Improved estimates of partial volume coefficients from noisy brain MRI using spatial context

José V. Manjón a,⁎, Jussi Tohka b, Montserrat Robles a

a Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain
b Department of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101, Tampere, Finland

Article history:
Received 30 October 2009
Revised 15 June 2010
Accepted 17 June 2010
Available online 25 June 2010

Keywords:
Partial volume
MRI
Segmentation
Non-Local Filtering
Markov Random Fields

ABSTRACT

This paper addresses the problem of accurate voxel-level estimation of tissue proportions in the human brain magnetic resonance imaging (MRI). Due to the finite resolution of acquisition systems, MRI voxels can contain contributions from more than a single tissue type. The voxel-level estimation of this fractional content is known as partial volume coefficient estimation. In the present work, two new methods to calculate the partial volume coefficients under noisy conditions are introduced and compared with current similar methods. Concretely, a novel Markov Random Field model allowing sharp transitions between partial volume coefficients of neighbouring voxels and an advanced non-local means filtering technique are proposed to reduce the errors due to random noise in the partial volume coefficient estimation. In addition, a comparison was made to find out how the different methodologies affect the measurement of the brain tissue type volumes. Based on the obtained results, the main conclusions are that (1) both Markov Random Field modelling and non-local means filtering improved the partial volume coefficient estimation results, and (2) non-local means filtering was the better of the two strategies for partial volume coefficient estimation.

© 2010 Elsevier Inc. All rights reserved.

Introduction

Quantitative MRI is becoming an important tool for early diagnosis and follow-up of brain diseases giving new insights to the pathological substrate by supplying objective information which can be very useful for the clinicians. The estimation of the brain tissue type volumes and their spatial distributions play a major role in the understanding of the normal and pathological brain structure and function (Garcia-Martí et al., 2008; Mietchen and Gaser, 2009). This kind of measures are obtained by using MRI segmentation algorithms, which have been extensively developed during the last years (Ashburner and Friston, 2005; Tohka et al., 2010; Pham and Bazin, 2009; Brouwer et al., 2010).

Segmentation algorithms assign a label to each voxel in an image volume according to voxel’s membership to a discrete tissue class. However, in medical imaging a single voxel may be composed of a mixture of tissue types due to the finite spatial resolution of imaging devices. This phenomenon is known as partial volume effect (PVE) and it is mainly observable at tissue boundaries. PVE is an important factor in the study of small brain structures or highly convoluted brain regions such as those within cerebral cortex. In fact, ignoring this effect can lead to volume measurement errors in the range 20–60% (Gonzalez-Ballester et al., 2000).

Due to this effect, a number of methods have been proposed to estimate the Partial Volume Coefficients (PVC) within each voxel representing the proportion of the different tissues that contribute to value of such voxel. One of the first attempts to statistical partial volume (PV) modelling based on single channel MR data was proposed by Santiago and Gage (1993). They modelled pure tissues assuming equal variance Gaussian distributions (motivated by the MR physics) while PV voxel intensities followed a different distribution based on the parameters of the pure distributions. Laidlaw et al. (1998) used a similar model but incorporated a more complex method that took into account neighbourhood voxels. Ruan et al. (2000) concluded that PV voxel intensities can be generally modelled by using Gaussian distributions. However, this approach has the drawback of not being able to compute PV fractions, since there is no explicit relation between the intensity of PV voxels and the parameters of the related pure tissue types. In the same way, other authors have used more than 3 Gaussian classes to better fit intensity histograms. For example, Grawbowski et al. (2000) proposed to use directly 5 Gaussian distributions with different variances to model 3 pure tissues and two mixture tissues. Shattuck et al. (2001) extended the model of Santiago and Gage (1993) into a general segmentation framework where the brain voxels were first classified in 6 different classes (3 for pure tissues and 3 for mixture tissues) and thereafter the PVCs of the different pure tissues types were calculated.

The PVCs of the different pure tissues types were calculated.

Tohka et al.
Several factors affect the accuracy of statistically based PVC estimation. The accuracy of tissue type parameter estimation has a clear and demonstrated effect on the accuracy of PVC estimates (Tohka et al., 2004). The bias field correction also affects the PVC estimates by changing noise characteristics (Gonzalez-Ballester et al., 2000). Image noise severely degrades the accuracy of the PVC estimates due to the under-constrained nature of the estimation problem. The noise reduction method has to be chosen carefully in order not to degrade salient features of the image.

Markov Random Fields (MRFs) have been used to model spatial information in various ways in order to reduce the impact of noise during PVC estimation. Choi et al. (1991) and Li et al. (2005) used their approaches, based on multichannel MR data, a quadratic MRF to model the similarities between the PVCs of neighboring voxels. The use of quadratic MRF can be criticized because it tends to oversmooth the sharp transitions in PVCs between neighboring voxels. Also, the straightforward use of quadratic MRFs is limited to 2-channel images in the three tissue class case (Kent and Mardia, 1998). In their two-stage procedures, Shattuck et al. (2001) and Tohka et al. (2004) applied an extended Potts model based MRF during the voxel classification stage, but choose not to apply spatial information during the second, PVC estimation stage. A modified version of this MRF was proposed in Kim et al. (2005). Several authors have suggested to discretize the PVC estimation, i.e. assume that only a discrete set of values for PVCs are possible (e.g. (0.0,0.1,0.2,...,1)) (Van Leemput et al., 2003; Ruan et al., 2002; Bricq et al., 2008). The challenge here is to set the discretization interval correctly. Pham (2001) has considered different spatial models for fuzzy C-means clustering. However, the fuzzy C-means objective follows the statistical PV model only in limited number of special cases (Pham and Prince, 1998). Shattuck et al. (2001) and Van Leemput et al. (2003) presented evidence that the MRF-modeling improves the accuracy of the PVC estimates.

