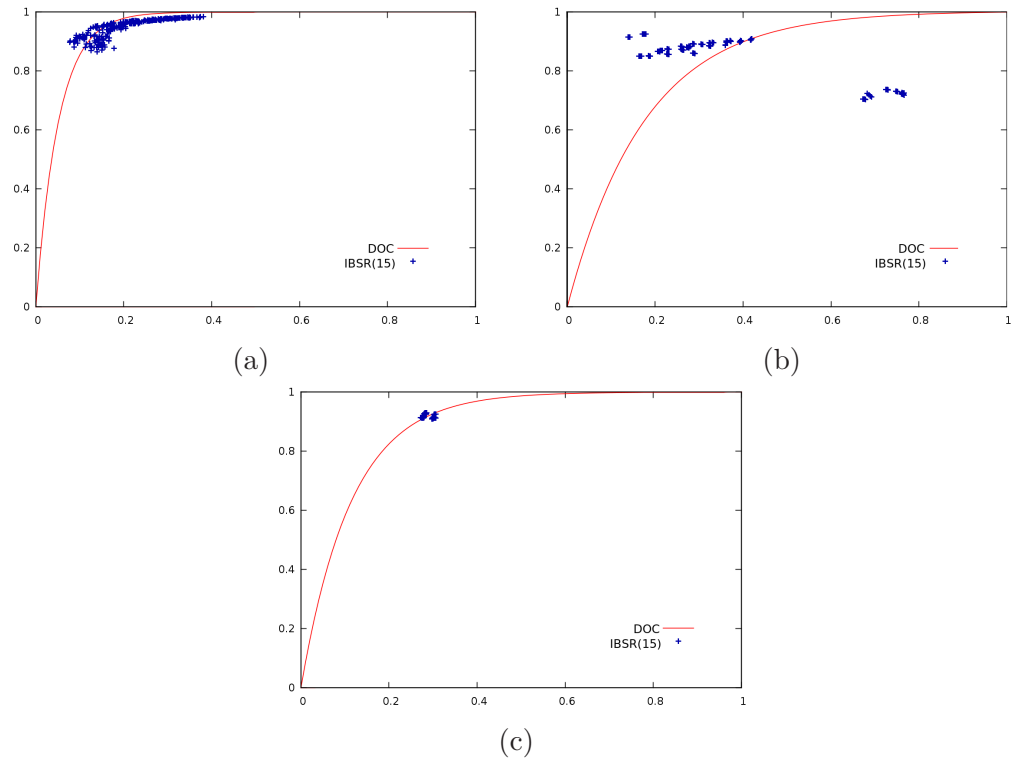


Figure 6: Samples of the operational points in image 15 of the IBSR database for each method and their interpolated DOC curves. (a) OPF, (b) FAST, and (c) PVC.

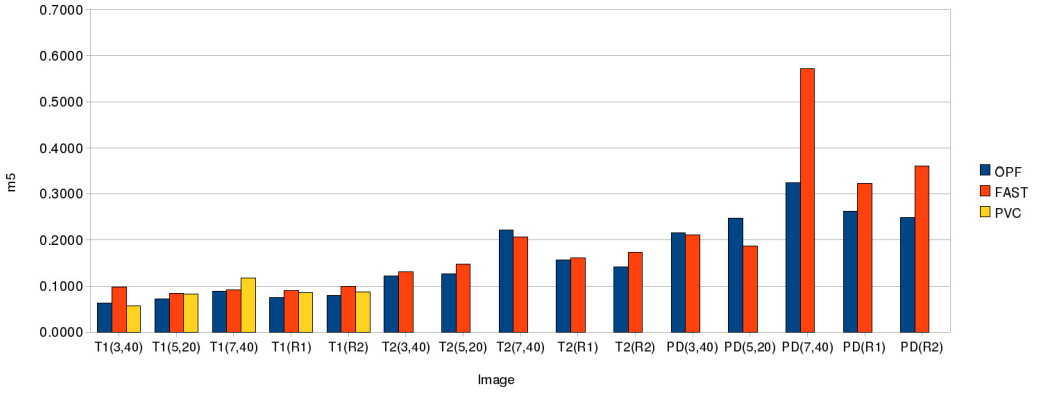


Figure 7: Measure m_5 for the OPF, Fast and PVC methods over the BWP database. The mean value $\bar{m_5}$ (label "R1") and the best parameter combination (label "R2") over the 18 images are also shown. Lower values indicate better performance.

