



LoAd: A locally adaptive cortical segmentation algorithm

M. Jorge Cardoso^{a,*}, Matthew J. Clarkson^{a,b}, Gerard R. Ridgway^{b,c}, Marc Modat^a,
Nick C. Fox^b, Sebastien Ourselin^{a,b}
and The Alzheimer's Disease Neuroimaging Initiative¹

^a Centre for Medical Image Computing (CMIC), University College London, UK

^b Dementia Research Centre (DRC), University College London, UK

^c Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, UK

ARTICLE INFO

Article history:

Received 16 September 2010

Revised 28 January 2011

Accepted 2 February 2011

Available online 23 February 2011

Keywords:

Gaussian mixture model
Expectation-maximization
Markov Random Field
Cortical segmentation
Partial volume effect

ABSTRACT

Thickness measurements of the cerebral cortex can aid diagnosis and provide valuable information about the temporal evolution of diseases such as Alzheimer's, Huntington's, and schizophrenia. Methods that measure the thickness of the cerebral cortex from in-vivo magnetic resonance (MR) images rely on an accurate segmentation of the MR data. However, segmenting the cortex in a robust and accurate way still poses a challenge due to the presence of noise, intensity non-uniformity, partial volume effects, the limited resolution of MRI and the highly convoluted shape of the cortical folds. Beginning with a well-established probabilistic segmentation model with anatomical tissue priors, we propose three post-processing refinements: a novel modification of the prior information to reduce segmentation bias; introduction of explicit partial volume classes; and a locally varying MRF-based model for enhancement of sulci and gyri. Experiments performed on a new digital phantom, on BrainWeb data and on data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) show statistically significant improvements in Dice scores and PV estimation ($p < 10^{-3}$) and also increased thickness estimation accuracy when compared to three well established techniques.

© 2011 Elsevier Inc. All rights reserved.

Introduction

The thickness of the cortex has been found to have an important correlation to various diseases such as Alzheimer's (Lerch et al., 2005; Du et al., 2007; Lehmann et al., in press), Huntington's (Rosas et al., 2008), schizophrenia (Nesvåg et al., 2008), and also to normal ageing (Shefer, 1973; Salat et al., 2004; Thambisetty et al., 2010). Automatic extraction of measurements from the cortex, such as thickness, has the potential to provide a biomarker for diagnosis and disease progression (Desikan et al., 2009). However, algorithms for the reliable extraction of the cortical layer are still in need of improvement. From a technical point of view, this problem is intrinsically complex due to the convoluted shape of the cortex and the fact that its normal thickness (2.5 ± 1.5 mm, (von Economo, 1929) is close to the typically acquired MRI voxel dimensions (≈ 1 mm isotropic). This task is further hampered by the presence of noise, partial volume (PV) effects and intensity non-uniformity (INU) across the image.

Segmentation of the brain into its different tissue types has been proposed using methods based on morphological operations (Mangin et al., 1995), edge detection (Tang et al., 2000), fuzzy c-means (Pham, 2002; Wang and Fei, 2009) and probabilistic models. Probabilistic mixture models fitted with the expectation maximisation (EM) algorithm form the basis of several image segmentation methods (Wells et al., 1996; Van Leemput et al., 1999b; Zhang et al., 2001; Ashburner and Friston, 2005). These EM-based image segmentation algorithms were shown to be among the most accurate and robust (Klauschen et al., 2009). Wells et al. (1996) segments the brain into three main tissue types (white matter, grey matter and cerebrospinal fluid), modelling each class as normal distribution after log transformation to make the bias field additive, and assumes a Gaussian distributed bias field model to correct for intensity non-uniformity. Van Leemput et al. (1999b) added a spatial consistency model based on a Markov Random Field (MRF), explicit modelling of the INU with polynomial basis functions, and some prior information about the brain anatomy to initialise and locally constrain the segmentation. Ashburner and Friston (2005) combined image registration with tissue classification, and bias field correction in an elegant unified framework. Despite these advances, the problems of intensity non-uniformity (INU), partial volume effect (PV), noise, image artefacts, limited resolution and the great degree of natural variability, mean that the local intensity difference is not enough to provide an accurate segmentation of fine structures. These problems can lead to an

* Corresponding author.

E-mail address: manuel.cardoso@ucl.ac.uk (M.J. Cardoso).