Image filtering is another method to reduce the noise level and improve the accuracy of the PVC estimation. Several random noise filters in the context of MRI have been proposed in the literature. These include the Anisotropic Diffusion Filter (Gerig et al., 1992), Wavelet filter (Pizurica et al., 2003) and recently the Non-local means filter (Coupé et al., 2008). The effects of filtering to PVC estimation are not well understood. Li et al. (2005) compared the use of anisotropic diffusion filter and their MRF based approach to reduce the effects of random noise. They found that filtering improved the segmentation accuracy in one case (the study sample consisted of two simulated images). However, they compared only crisp segmentations, not the PVC estimates.

Before the PVC estimation can take place, the intensity characteristics of the pure tissue types have to be known. An interesting alternative to modelling the contribution of the PV voxels in image histograms (Santago and Gage, 1993; Tohka et al., 2007) is to discard them from the pure tissue parameter estimation process. In this sense, Shattuck et al. (2001) proposed to use a set of heuristics and Tohka et al. (2004) proposed the use of robust statistics to overcome the PVC problem. In the latter work, two methods known as Minimum Covariance Determinant (MCD) and Minimum Volume Ellipsoid (MVE) were used over a set of pre-classified and trimmed data. Recently, an extension of this method has been developed (Manjón et al., 2008) that improves the tissue parameter estimation step by applying a new trimming procedure based on local gradient information to select pure tissue samples from the data.

The main aim of the present work is to combine robust parameter estimation methods (Tohka et al., 2004; Manjón et al., 2008) to strategies to reduce random noise present in MR images to formulate an accurate method to estimate the PVCs. Concretely, to reduce the contribution of image noise, the applications of a novel MRF type strategy capable to overcome the limitations of the MRF-type methods described above and a state-of-the-art adaptive filtering method based on non-local means modelling are studied and compared. We will show in our experiments that both the MRF-based technique and non-local means filtering improve the PVC estimation results. Further, the non-local means filtering was the more effective technique in our experiments.

**Methods**

In MRI brain segmentation, after a brain extraction operation, the intracranial cavity is supposed to be composed mainly of 3 tissue types, which are white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF). However, brain tissues exhibit natural variations in their composition along the brain and in the tissue boundaries due to PVE. To reflect such heterogeneity some authors have used the *mixel* model that reflects the intensity emitted by a voxel as a linear combination of the mean values of each tissue contained in the voxel weighted by their relative content within the voxel (Choi et al., 1991). Therefore, the measured image *y* can be modelled as follows:

\[
y_i = \Phi (\mu_i, \alpha_i) \beta_i + \epsilon_i
\]

\[
\Phi (\mu_i, \alpha_i) = \sum_{j=1}^{3} \alpha_{ij} \mu_{ij} \sum_{j=1}^{3} \alpha_{ij} = 1, \alpha \in [0,1]
\]

where \( \Phi (\mu_i, \alpha_i) \) is a linear mixing function that gives the intensity of a voxel based on a linear combination of the different tissue means \( \mu_i \) and partial volume coefficients (PVCs) \( \alpha_{ij} \) (relative proportions of each tissue *j* on the voxel *i*), \( \beta_i \) is a multiplicative low frequency noise commonly known as bias field (Sled et al., 1998; Manjón et al., 2007) and \( \epsilon \) is the random noise from the measuring process. In magnitude MR images, \( \epsilon \) follows a Rician distribution (Sijbers, 1998). Note that, in the absence of spatial constraints, the estimation of the proportions of three tissue types within a voxel requires data from at least two acquisitions with different pulse sequences.

Some authors include a tissue dependent noise to reflect the anatomic variability within the tissue (Tohka et al., 2004; Tohka et al., 2007; Prima et al., 2001). We make the assumption that this variability can be considered as a special case of PVE as done in other works (Santago and Gage, 1993; Laidlaw et al., 1998).

Since the combination of more than two tissues in a voxel is very rare for the resolution used in the clinical practice, a simpler model can be used by assuming that only two tissue types can be present in a voxel. This simplification has been previously used in other works (Santago and Gage, 1993; Laidlaw et al., 1998; Ruan et al., 2000) and it allows for the estimation of tissue proportions in a voxel based on a single MR volume. Then, the simplified \( \Phi \) function is:

\[
\Phi (\mu_i, \alpha_i) = \alpha_{ij} \mu_j + (1-\alpha_{ij}) \mu_h
\]

where the tissue type different from \( j \) and \( h \) have a zero proportion.

The first step—termed PV classification in Tohka et al. (2004)—in applying the model (3) is solving which tissue types are present in each voxel. At least two approaches to solve the tissue types present in a voxel have been applied in a single channel case. The first one (termed naive in this work) assigns tissue types \( j \) and \( h \) with successive mean values to a voxel \( i \) if \( y_i \in [\mu_j, \mu_h] \) (Brouwer et al., 2010). The second approach is based on the Bayes classifier (Duda et al., 2000). The class conditional densities for pure tissue classes are modelled by the normal density and for the partial volume classes they are constructed by using a marginalization technique developed originally in Santago and Gage (1993). This technique was applied in Shattuck et al. (2001) and Tohka et al. (2004) in a combination with...
an MRF-based prior which can be seen as an extension of the Potts model. In this work, we call this method as MRF-based PV classification. Clearly, methods for PVC estimation rely strongly on the correctness of the PV classification. It is also worth noting that no CSF/WM voxels are allowed in the described classification. However, since the number of voxels falling into this class is small, this assumption has only a minimal effect.