UNICAMP and PATIENTS data sets. The results are summarized on Table 3, where the P values are listed. We observe that the OPF's superior performance was statistically

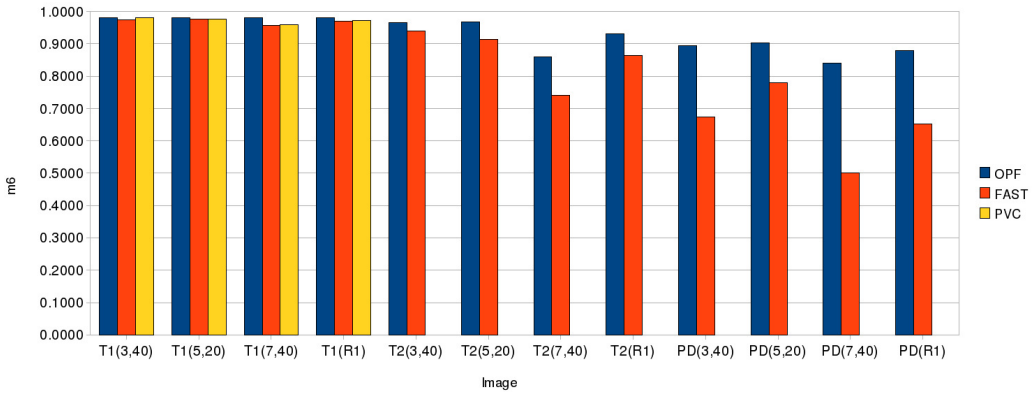


Figure 8: Measure m_6 for the OPF, Fast, and PVC methods over the BWP database. The mean value $\bar{m_6}$ (label "R1") is also shown. Higher values indicate better performance.

Pair	Metric	IBSR	BWP	
			T1	T2 and PD
OPF vs FAST	m_5	0.00	0.00	0.25
	m_6	0.00	0.06	0.01
OPF vs PVC	m_5	0.00	0.02	—
	m_6	0.00	0.15	—
FAST vs PVC	m_5	0.00	0.39	—
	m_6	0.00	0.21	—

Table 1: Comparison of methods pairwise on the IBSR and BWP data sets via paired t-tests. P values are listed.

Data set	Method	Image										
		0	1	2	3	4	5	6	7	8	9	Mean
UPENN	OPF	4.50	5.00	4.50	4.25	4.50	4.50	5.00	4.75	4.50	4.25	4.58
UPENN	FAST	3.50	5.00	4.00	4.50	4.50	5.00	4.00	3.00	4.50	5.00	4.30
UNICAMP	OPF	4.50	4.50	4.50	5.00	4.50	4.50	4.25	5.00	4.40	4.75	4.59
UNICAMP	FAST	3.50	4.50	5.00	4.50	4.50	5.00	4.50	5.00	4.50	4.50	4.55
UNICAMP	PVC	3.50	4.50	3.50	5.00	4.50	5.00	4.50	5.00	4.50	4.50	4.45
PATIENTS	OPF	4.25	4.25	5.00	4.00	4.75	4.50	4.25	4.50	4.25	5.00	4.48
PATIENTS	FAST	3.50	4.25	5.00	4.25	4.00	4.00	4.00	4.25	4.00	4.00	4.13
PATIENTS	PVC	3.00	4.25	5.00	4.00	4.00	3.75	3.75	4.25	4.50	3.25	3.98

Table 2: The mean grades of the OPF, FAST and PVC methods ascertained by two specialists and the final mean grade over the MRT1-images from UNICAMP, UPENN and PATIENTS data sets.

significant ($P \leq 0.05$) on PATIENTS data set. This indicate that the OPF methodology can segment abnormal tissues better than the FAST and PVC methods by the the spetialists' view point.