¹ Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://www.loni.ucla.edu/ADNI>). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Citation.shtml).

incorrect delineation of problematic areas like PV-corrupted grey matter folds, resulting in incorrect segmentations. The use of prior knowledge may also cause problems in areas that have a high degree of natural variability, as the prior information is representative of a sample of a normal population and might not describe a particular subject. The use of probabilistic priors becomes more problematic when an atlas derived from a normal population is used to segment patients with different anatomical or pathological characteristics.

The methods described above are global brain segmentation methods, and are not specifically designed for the cortical layer. In this paper we are interested specifically in cortical segmentation as an input to a voxel-based cortical thickness algorithm. Cortical thickness estimation methods can be broadly categorised into two types: surface-based (Fischl and Dale, 2000; Kim et al., 2005) and voxel-based methods (Jones et al., 2000; Hutton et al., 2008; Lohmann et al., 2003; Acosta et al., 2009). Surface-based approaches fit a triangulated mesh to the internal and external surface of the cerebral cortex. These surface-based methods work in the continuous domain and can achieve sub-voxel accuracy and robustness to image noise due to mesh smoothness constraints. However, these methods are computationally very demanding (normally above 10 h), and often require laborious manual interaction at several stages. Surface-based methods can also produce biased results due to the implicit surface model and topology constraints (MacDonald et al., 2000; Srivastava et al., 2003; Kim et al., 2005; Thompson et al., 2005; Scott et al., 2009).

In contrast, voxel-based techniques that work directly in the 3D voxel grid are much more computationally efficient but are more prone to noise, PV and INU effects and topological errors. To locally improve the detection of PV corrupted sulci, Han et al. (2004) and Acosta et al. (2008) used the information derived from a distance based cost function as a post processing step to try to solve this problem. Hutton et al. (2008) used a layering method based on mathematical morphology to detect deep sulci. However, these approaches are post processing steps; they do not take the new information into account to improve the segmentation. They are also only concerned with improvements in the delineation of deep sulci though the same problems can occur in thinned gyri due to white-matter tissue loss, PV effects and structural readjustments.

In this paper we improve a probabilistic segmentation framework with three novel modifications in order to reduce the influence of the priors in an anatomically coherent way and improve the PV estimation and the delineation of deep sulci and gyri (Fig. 1). Both the solution of the EM algorithm and the information derived from a geodesic distance function are used to locally modify the priors and the weighting of the MRF, enabling the detection of small variations in intensity while maintaining robustness to noise. An MRF energy matrix derived from the anatomical properties of the brain is used to add topological and shape knowledge to the MRF. Although full topological correctness is not ensured, the proposed MRF energy matrix improves the topological characteristics of the segmentation and reduces the PV layer thickness, making it more in line with the theoretical anatomical limit. The implicit modelling of PV and the reduction of the PV layer thickness obviates the need for an empirical

threshold to distinguish between pure and mixed voxels and eases the problem of achieving subvoxel accuracy when calculating the cortical thickness.

Method

Intensity model and MRF regularisation

Starting from the image model developed by Van Leemput et al. (1999b), let $i \in \{1, 2, \dots, n\}$ index the n voxels of an image domain. For coregistered multimodal datasets, intensities form feature vectors $y_i \in \mathcal{R}^m$; for simplicity, we assume unimodal data with $m = 1$. Let z_i denote the tissue type to which voxel i belongs. For K tissue types, $z_i = e_k$ for some k , $1 \leq k \leq K$ where e_k is a unit vector with the k th component equal to one and all the other components equal to zero.

As in Van Leemput et al. (1999a) we represent an INU bias field as a linear combination $\sum_{j=1}^J c_j \phi_j$ of J smoothly varying basis functions $\phi_j(x)$, where x denotes the spatial position and $C = \{c_1, c_2, \dots, c_j\}$ denote the bias field parameters. For mathematical convenience and similarly to Garza-Jinich et al. (1999), Wells et al. (1996), Van Leemput et al. (1999b) and Zhang et al. (2001), we assume that the intensity of the voxels that belong to class k are normally distributed after log transformation with mean μ_k and standard deviation σ_k grouped in $\theta_k = \{\mu_k, \sigma_k\}$. Let $\Phi_y = \{\theta_1, \theta_2, \dots, \theta_K, C\}$ represent the overall model parameters. This log transformation of the data makes the multiplicative bias field additive, ameliorating problems with numerical stability and enabling the existence of a linear least square solution for the coefficient optimisation (Van Leemput et al., 1999b).