Then, assuming that intensity inhomogeneities and random noise have been previously minimized by using specific filters, the relative proportion of each tissue on each voxel can be derived from Eq. (3) using the maximum likelihood principle (assuming equal tissue type variances). If we assume that voxel \(i\) is a mixture of tissue types \(j\) and \(h\) then:

\[
\alpha_{ij} = \frac{|y_i - \mu_j|}{|y_i - \mu_h|}
\]

and \(\alpha_{ik} = 0\) for \(k \neq j, h\). In the special case where \(y_i\) is less than the smaller mean value \(\mu_j\) or the higher mean value, the proportions are assigned to a 100% tissue content of their corresponding tissue. This second step is termed PVC estimation in accordance to Tohka et al. (2004).

**Minimization of random noise effect on PVC estimation**

To reduce the effect of random noise in the PVC estimation, two main alternatives can be chosen, spatial modelling and filtering. In the present work, both approaches are compared to discover the best choice for the PVC estimation.

**Spatial modelling**

Modelling the spatial dependencies among PVC estimates of neighbouring voxels is one way to cope with the image noise during the PVC estimation. Markov Random Fields are a widely applied technique for such modelling. However, as described in Introduction, most MRF based techniques applied to PVC estimation have limitations. We therefore propose a new MRF-like method to aid PV estimation step; note that this MRF will be different from the Potts model applied during the PV classification (that has been described in Tohka et al., 2004). The details of the method are described in the appendix A due to the complexity of its mathematical description and only a short summary is presented in this section.

This new MRF-based method combines the information from the PVC classification step to the prior information on the spatial layout of PVCs in the PVC estimation step. The approach also allows for sharp transitions of PVCs near the tissue type boundaries. The method is formulated as a penalized likelihood optimization, that is, we aim to minimize a certain penalized likelihood criterion that is a weighted form of the likelihood. The method does not penalize the contribution of a voxel according to its geometrical distance of the voxel to the voxel being filtered, as for example occurs with the bilateral filter (Tomasi and Manduchi, 1998). Although the method can use the whole image to find denoised estimate for every single voxel intensity in the image, only a local region \(\Omega\) surrounding each voxel is processed for the estimate of the voxel intensity due to computational reasons.

Given an image \(Y\), the filtered value at a point \(p\) using the NLM method is calculated as a weighted average of all the voxels in a given neighbourhood \(\Omega\) defined as a 3D region surrounding the voxel being filtered:

\[
\text{NLM}(Y(p)) = \sum_{q \in \Omega} w(p,q) Y(q)
\]

where the weights are inversely proportional to the intensity distance of the patches surrounding the voxels being compared:

\[
w(p,q) = \frac{1}{Z(p)} e^{-\frac{\|N_p - N_q\|^2}{\nu}}
\]

where \(Z(p)\) is normalization constant:

\[
Z(p) = \sum_{q \in \Omega} e^{-\frac{\|N_p - N_q\|^2}{\nu}} d(N_p, N_q) = \|N_p - N_q\|^2
\]

where \(N_p\) denotes a vector of the intensity values in the 3D neighbourhood of the voxel \(Y(p)\). The parameter \(\nu\) controls the strength of the filtering.

In Eq. (6) there is a special case when \(p = q\). As the self similarity is very high, it can produce an over-weighting effect. To solve this situation Buades suggested to calculate \(w(p,q)\) as the maximum \(w(p,q)\) for all \(q \neq p \in \Omega\). A detailed description of the method can be found in Coupé et al. (2008).

In the present work, we used a 3D voxel-wise version of the Non-local means method as proposed by Coupé et al. (2008) with two modifications which allow reducing the computation time by 25% while producing slightly better results.

First, the weight computation has been made symmetric \((w(p,q) = w(q,p))\) and therefore we have not to calculate each weight twice resulting in a speed up factor of two. In the original NLM method \(w(p,p)\) was set to the maximum weight in the neighbourhood \(\Omega\). However this solution has the disadvantage of making the weights not symmetric since \(w(p,p)\) depends on the neighbourhood \(\Omega\) which is different for \(p\) and \(q\). To avoid this problem, we propose a new distance measure that overcomes the overweighing effect of the original formulation and makes unnecessary to calculate the self distance in function of the rest of the distances of the local neighbourhood.

\[
d(N_p, N_q) = \max(0, \|N_p - N_q\|^2 - 2\sigma^2)
\]

This modification takes in consideration the fact that the distance of two exactly equal patches with different instances of noise is approximately equal to \(2\sigma^2\). Therefore, this minimizes the over-weighting effect of \(w(p,p)\) which can be set to its theoretical optimum

\[
\alpha_{ij} = |y_i - \mu_j| / |y_i - \mu_h|
\]
value (i.e. \( w(p,n) = 1 \)). This makes the weight calculation independent of the search area and symmetric by definition.