Although the subjective evaluation could distinguish the best method in some cases, they

Pair	UPENN	UNICAMP	PATIENTS
OPF vs FAST	0.14	0.39	0.01
OPF vs PVC	—	0.20	0.02
FAST vs PVC	—	0.27	0.10

Table 3: Comparison of methods pairwise on the UPENN, UNICAMP and PATIENTS data sets via paired t-tests of the grades given by the specialists. P values are listed.

seem to have similar performance in the mean case from the viewpoint of these specialists. On the other hand, the specialists' comments in the subjective analysis pointed out the weaknesses of each method. The main problems are described in Table 4. Figure 9 illustrates the quality of these segmentations.

Expert	Database	Method	Main Problem
1	UPENN	OPF	Loses GM to WM in putamen.
2	UPENN	OPF	Exchanges GM with WM in the cerebellum.
1	UPENN	FAST	Is more sensitive to noise.
2	UPENN	FAST	Loses GM to CSF.
1	UNICAMP	OPF	Loses GM in the cortex.
2	UNICAMP	OPF	Loses GM to CSF.
1	UNICAMP	FAST	Loses GM to CSF.
2	UNICAMP	FAST	Is more sensitive to noise.
1	UNICAMP	PVC	Loses WM all over the brain.
2	UNICAMP	PVC	Is more sensitive to noise.
1	PATIENTS	OPF	Exchanges WM with GM in the temporal lobes.
2	PATIENTS	OPF	Loses GM to WM. in the girus
1	PATIENTS	FAST	Loses GM to WM in the cortex.
2	PATIENTS	FAST	Is more sensitive to noise.
1	PATIENTS	PVC	Loses GM to WM in the cortex.
2	PATIENTS	PVC	Is more sensitive to noise.

Table 4: Main problems of each method as noted by the specialists during visual analysis.

5.3 Timing issues

We have also evaluated the execution time of the three methods. We ran the methods over IBSR, BWP-T1 and UNICAMP data sets. That is because they contain images of different sizes and because all methods support their protocol (i.e. T1). As we wanted to avoid pathological parameters selection, we used the default parameters of the methods. The tests were executed in a AMD Phenom X4 9650 PC with 8GB of memory.

Table 5 show the minimal, mean and maximal execution time of OPF, FAST and PVC.

The fastest approach, including the time for inhomogeneity correction, is PVC, and the slowest one is FAST. OPF follows PVC closely, taking from 15 seconds to 1 minute to