Defining Φ_y as the model parameters, the overall probability density for y_i is

$$f(y_i | \Phi_y) = \sum_k f(y_i | z_i = e_k, \Phi_y) f(z_i = e_k) \quad (1)$$

with

$$f(y_i | z_i = e_k, \Phi_y) = G_{\sigma_k} \left(y_i - \mu_k - \sum_j c_j \phi_j(x_i) \right) \quad (2)$$

where $G_{\sigma_k}(\cdot)$ denotes a zero-mean normal distribution with standard deviation σ_k . Eq. (1) can be seen as a mixture of normal distributions.

Thus, by assuming statistical independence between voxels, the overall probability density for the full image can be given by

$$f(y | \Phi_y) = \prod_i f(y_i | \Phi_y) \quad (3)$$

The Maximum Likelihood (ML) parameters for Φ_y can be found by maximisation of $f(y | \Phi_y)$, giving the following update equations for the model parameters:

$$\mu_k^{(m+1)} = \frac{\sum_{i=1}^n p_{ik}^{(m+1)} \left(y_i - \sum_{j=1}^J c_j^{(m)} \phi_j(x_i) \right)}{\sum_{i=1}^n p_{ik}^{(m+1)}} \quad (4)$$



Fig. 1. Segmentation of a BrainWeb T1-weighted dataset with 3% noise and 20% INU: left) BrainWeb ground truth segmentation; centre) MAP with MRF but without the proposed improvements; right) proposed method.

$$\sigma_k^{(m+1)} = \sqrt{\frac{\sum_{i=1}^n p_{ik}^{(m+1)} (y_i - \mu_k^{(m+1)} - \sum_{j=1}^J c_j^{(m)} \phi_j(x_i))^2}{\sum_{i=1}^n p_{ik}^{(m+1)}}} \quad (5)$$

where

$$p_{ik}^{(m+1)} = \frac{f(y_i | z_i = e_k, \Phi_y^{(m)}) f(z_i = e_k)}{\sum_{j=1}^K f(y_i | z_i = e_j, \Phi_y^{(m)}) f(z_i = e_j)} \quad (6)$$

is a weight at the index i and class k and m denotes the iteration number. The estimation of $c_j^{(m+1)}$ is provided by Van Leemput et al. (1999b).

Anatomical priors that incorporate probabilistic information derived from a digital brain atlas are added to the model in order to condition the posterior probabilities and indirectly condition the model parameters. These atlases are brought into correspondence using an affine registration (Ourselin et al., 2000) followed by a free-form non-rigid registration algorithm (Modat et al., 2010)² and are introduced as a weight π_{ik} , integrated in Eq. (1) by making $f(z_i = e_k) = \pi_{ik}$. Eqs. (4)–(6) remain valid and the initial values for p_{ik}^0 , μ_k^0 and σ_k^0 are given by their equations with $c_j = 0$ and $f(y_i | z_i = e_k, \Phi_y) = 1$.

We assume skull stripped images and initially model the problem with $K=6$ classes, each one with a corresponding digital atlas prior probability for white matter (WM), cortical grey matter (cGM), deep grey matter (dGM), external cerebrospinal fluid (eCSF), internal cerebrospinal fluid (iCSF) and dura (DU) respectively at every voxel position. These priors are derived from the ICBM Tissue Probabilistic Atlas³ and are created by merging several priors from several areas together. The images were skull stripped using a semi-automated method (Freeborough et al., 1997) and dilated then filled to include the ventricles and sulci.

The cortical and deep GM are modelled as separate classes to enable thickness calculation over the cortical structures and to enable the segmentation of a broader range of pulse sequences (e.g. new quantitative MR sequences that look at iron concentration – R2 and R2* maps (Gelman et al., 1999)), that have differing intensities for dGM and cGM. The distinction between deep and cortical GM and internal and external CSF also enables different topological and connectivity properties to be assigned to each class. For example the iCSF, i.e. the CSF within the ventricles, can be next to WM or dGM voxels while the eCSF can only be next to cGM voxels. Finally, the dura class is used to compensate for problematic skull stripping situations.