Second, in contrast with the preselection proposed by Coupé et al. (2008) based on the first and second order image moments, we use a statistically driven approach based on the distance of the mean values of the 3D patches. We propose to preselect all those voxels with a difference of their first local moment (mean value of a \( 3 \times 3 \times 3 \) image patch) smaller than the \( 3\sigma/\sqrt{n} \) which corresponds to the third quantile of a standard normal distribution (\( \sigma \) is the noise standard deviation in the volume and \( n \) the number of pixels used to compute the mean, 27 in our case) in each volume:

\[
w(p,q) = \begin{cases} 
1 & \text{if } \left| \mu_p - \mu_q \right| < 3\sigma/\sqrt{n} \\
0 & \text{otherwise}
\end{cases} \quad (9)
\]

Patches with mean value differences higher than this threshold have a very small probability to be similar to the current patch. In all the experiments described in this paper, the filter was applied with its default parameters, search volume size \( \Omega = [11 \times 11 \times 11] \), patch size \( 3 \times 3 \times 3 \) and \( h^2 = 2\sigma^2 \) (see (Coupé et al., 2008) for further details).

**Compared methods**

Several methods combining pre-processing, parameter estimation, PV classification and PVC estimation were compared with the state of art methods.

As reference, we used the Trimmed Minimum Covariance Determinant method (TMCD) (Tohka et al., 2004) and the Trimmed Mean Segmentation (TMS) method (Manjón et al., 2008). The TMS method does not supply PVCs but the mean intensities of different brain tissue types. Here, after the tissue parameter estimation using the TMS method, the PVC estimation was performed using Eq. (4) and assuming naive PV classification. The TMCD method was exactly as in Tohka et al. (2004) except the initial segmentation was performed by the genetic algorithm-based finite mixture model (Tohka et al., 2007) instead of the neural network approach of (Tohka et al., 2004). The TMCD method combines a robust estimation based tissue type parameter estimation, an MRF-based PV classification into three pure and three mixed tissue types, and maximum likelihood based PVC estimation which can assume different variances for pure tissue types. The TMCD parameter estimation is based on the estimation of brain tissue type parameters (mean and variance) from a trimmed subset of intensities using the robust Minimum Covariance Determinant estimator. The original set of intensities is obtained based on a rough segmentation of the image into the three principal tissue classes. The TMS parameter estimation method is an extension of the TMCD that improves the estimation of the mean value of each tissue type by excluding from the estimation process voxels with a high local gradient (as they are likely to be PV affected voxels). Further details about TMCD and TMS methods can be found in Tohka (2004) and Manjón (2008).

As new methods, we tested the following combinations (see Table 1):

- **TMS + Filtering**: TMS method over a previously denoised volumes using the described non-local means filter.
- **TMS + TMCD**: TMS for parameter estimation and TMCD for classification.
- **TMS + TMCD + Filtering**: TMS for parameter estimation and TMCD for classification over a previously denoised volumes using the described non-local means filter.
- **TMS + TMCD + MRF**: TMS for parameter estimation, TMCD for classification and the new 3D Markov Random field approach described in this paper for PVC estimation.

**Implementation details**

All the experiments described in this paper were performed using MATLAB Environment (version 2007a) in a normal PC (Intel Core 2 Quad with 8 Gb of RAM). Due to the high computational burden of the NLM filter this was implemented using a multithreading approach while the rest of the methods were single thread.

The processing times for each of the methods were 3 min for the PV classification, 11 min for the MRF-based PVC estimation and 5 s for the tissue parameter estimation. The optimized NLM filter proposed in this paper takes 18 min to perform the filtering. All the methods were run with their default parameters.

**Experiments and results**

Experiments with synthetic brain MR data were performed to evaluate the accuracy and robustness of the different approaches. Results were compared with those obtained with SPM5 software (http://www.fil.ion.ucl.ac.uk/spm/) as it has become one of the most used methods in recent bibliography (Ashburner and Friston, 2005). This software produces the tissue maps corresponding to the probability of observing different tissue types in each voxel. Note that fuzzy segmentations in SPM5 do not model the tissue fractions but the tissue type likelihoods—the probability of observing a certain tissue type at a voxel. Therefore, the results of SPM5 should be taken as an indicator how the incorrect model choice affects the results. A more detailed analysis of model differences is presented in Appendix B. This analysis does not concern only SPM5, but also any other tissue classification method producing tissue type likelihoods as output.

To simplify the analysis of the results, we will consider MR intensity uniform volumes. Intensity inhomogeneity correction can be performed by applying specific filters (Sled et al., 1998; Belaroussi et al., 2006; Manjón et al., 2007).

**Original BrainWeb phantoms**

The proposed methods were evaluated with synthetic data using the Montreal Neurological Institute (MNI) BrainWeb phantom (Kwan et al., 1999; Collins et al., 1998). The input for these simulations is a fuzzy realistic brain phantom (Collins et al., 1998) as opposed to a discrete phantom, which makes the evaluation of PVC estimation...
algorithms possible. A skull stripped (only CSF, GM and WM labelled voxels) T1-weighted BrainWeb phantom of 1 mm³ voxel size with no inhomogeneity was used with different levels of Rician distributed random noise (0%, 1%, 3%, 5%, 7% and 9% of the maximum image intensity). The noiseless phantoms were considered to be complex valued with the imaginary part equal to zero for the generation of Rician noise. Noise was generated by adding Gaussian noise to the real and imaginary parts and then computing the complex modulus, thus forming a magnitude image.

A comparison of the proposed NLM based filtering approach with Coupe’s method is presented in Fig. 1 where both filters were applied to the T1-w phantom for different levels of random noise. As can be noted using the Mean Absolute Difference (MAD) measure, the proposed method slightly improves Coupe’s method but while running a 25% faster.