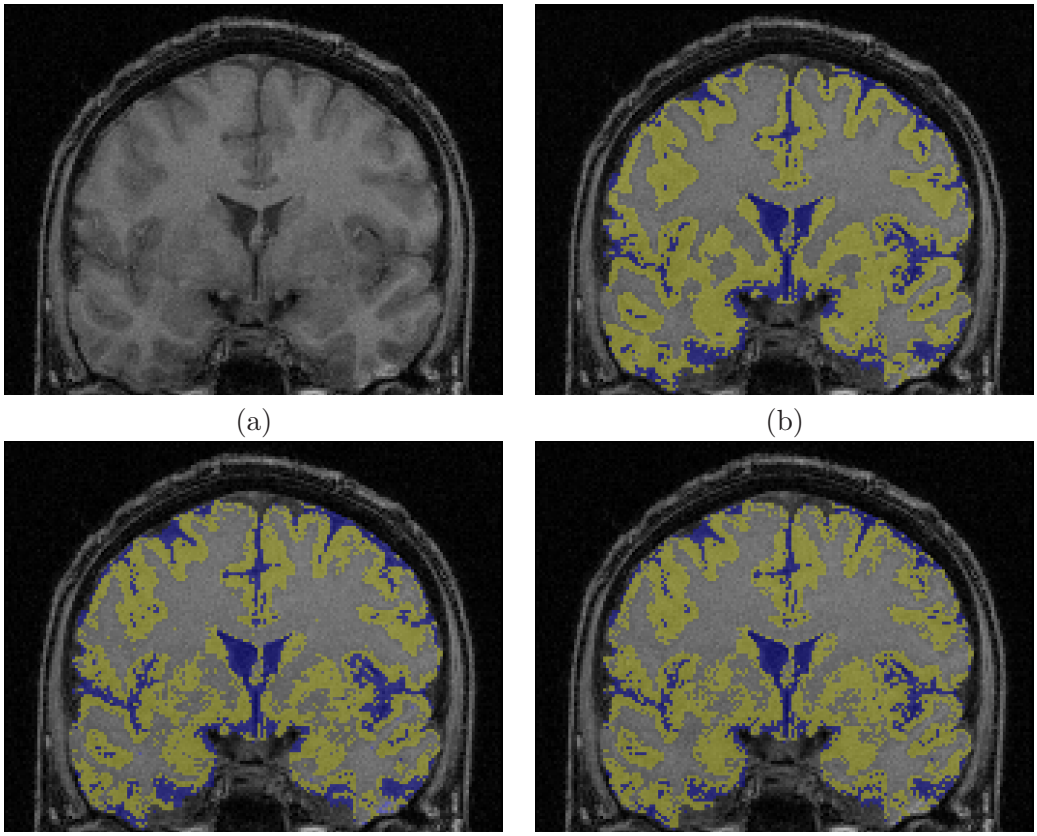


Figure 9: A sample slice (a) of a 3D image from the UNICAMP data set and its brain tissue segmentation by (b) OPF, (c) FAST and (d) PVC.

Method	Minimum(min:sec)	Mean(min:sec)	Maximum(min:sec)
PVC	0:43	2:09	1:05
OPF	0:47	2:29	1:26
FAST	7:16	13:42	9:41

Table 5: Main problems of each method as noted by the specialists during visual analysis.

segment CSF, GM and WM, depending on the input image. The rest of the OPF execution time comes from the inhomogeneity correction. As compared to the GM/WM segmentation by OPF reported in [57, 51], the current algorithm has reduced it from 1.5 minutes to 15 seconds.

6 Conclusion

We have presented a more efficient, accurate, general and robust approach for brain tissue segmentation based on optimum-path forest clustering (OPF). The method includes a novel field-inhomogeneity correction algorithm, a robust scheme for data sampling and tissue labeling from the OPF clusters, ability to handle different MRI protocols (T1, T2 and PD), a suitable strategy to separate WM, GM and CSF, and an extensive and objective parameter-free evalutation strategy and experiments with synthetic and real images.

The results from the objective evaluation based on the DOC curve [58] indicate that the proposed segmentation pipeline is more accurate and robust than two of the most popular brain tissue segmentation methods in most of the cases [16, 17](FAST and PVC). This analysis also suggests that the parameters of the OPF method can be tuned more easily than in the other methods. The subjective evaluation reveals the main problems of each approach and indicates that all methods can present similar good results on average. On preoperative and postoperative brains, OPF did better, indicating that it is robust to anatomical variations.

OPF and PVC can complete segmentation in less than 3 minutes, PVC being slightly faster than OPF. However, PVC works only on T1 images, while OPF handles T1, PD and T2 images with a better accuracy in most of the cases.

References

[1] J. Kesslak, O. Nalcioglu, C. Cotman, Quantification of magnetic resonance scans for hippocampal and parahippocampal atrophy in Alzheimer's disease, Neurology 41 (1) (1991) 51.

[2] C. Jack, R. Petersen, P. O'Brien, E. Tangalos, MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease, Neurology 42 (1) (1992) 183–188.

[3] K. Juottonen, M. Lehtovirta, S. Helisalmi, P. Sr, H. Soininen, Major decrease in the volume of the entorhinal cortex in patients with Alzheimer's disease carrying the apolipoprotein E $\varepsilon 4$ allele, Journal of Neurology, Neurosurgery & Psychiatry 65 (3) (1998) 322–327.

[4] J. Udupa, L. Nyul, Y. Ge, R. Grossman, Multiprotocol MR Image Segmentation in Multiple Sclerosis Experience with Over 1,000 Studies, Academic Radiology 8 (11) (2001) 1116–1126.

[5] G. Cascino, C. Jack Jr, J. Parisi, F. Sharbrough, K. Hirschorn, F. Meyer, W. Marsh, P. O'Brien, Magnetic resonance imaging-based volume studies in temporal lobe epilepsy: pathological correlations, Annals of Neurology 30 (1).