Unfortunately, the intensity model alone only works in relatively ideal conditions because it classifies the voxels of the image based solely on intensity and on the assumption that neighbouring voxels are independent. This makes the segmentation very prone to noise and image artefacts. Therefore, the model has to be made more robust to noise by augmenting the spatial tissue priors with additional prior knowledge about topology and spatial smoothness. This can be achieved by the using an MRF which assumes that the probability that voxel i belongs to tissue k depends on its first-order 3D neighbours \mathcal{N}_i . Using the mean field approximation as described in Zhang (1992) and Van Leemput et al. (1999b), Eq. (6) becomes

$$p_{ik}^{(m+1)} = \frac{f(y_i | z_i = e_k, \Phi_y^{(m+1)}) f(z_i = e_k | p_{\mathcal{N}_i}^{(m)} \Phi_z^{(m)})}{\sum_{j=1}^K f(y_i | z_i = e_j, \Phi_y^{(m+1)}) f(z_i = e_j | p_{\mathcal{N}_i}^{(m)} \Phi_z^{(m)})} \quad (7)$$

with,

$$f(z_i = e_k | p_{\mathcal{N}_i}^{(m)} \Phi_z^{(m)}) = \frac{\pi_{ik} e^{-\beta_i U_{\text{MRF}}(e_k | p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)})}}{\sum_{j=1}^K \pi_{ij} e^{-\beta_i U_{\text{MRF}}(e_j | p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)})}} \quad (8)$$

where $U_{\text{MRF}}(z_i | p_{\mathcal{N}_i}, \Phi_z)$ is an energy function dependent on the parameters Φ_z and, at this stage $\beta_i = 1 \forall i$. Due to the possibility of anisotropic voxel size and slice spacing, the interaction between neighbours in each direction should be different. To take this property into account, a connection strength factor s is introduced as $s = \{s_x, s_y, s_z\} = \{\frac{1}{d_x}, \frac{1}{d_y}, \frac{1}{d_z}\}$, where d is the real-world distance between the centre of neighbouring voxels in each direction. This approach leads to higher weights in the MRF when voxels are closer together. Under this framework,

$$U_{\text{MRF}}(e_k | p_{\mathcal{N}_i}, \Phi_z) = \sum_{j=1}^K G_{kj} \left(\sum_{l \in \mathcal{N}_i^x} s_x p_{lj} + \sum_{l \in \mathcal{N}_i^y} s_y p_{lj} + \sum_{l \in \mathcal{N}_i^z} s_z p_{lj} \right) \quad (9)$$

where $\Phi_z = \{G, s\}$, with G as a $K \times K$ matrix with element G_{kj} containing the transition energy between tissue k and j , and with the MRF neighbourhood system defined as $\mathcal{N}_i = \{\mathcal{N}_i^x, \mathcal{N}_i^y, \mathcal{N}_i^z\} = \{\{i^n, i^s\}, \{i^e, i^w\}, \{i^t, i^b\}\}$.

Although G can be estimated and updated using a mean field theory based approximation, these estimates are only representative of the global image statistics and not of the known brain anatomy. Furthermore, the presence of noise can hamper the correct estimation of the MRF energy matrix. Instead of estimating and updating G at each iteration, we assume constant values based on anatomical proprieties of the brain. The MRF class connectivity network is represented in Fig. 2. The classes connected with arrows are considered neighbouring classes, and the ones that are not connected are considered distant classes. Even though this connectivity matrix is representative of most anatomical neighbouring features, in areas like the ventricle edges, a layer of GM will be assigned to the glial tissue and the PV corrupted voxels in the interface between WM and CSF. This will also happen in areas like the pons. These small anatomical incoherences are related to the constant MRF energy matrix G . A spatially varying MRF energy matrix could be used to spatially change the neighbouring rules, however, this would greatly increase the computational complexity. One should also bear in mind that the neighbouring rules are not a hard constraint. Matrix G is defined as:

$$G_{kj} = \begin{cases} 0 & \text{if class } k \text{ is the same as } j \\ \alpha & \text{if class } k \text{ is neighbouring } j \\ \gamma & \text{if class } k \text{ is distant from } j \end{cases} \quad (10)$$

with

$$0 \leq \alpha \leq \gamma \quad (11)$$

where γ is a penalty factor for anatomically distant classes (e.g. eCSF and WM) and α is a penalty factor for anatomically neighbouring classes (e.g. eCSF and cGM). Under these assumptions, a bigger γ leads to a lower probability that two voxels with anatomically distant labels would be together and a bigger α would increase the sharpness of the transitions between neighbouring tissues, leading to more homogeneous but less detailed segmentations. The values for α and γ used in this paper are 0.5 and 3 respectively.

Segmentation refinement

The model described above is only based on global parameters. However, in some situations, due to lack of image contrast, intensity non-uniformity, partial volume effect and noise, these global

² <http://sourceforge.net/projects/niftyreg/>.