The PVC estimation experimental results were evaluated using root mean square error (RMSE) of each tissue type:

\[ RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^{M} (E(i) - T(i))^2} \]  

where \( E(i) \) is the estimated PVC for a specific tissue type at the voxel \( i \), \( T(i) \) is the true PVC value at the voxel \( i \) and \( M \) is the total number of brain voxels. The global RMSE was obtained by averaging the tissue type specific RMSE.

In Fig. 2, the RMSE of PVCs for each tissue type and averaged RMSE are presented for all compared methods. Example PVC maps of the compared methods are presented in Fig. 3 for a high noise level.
The different tissue volumes were calculated from the PVCs obtained with the different methods by integration to find out their effect in the measurement of the different tissue volumes. Volume error (in mm$^3$) for each tissue type $k$ was calculated as the real volume of this tissue type minus the estimated volume:

$$VE(k) = \sum_{i=1}^{M} T_k(i) - \sum_{i=1}^{M} E_k(i)$$ (11)

Normalized brain Volume Error (in %) was also computed as the absolute VE of the brain (grey and white matter) divided by the true brain volume:

$$NVE = 100 \left( \frac{\sum_{i=1}^{M} (T_g(i) + T_w(i)) - \sum_{i=1}^{M} (E_g(i) + E_w(i))}{\sum_{i=1}^{M} (T_g(i) + T_w(i))} \right)$$ (12)

Results of such measurements for the different levels of noise applied and with the different methods are shown in Fig. 4.

**New BrainWeb Phantoms**

As no PVC error estimation can be performed on real clinical data due to the lack of the gold standard, a PVC error estimation was performed over the 20 example brain images supplied by the MNI (Aubert-Broche et al., 2006) to evaluate the effectiveness of the different methods over different anatomies. In our experiments, we used 20 noisy (2% of Rician noise) T1-weighted simulated volumes representing a spoiled FLASH sequence with $TR = 22$ ms, $TE = 9.2$ ms, flip angle = 30 deg and 1 mm$^3$ isotropic voxel size) available from the BrainWeb site (http://mouldy.bic.mni.mcgill.ca/brainweb/anatomic_normal_20.html).

The corresponding 20 PVC maps (0.5 mm $\times$ 0.5 mm $\times$ 0.5 mm) were resampled to 1 mm$^3$ by averaging the 8 true PVC values within a 1 mm $\times$ 1 mm $\times$ 1 mm voxel to match with the resolution of the simulated images (1 mm$^3$). The PVC estimation was performed with the compared methods. Results are summarized in Table 2 and Figs. 5 and 6.

**Discussion**

In the present work, novel methods for PVCs estimation have been introduced and compared with existing state of the art methods. The new methods combine either non-local means filter or an MRF-based spatial modelling of PVCs to a recently introduced MRI tissue type parameter estimation method (TMS, Manjón et al., 2008). Based on the obtained results it can be concluded that the non-local means filtering based methods jointly with the TMS parameter estimation provided the best results in the comparative and therefore seemed to be the most suitable for PVC estimation.

Concretely, for the original BrainWeb data in Original BrainWeb phantoms, the TMS + Filtering method obtained the best PVC estimation results in all the cases closely followed by the TMS + TMCD + Filtering method. The methods differed in how the PV classification step was performed. Therefore, with these data, the more advanced MRF-based PV classification was not helpful. Instead, the TMCD method (Tohka et al., 2004) obtained better results than the TMS method. This can be attributed largely to the advantages of MRF-based PV classification over the naive PV classification when working with noisy images. Surprisingly, the TMCD method also outperformed the TMS + TMCD method. The main difference between the two methods is in the tissue type parameter estimation and TMS has been demonstrated to provide more accurate estimations of tissue type means of the two (see (Manjón et al., 2008) where the two parameter estimation methods were compared). The superiority of TMCD over TMS + TMCD in PVC estimation should be attributed to a more accurate tissue type variance estimation in a combination with an advanced PV classification. The TMS + TMCD + MRF method obtained better results than the other methods not based on filtered data but the results were still worse than with filtering based methods. Also, the results of SPM5 illustrate that the distinction between PVC estimation and the tissue likelihood estimation is important also in practise and the tissue likelihoods of the SPM5 should not be interpreted as tissue fractions. The converse is also true, i.e. the PVC...
estimates are not tissue type likelihoods. In the case of the new 20 brain phantoms, again, filtering based methods (TMS + TMCD + Filtering and TMS + Filtering) obtained the best estimates in all the cases although in this case TMS + TMCD + Filtering was the best method (see New BrainWeb phantoms). Otherwise, the ranks of the methods were the same as with the original BrainWeb data.