[6] F. Cendes, F. Andermann, P. Gloor, A. Evans, M. Jones-Gotman, C. Watson, D. Melanson, A. Olivier, T. Peters, I. Lopes-Cendes, MRI volumetric measurement of amygdala and hippocampus in temporal lobe epilepsy, Neurology 43 (4) (1993) 719.

[7] P. Barta, G. Pearlson, R. Powers, S. Richards, L. Tune, Auditory hallucinations and smaller superior temporal gyral volume in schizophrenia, American Journal of Psychiatry 147 (11) (1990) 1457–1462.

[8] G. Moonis, J. Liu, J. Udupa, D. Hackney, Estimation of tumor volume with fuzzy-connectedness segmentation of MR images, American Journal of Neuroradiology 23 (3) (2002) 356.

[9] D. Chan, N. Fox, R. Scahill, W. Crum, J. Whitwell, G. Leschziner, A. Rossor, J. Stevens, L. Cipolotti, M. Rossor, Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease, Annals of Neurology 49 (4) (2001) 433–442.

[10] M. Rocca, A. Ceccarelli, A. Falini, B. Colombo, P. Tortorella, L. Bernasconi, G. Comi, G. Scotti, M. Filippi, Brain gray matter changes in migraine patients with T2-visible lesions: a 3-T MRI study, Stroke 37 (7) (2006) 1765.

[11] A. Zijdenbos, B. Dawant, Brain segmentation and white matter lesion detection in MR images, CRBE 22 (5-6) (1994) 401–65.

[12] V. Khoo, D. Dearnaley, D. Finnigan, A. Padhani, S. Tanner, M. Leach, Magnetic resonance imaging (MRI): considerations and applications in radiotherapy treatment planning., Radiother Oncol 42 (1) (1997) 1–15.

[13] K. Nobuhara, G. Okugawa, T. Sugimoto, T. Minami, C. Tamagaki, K. Takase, Y. Saito, S. Sawada, T. Kinoshita, Frontal white matter anisotropy and symptom severity of late-life depression: a magnetic resonance diffusion tensor imaging study, Journal of Neurology, Neurosurgery & Psychiatry 77 (1) (2006) 120–122.

[14] L. Campbell, E. Daly, F. Toal, A. Stevens, R. Azuma, M. Catani, V. Ng, T. Van Amelsvoort, X. Chitnis, W. Cutter, et al., Brain and behaviour in children with 22q11. 2 deletion syndrome: a volumetric and voxel-based morphometry MRI study, Brain 129 (5) (2006) 1218.

[15] M. Vernooij, M. Ikram, H. Tanghe, A. Vincent, A. Hofman, G. Krestin, W. Niessen, M. Breteler, A. van der Lugt, Incidental findings on brain MRI in the general population, New England Journal of Medicine 357 (18) (2007) 1821.

[16] D. Shattuck, S. Sandor-Leahy, K. Schaper, D. Rottenberg, R. Leahy, Magnetic resonance image tissue classification using a partial volume model, NeuroImage 13 (5) (2001) 856–876.

[17] Y. Zhang, M. Brady, S. Smith, Segmentation of brain MR images through a hidden markov randomfield model and the expectation-maximization algorithm, IEEE TMI 20 (1) (2001) 45–57.

[18] L. Dade, F. Gao, N. Kovacevic, P. Roy, C. Rockel, C. O'Toole, N. Lobaugh, A. Feinstein, B. Levine, S. Black, Semiautomatic brain region extraction: a method of parcellating brain regions from structural magnetic resonance images, Neuroimage 22 (4) (2004) 1492–1502.

[19] J. Ashburner, K. Friston, Unified segmentation, Neuroimage 26 (3) (2005) 839–51.

[20] J. Sled, A. Zijdenbos, A. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, IEEE TMI 17 (1) (1998) 87–97.

[21] L. Zhou, Y. Zhu, C. Bergot, A. Laval-Jeantet, V. Bousson, J. Laredo, M. Laval-Jeantet, A method of radio-frequency inhomogeneity correction for brain tissue segmentation in MRI., CMIG 25 (5) (2001) 379.

[22] E. Lewis, N. Fox, Correction of differential intensity inhomogeneity in longitudinal MR images, Neuroimage 23 (1) (2004) 75–83.