Our proposed methods based on filtering clearly outperformed previous proposed methods evaluated using the original BrainWeb phantom (Shattuck et al., 2001; Tohka et al., 2004); the latter is the TMCD method of this work. In the case of Shattuck method, we are aware that they evaluated the whole segmentation process including non-uniformity correction so the comparison can be only partial. Recent work from Brouwer et al. (2010) shows comparable RMSE-results to our methods (using classical BrainWeb phantom) for low noise levels (3%) (RMSE was 0.093 for GM, 0.077 for WM and 0.052 for CSF). The corresponding RMSEs for our TMS + Filtering method were 0.098 for GM, 0.072 for WM and 0.093 for CSF. However, our methods show clearly improved results for medium (5%) noise levels for GM (RMSE was 0.112 for TMS + Filtering vs. 0.143 by Brouwer et al.), WM (RMSE was 0.084 for TMS + Filtering vs. 0.127 by Brouwer et al.) and slightly worse for CSF (RMSE was 0.099 for TMS + Filtering vs. 0.068 by Brouwer et al.). Brouwer et al. (2010) provided no results for higher noise levels. Also, their MRI simulations using 20 new phantoms (0.5 mm³ voxel size with 1% noise level) were different from the ones provided in the BrainWeb site and described in Aubert-Broche et al. (2006) and thus no direct comparison to our results can be made with these data.

In the present work, we used RMSE on the PVC estimates as the figure of merit. This measure was chosen because it has been used in

<table>
<thead>
<tr>
<th>Methods</th>
<th>CSF</th>
<th>GM</th>
<th>WM</th>
<th>Global</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMS</td>
<td>0.152 ± 0.001</td>
<td>0.145 ± 0.005</td>
<td>0.120 ± 0.004</td>
<td>0.140 ± 0.004</td>
</tr>
<tr>
<td>TMS + Filtering</td>
<td>0.149 ± 0.003</td>
<td>0.115 ± 0.006</td>
<td>0.086 ± 0.005</td>
<td>0.119 ± 0.004</td>
</tr>
<tr>
<td>SPM5</td>
<td>0.225 ± 0.020</td>
<td>0.185 ± 0.007</td>
<td>0.128 ± 0.009</td>
<td>0.184 ± 0.009</td>
</tr>
<tr>
<td>TMCD</td>
<td>0.140 ± 0.003</td>
<td>0.135 ± 0.003</td>
<td>0.107 ± 0.006</td>
<td>0.128 ± 0.004</td>
</tr>
<tr>
<td>TMS + TMCD</td>
<td>0.139 ± 0.003</td>
<td>0.138 ± 0.005</td>
<td>0.110 ± 0.005</td>
<td>0.130 ± 0.004</td>
</tr>
<tr>
<td>TMS + TMCD + Filtering</td>
<td>0.136 ± 0.004</td>
<td>0.116 ± 0.005</td>
<td>0.097 ± 0.005</td>
<td>0.115 ± 0.004</td>
</tr>
<tr>
<td>TMS + TMCD + MRF</td>
<td>0.139 ± 0.004</td>
<td>0.133 ± 0.005</td>
<td>0.104 ± 0.005</td>
<td>0.126 ± 0.004</td>
</tr>
</tbody>
</table>

Fig. 4. Volume measurement errors for the different methods compared.
several studies cited above to evaluate the accuracy of PVC estimation. The PVC estimates can be used to classify each voxel to the most representative tissue type (CSF, GM, or WM), for example by selecting the tissue type with the maximal PVC to represent the voxel. In MRI segmentation literature, the quality of these kind of hard segmentations is typically measured using Dice or Jaccard coefficients. However, we were interested about the accuracy of PVC estimates, not the accuracy of hard segmentations based on them. The PVCs have themselves several applications—including voxel based morphometry—in which the conversion to hard labels is not desired. In addition, this conversion is usually based on a simple maximal PVC criterion described above (see e.g. Tohka et al., 2004), which may be suboptimal. Indeed, if the hard segmentation is required, it can be preferable to directly estimate the hard segmentation instead of PVC estimates. This was evident in the experimental results in Tohka et al. (2010) where the issue is further discussed. In other words, the quality of the PVC estimates cannot be evaluated based on the quality of the hard segmentation generated based on these PVC estimates.

The effect of the different methods for PVC estimation in brain tissue volumetry was explored in this work. These experiments showed that most of the methods tended to overestimate CSF volume (especially the SPM5 method) with the exception TMS method. Regarding the brain volume error, TMS method showed the lower error for noise levels below 5% while the TMS + TMCD method produced the best results for higher noise levels. This can be explained by the fact that TMCD is able to produce a better classification of brain voxels for high noise levels. While the performance of the TMS method degraded with the noise level as expected due to voxel misclassification and the asymmetry of the Rician noise distribution, the application of the NLM filter highly improved the

![Fig. 5. RMSE of PVCs of the MNI 20 example brains calculated with the different methods. Top-Left: Global RMSE for all tissue types. Top-right: RMSE for cerebral spinal fluid tissue type. Bottom-left: RMSE for gray matter tissue type. Bottom-right: RMSE for white matter.](image)

![Fig. 6. Normalized Brain Volume Error (%). Note that the TMS method was the only method that provided unbiased estimates. The rest of the methods also show good results except the SPM5 method which exhibited the higher variability.](image)
The spatial PVC content will be taken into account, and more specifically we consider triplets such as \((i, ix-, ix+)\), \((i, iy-, iy+)\) and so forth, although we can conclude that the TMS+ filtering method seems to be the method of choice for most common research and clinical settings.

Acknowledgments

This work has been partially supported by the Spanish Health Institute Carlos III through the RETICS Combiomed, RD07/0067/2001. J. Tohka’s work was funded in part by the Academy of Finland (application numbers 130275 and 129657, Finnish Programme for Centres of Excellence in Research 2006–2011).