[23] A. Dale, B. Fischl, M. Sereno, Cortical surface-based analysis: I. Segmentation and surface reconstruction, Neuroimage 9 (2) (1999) 179–194.

[24] Y. Zhuge, J. Udupa, J. Liu, P. Saha, T. Iwanage, Scale-based method for correcting background intensity variation in acquired images, in: SPIE, Vol. 4684, 2002, p. 1103.

[25] J. Milles, Y. Zhu, G. Gimenez, C. Guttmann, I. Magnin, MRI intensity nonuniformity correction using simultaneously spatial and gray-level histogram information, CMIG 31 (2) (2007) 81–90.

[26] M. Ahmed, S. Yamany, N. Mohamed, A. Farag, T. Moriarty, A modified fuzzy c-means algorithm for bias field estimation andsegmentation of MRI data, IEEE TMI 21 (3) (2002) 193–199.

[27] S. Kim, S. Ng, G. McLachlan, D. Wang, B. Lovell, Segmentation of brain MR images with bias field correction (2003).

[28] R. Bansal, L. Staib, B. Peterson, Correcting nonuniformities in MRI intensities using entropy minimization based on an elastic model, Lecture Notes in Computer Science (2004) 78–86.

[29] E. Learned-Miller, P. Ahammad, Joint MRI bias removal using entropy minimization across images, in: Neural Information Processing Systems, Vol. 17, 2005, pp. 761–768.

[30] U. Vovk, F. Pernuš, B. Likar, Intensity inhomogeneity correction of multispectral MR images, Neuroimage 32 (1) (2006) 54–61.

[31] C. Lee, S. Huh, T. Ketter, M. Unser, Unsupervised connectivity-based thresholding segmentation of midsagittal brain MR images, Computers in Biology and Medicine 28 (3) (1998) 309–338.

[32] V. Grau, A. Mewes, M. Alcaniz, R. Kikinis, S. Warfield, Improved watershed transform for medical image segmentation using prior information, IEEE TMI 23 (4) (2004) 447–458.

[33] S. Warfield, M. Kaus, F. Jolesz, R. Kikinis, Adaptive, template moderated, spatially varying statistical classification, MIA 4 (1) (2000) 43–55.

[34] W. Wells III, W. Grimson, R. Kikinis, F. Jolesz, Adaptive segmentation of MRI data, IEEE TMI 15 (4) (1996) 429–442.

[35] Z. Song, S. Awate, J. Gee, Nonparametric Markov priors for tissue segmentation, in: IEEE ISBI, 2008, pp. 73–76.

[36] C. Cocosco, A. Zijdenbos, A. Evans, A fully automatic and robust brain MRI tissue classification method, MIA 7 (4) (2003) 513–527.

[37] J. Sijbers, P. Scheunders, M. Verhoye, A. Van der Linden, D. Van Dyck, E. Raman, Watershed-based segmentation of 3D MR data for volume quantization, Magnetic Resonance Imaging 15 (6) (1997) 679–688.

[38] K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens, Automated model-based tissue classification of MR images of thebrain, IEEE TMI 18 (10) (1999) 897–908.

[39] A. Zijdenbos, R. Forghani, A. Evans, Automatic" pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis, IEEE TMI 21 (10) (2002) 1280–1291.

[40] J. Yang, L. Staib, J. Duncan, Neighbor-constrained segmentation with level set based 3-D deformable models, IEEE TMI 23 (8) (2004) 940–948.

[41] K. Pohl, J. Fisher, R. Kikinis, W. Grimson, W. Wells, Shape based segmentation of anatomical structures in magnetic resonance images, ICCV.

[42] M. Prastawa, J. Gilmore, W. Lin, G. Gerig, Automatic segmentation of MR images of the developing newborn brain, MIA 9 (5) (2005) 457–66.

[43] S. Awate, T. Tasdizen, N. Foster, R. Whitaker, Adaptive markov modeling for mutual-information-based, unsupervised MRI brain-tissue classification, MIA 10 (5) (2006) 726–39.

[44] U. Yoon, V. Fonov, D. Perusse, A. Evans, The effect of template choice on morphometric analysis of pediatric brain data, Neuroimage.