Appendix A

In this Appendix, we will present the penalized likelihood method to estimate partial volume coefficients \( \alpha_k \) based on the image intensities and the knowledge of the tissue types present in each voxel. For a notational convenience, we define \( \alpha = [\alpha_1, \ldots, \alpha_k]^T \) and \( A = \{a_i : i = 1, \ldots, M\} \). The context image \( C = [c_i : i = 1, \ldots, M, c_i \in \hat{I}] \) consisting of the labels of each voxel. Labels come from the set \( \hat{I} = \{1, \ldots, M\} \). Labels of the type \((j, k)\) indicate mixed voxels of the tissue types \(j\) and \(k\). Note that we allow maximum of two types of tissue per voxel. This is an important assumption for later developments. Define the negative of log-likelihood as

\[
I(A|C, Y) = (\gamma - \sum_{j=1}^3 \alpha_j k_j)^2 / \sigma^2,
\]

where \( Y = \{y_i : i = 1, \ldots, M\} \) are the image intensities, \( \mu_j, j=1,2,3 \), are the mean intensities of the tissue types and \( \sigma^2 \) is the variance of white Gaussian noise \( \epsilon \). Note that the minimization of the negative log-likelihood leads to the PVC estimates of Eq. (4). We introduce a spatial regularization term \( U(A|C) \) to deal with image noise. The PVC estimates are solved by minimizing a penalized likelihood criterion:

\[
A^* = \arg\min_{A} I(A|C, Y) + \lambda U(A|C)
\]

where \( \lambda \) is regularization parameter controlling the strength of the regularization term.

The rest of the appendix is concerned with the technical definition of \( U(A|C) \), which can be seen as the energy function of a Markov Random Field (MRF) and is meant to discourage PVC configurations that are unrealistic. This definition is presented in a concise manner and it uses terminology drawn from MRF literature. Readers unfamiliar with MRFs are referred to Li (2009). We consider a slightly unusual neighborhood system, where the voxel \( i = [ix, iy, iz] \) has 12 neighbors: \( ix-, ix+, iy-, iy+, iy-, iy+, iz-, iz+, iz-, iz+, \) where \( ix- = [ix-1, iy, iz], ix- = [ix-2, iy, iz], iz- = [ix, iy+1, iz], iz- = [ix, iy, iz] \) and similarly for others. The set of neighbors of \( i \) is denoted by \( N_i \). Now the cliques of the neighborhood system (with a non-zero clique potential value) consist of triplets \((i, ix-, ix+), (i, iy-, iy+)\) and so forth and \( U(A|C) \) is defined in terms of these clique potentials:

\[
U(A|C) = \sum_{i=1}^{N} [U_{ix}(\alpha_i, \alpha_{ix-} + \alpha_{ix+} + \alpha_{ix} + \alpha_{ix} \alpha_{ix-} \alpha_{ix+}] + U_{iy}(\alpha_i, \alpha_{iy-} + \alpha_{iy+} + \alpha_{iy} + \alpha_{iy} \alpha_{iy-} \alpha_{iy+}] + U_{iz}(\alpha_i, \alpha_{iz-} + \alpha_{iz+} + \alpha_{iz} + \alpha_{iz} \alpha_{iz-} \alpha_{iz+}]])
\]

This is a 3rd order MRF meaning that interactions consist of triplets of sites whereas typically in MRI segmentation 2nd order MRFs have been considered instead. Let now \( e_j \) be an unity vector whose \( j \)th element is one and all other elements are zero. Let \( \delta \) be the Kronecker’s delta function. Then,

\[
U_{ix}(\cdot) = \begin{cases} -\ln \delta(\alpha_i - e_j) & \text{if } c_i = j \\ \delta_{jk}(\alpha_i, \alpha_{ix-} + \alpha_{ix+} + \alpha_{ix}) & \text{if } c_i = k \end{cases}
\]

where \( \delta_{jk} \) will be defined momentarily, and similar definitions hold for \( U_{iy}, U_{iz} \). Above, we have used the convention that \( \ln(0) = -\infty \). The interpretation of this definition is that if the label of the voxel \( i \) is a pure label, say \( c_i = j \), then \( \alpha_i = e_j \). However, if the label \( c_i \) is a mixed label, then the spatial PVC content will be taken into account, and more specifically we consider triplets such as \( i, ix-, ix+ \).
Denote $D_{jk} = \{ j, k \}$. Define
\[
\Delta_{jk}(a, b) = \begin{cases} 1 & \text{if} \quad (a = \{ j, k \} \land b = D_{jk}) \lor (b = \{ j, k \} \land a = D_{jk}) \\ 0 & \text{otherwise} \end{cases}
\]

Now
\[
\phi_j(\alpha, \alpha_{x-\cdot} | C) = \Delta_j(c_{ix} + . . c_{ix-\cdot})(\alpha_{ix} - 0.5(\alpha_{ix-\cdot} + \alpha_{ix}))^2
\]
and $\phi_j(\alpha, \alpha_{x-\cdot} | C)$ and $\phi_j(\alpha, \alpha_{x\cdot-} | C)$ are defined in the same way. This means that $\phi_j(\alpha, \alpha_{x\cdot-} | C)$ has a nonzero value only if one of the $c_{ix+} , c_{ix}$ has the same mixed label as $c_j$. This is because if both $i x +$ and $i x -$ have a different label than $i$, their PV coefficients cannot be expected to provide any information on the PV coefficient $\alpha_i$. Moreover, we require that if only one of the $c_{ix+} , c_{ix}$ is equal to $c_j$, another must have a relevant pure label in order to $\phi_j(\alpha, \alpha_{x\cdot-} | C)$ be non-zero. Note also that if $c_i = \{ j, k \}$ and $c_{ix+ \cdot} , c_{ix\cdot} \in D_{jk}$, then
\[
(\alpha_{ix\cdot} - 0.5(\alpha_{ix-\cdot} + \alpha_{ix}))^2 = (\alpha_{ix\cdot} - 0.5(\alpha_{ix-\cdot} + \alpha_{ix-\cdot}))^2,
\]
and
\[
(\alpha_{ix\cdot} - 0.5(\alpha_{ix-\cdot} + \alpha_{ix}))^2 = 0 \quad \text{for} \quad i \neq D_{jk}
\]