[45] K. Nakamura, E. Fisher, Segmentation of brain magnetic resonance images for measurement of gray matter atrophy in multiple sclerosis patients, Neuroimage 44 (3) (2009) 769–776.

[46] D. Pham, C. Xu, J. Prince, A survey of current methods in medical image segmentation, Annual Review of Biomedical Engineering 2 (2000) 315–337.

[47] B. Belaroussi, J. Milles, S. Carme, Y. Zhu, H. Benoit-Cattin, Intensity non-uniformity correction in MRI: Existing methods and their validation, MIA 10 (2) (2006) 234–246.

[48] U. Vovk, F. Pernus, B. Likar, A review of methods for correction of intensity inhomogeneity in MRI, IEEE TMI 26 (3) (2007) 405–421.

[49] S. Bouix, M. Martin-Fernandez, L. Ungar, M. Nakamura, M. Koo, R. McCarley, M. Shenton, On evaluating brain tissue classifiers without a ground truth, Neuroimage 36 (4) (2007) 1207–1224.

[50] P. Teo, G. Sapiro, B. Wandell, Creating connected representations of cortical gray matter forfunctional MRI visualization, IEEE TMI 16 (6) (1997) 852–863.

[51] L. Rocha, F. Cappabianco, A. Falcão, Data Clustering as an Optimum-Path Forest Problem with Applications in Image Analysis, International Journal of Imaging Systems and Technology 19 (2) (2009) 50–68.

[52] P. Miranda, A. Falcão, J. Udupa, Cloud bank: A multiple clouds model and its use in MR brain image segmentation, in: IEEE ISBI, 2009, pp. 506–509.

[53] A. Falcão, F. Bergo, Interactive volume segmentation with differential image foresting transforms, IEEE TMI 23 (9) (2004) 1100–1108.

[54] S. Sandor, R. Leahy, Surface-based labeling of cortical anatomy using a deformable database, IEEE TMI 16 (1) (1997) 41–54.

[55] S. Smith, Fast robust automated brain extraction, Human Brain Mapping 17 (3).

[56] F. Segonne, A. Dale, E. Busa, M. Glessner, D. Salat, H. Hahn, B. Fischl, A hybrid approach to the skull stripping problem in MRI, Neuroimage 22 (3) (2004) 1060–1075.

[57] F. Cappabianco, A. Falcão, R. L.M., Clustering by optimum path forest and its application to automatic GM/WM classification in MR-T1 images of the brain, in: IEEE ISBI, 2008, pp. 428–431.

[58] J. Udupa, Y. Zhuge, Delineation operating characteristic (DOC) curve for assessing the accuracy behavior of image segmentation algorithms, in: SPIE, Vol. 5370, 2004, p. 640.

[59] S. Bricq, C. Collet, J. Armspach, Triplet markov chain for 3D MRI brain segmentation using a probabilistic atlas, IEEE ISBI (2006) 386–389.

[60] K. Pohl, S. Bouix, M. Nakamura, T. Rohlfing, R. McCarley, R. Kikinis, W. Grimson, M. Shenton, W. Wells, A hierarchical algorithm for MR brain image parcellation, IEEE TMI 26 (9) (2007) 1201–1212.

[61] D. Comaniciu, An algorithm for data-driven bandwidth selection, IEEE TPAMI 25 (2) (2003) 281–288.

[62] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE TPAMI 22 (8) (2000) 888–905.

[63] A. Falcão, J. Stolfi, R. Lotufo, The image foresting transform: theory, algorithms, and applications, IEEE TPAMI 26 (1) (2004) 19–29.

[64] D. Collins, A. Zijdenbos, V. Kollokian, J. Sled, N. Kabani, C. Holmes, A. Evans, Design and construction of a realistic digital brain phantom, IEEE TMI 17 (3) (1998) 463–468.

[65] J. Udupa, V. LeBlanc, Y. Zhuge, C. Imielinska, H. Schmidt, L. Currie, B. Hirsch, J. Woodburn, A framework for evaluating image segmentation algorithms, CMIG 30 (2) (2006) 75–87.

[66] D. Huttenlocher, G. Klanderman, W. Rucklidge, Comparing images using the Hausdorff distance, IEEE Transactions on pattern analysis and machine intelligence 15 (9) (1993) 850–863.