Hence, it suffices to examine only the $jth$ element of the PV coefficients to compute $\phi_j$. As can be seen in Eq. (9), triplet interactions—instead of pairwise interactions—are a necessary feature of our MRF model due to the extra information contained in the PV labeling C. The particular neighborhood system was selected due to conceptual and computational simplicity. We consider only triplets whose embedding on the image is a straight line passing through the voxel of interest.

Unfortunately, it is impossible to find a closed form solution for $A^*$ and we have to resort to a numerical method for solving it. It is straightforward to derive an Iterative Conditional Modes (ICM)-like algorithm for the purpose, which we give without derivation (see Besag (1986) and Li (2009) for a more general discussion about ICM algorithms):

Define $r(x) = \{ x \}$ when $x \in [0, 1], r(x) = 0$ when $x < 0$, and $r(x) = 1$ when $x > 1$. For all $i$ with $c_i = \{ j, k \}$, set
\[
\alpha_{ij}^0 = r\left( \frac{h_{ik} - y_j}{h_{ik} - h_j} \right).
\]

For all $i$ with $c_i = \{ j, k \}$, iterate until convergence:
\[
\alpha_{ij}^t \leftarrow r\left( \frac{h_{ik} - y_j}{h_{ik} - h_j} \right) + \lambda \alpha_i^t \sum_{l \in N_{I,i} \cap N_{P,j}} \Delta_{jk}(c_i, c_{il}) (\alpha_{il}^t + \alpha_{il}^t) + \sum_{l \in N_{I,i} \cap N_{P,j}} \Delta_{jk}(c_i, c_{il}) (\alpha_{il}^t - 0.5 \alpha_{il}^t)
\]
\[
\alpha_{ij}^t \leftarrow r(\alpha_{ij}^t)
\]
where $t$ denotes the most current estimate available, and $N_{P,i} = \{ (ix+ix-, iy+iy-, iz+iz-) \}$ is the set of first order neighbor pairs and $N_{P,j} = \{ (ix+ix+, iy+iy+, iz+iz+) \}$ is the set of second order neighbor pairs. Note that the neighbor pairs in sets $N_{P,i}$ and $N_{P,j}$ are ordered. In our experiments, $\lambda$ was set to 10. This value was empirically determined by using a single BrainWeb image with 5% noise level as test material. Note that the second order pairs enter into the update Eq. (16) because the potential of the cliques of the type $(ix-, i, ix-), (ix+, i, ix+)$ and so forth depend on $\alpha_i$; see also Chapter 2 of (Li, 2009).

Appendix B

In this Appendix, we explain the difference between the PVC estimation and the estimation of tissue type probabilities akin to SPM5 (Ashburner and Friston, 2005) and various other methods in literature. The most important of these is the conceptual difference between the PVCs, which model the relative content of tissue types within a voxel, and the tissue type probabilities, which model the probability that a majority of a voxel content is of a certain tissue type. Still, tissue type probability modeling allows only observing voxels of a specified label while PVC estimation allows observing voxels which are, for example, 27% of WM and 73% of GM. Modern tissue classification methods based on the tissue type probability estimation often account for partial volume effect e.g. by using mixture of Gaussians to model a single tissue type (SPM5) or by non-parametric modeling of tissue type intensities (Cocosco et al., 2003; Anbeek et al., 2005), however, the conceptual difference still remains. More practically, the estimate of the probability of observing label $j$ at the voxel $i$ becomes:

\[
P(c_i = j | y_i) = \frac{P(c_i = j, y_i)}{\sum_{k=1}^{3} P(c_i = k, y_i)}
\]

where $P(c_i = j | y_i)$ and $P(c_i = j, y_i)$ are the posterior probability that the voxel $i$ is of the label $j$ when image intensity $y_i$ is observed and the joint probability of that $c_i = j$ and image intensity is $y_i$. This equation is valid independent on the chosen modelling scheme (parametric or non-parametric) and it should be compared to the PVC estimate of Eq. (4). As a concrete illustration, consider the case where the tissue type intensities are modelled by the Gaussian densities with equal
voxel intensity varies. The intensity model can be seen in Fig. 7 (Gray Matter tissue type). In Fig. 7, the GM and WM mean intensities have been chosen according to the original noiseless BrainWeb phantom. 3%, 5% and 9% variances and each label is a-priori equally probable. Then, Eq. (17) leads to

\[
P(c_i = j | y_i) = \frac{1}{3} \theta^{(-0.5(y_i - \bar{y}_j)^2 / \sigma^2)}
\]

A theoretical example on the solutions based on Eq. (18) compared to the ML-based PVC estimation result using the same intensity model can be seen in Fig. 7 (Gray Matter tissue type). In Fig. 7, the GM and WM mean intensities have been fixed and the voxel intensity varies.

References


Santago, P., Gage, H.D., 1993. Quantification of brain structures in MRI including con…