³ Available from http://www.loni.ucla.edu/ICBM/ICBM_Probabilistic.htm.

parameters are not enough to provide an accurate and topologically aware segmentation of fine structures. Three refinement levels were added to compensate for three main problems. First, a method was created to iteratively relax the constraints embedded within the prior information, compensating for problems in areas with high degree of natural anatomical or pathological variability. Second, an explicit modelling of PV was added and the MRF energy matrix was altered in order to incorporate the new classes. This refinement step obviates the need for an artificial threshold to separate pure and mixed voxels and allows different MRF behaviour between pure and PV corrupted areas. Finally, in order to add topological information to the segmentation and to increase the detail of the segmentation, a method to enhance the delineation of PV-corrupted grey matter folds is performed in an iterative manner until convergence. The algorithm's flowchart is shown in Fig. 3.

First level: prior probability relaxation

The EM algorithm is known to converge to a local maximum. In an ML approach, the prior probability drives the EM algorithm to a sensible solution, making it more robust to noise and INU. However, in areas with high anatomical variability, prior driven ML approaches can lead to an erroneous solution because the prior probability for the expected class might be too close to 0 to allow the EM to converge to the desired solution. It can also bias the segmentation towards the template, possibly overshadowing some anatomical differences. We propose a method where the prior probabilities are changed iteratively at each convergence of the EM algorithm, in an anatomically coherent way. As our model parameters become closer to the true solution, we are able to locally relax our prior probability without robustness to noise, INU and PV. This is analogous to coarse-to-fine refinement of regularisation in image registration, for example the gradual reduction of prior influence over the outer iterations in DARTEL (Ashburner and Friston, 2009).

After the EM algorithm converges, the model parameters Φ_y are closer to the true solution. However, due to the *a priori* spatial constraints, the segmentation of patients with different anatomical

and structural characteristics might not converge to the correct solution. In order to relax these constraints and make the segmentation less dependant on these priors, one possible solution might be to smooth the priors and consequently smooth these constraints. However, because these relaxed priors would then be similar to uninformative priors, the problem would become similar to a Maximum Likelihood formulation, resulting in erroneous segmentations in patients with white matter hypo and hyper-intensities. Instead, similarly to Seghier et al. (2008), patient specific priors are generated using an ad hoc transformation over the posteriors. These updated atlases cannot be considered as priors in a strict mathematical sense as they are derived from the data, however they behave as such in this segmentation framework. The patient specific relaxed anatomical atlases are generated as a combination of the current estimates of the posteriors smoothed over anatomically neighbouring classes as described by

$$\pi_{ik} = \frac{p_{ik} + \sum_{j=1}^K H_{kj} \tau_{ik} p_{ij}}{\sum_{l=1}^K (p_{il} + \sum_{j=1}^K H_{lj} \tau_{il} p_{ij})} \quad (12)$$

with

$$H_{kj} = \begin{cases} 0 & \text{if class } k \text{ is the same as } j \\ R_f & \text{if class } k \text{ is next to } j \\ 0 & \text{if class } k \text{ is distant from } j \end{cases} \quad (13)$$

and

$$\tau_{ik} = \frac{1}{1 + \mathcal{E}(p_{ik_{0.5}})} \quad \text{and } 0 \leq R_f \leq 1. \quad (14)$$

Here, τ_{ik} is inversely proportional to $\mathcal{E}(p_{ik})$, defined as the Euclidean distance from point i to the closest hard classified voxel where $p_{ik} > 0.5$. Thus τ_{ik} will be equal to 1 where $p_{ik} > 0.5$ and will have a decreasing value as the distance to the hard classified set increases. The parameter R_f controls the amount of prior probability sharing and H is a matrix containing the same anatomical neighbouring rules as the MRF.

Second level: explicit PV modelling

In PV segmentation, it is common to assume that if two tissues mix in a voxel, all mixing proportions are equally likely. The PV probability can be seen as a number of mixed Gaussians in between the two pure classes, corresponding to all the possible tissue proportions within a voxel (Van Leemput et al., 2003). Ruan et al. (2000) showed that, for brain imaging and for the signal-to-noise ratio and contrast-to-noise ratio levels of the current MRI systems, the density of all these PV Gaussian classes can be approximated by a single Gaussian with a small risk ($\alpha < 1$ for D'Agostino–Pearson normality test). Under this assumption, we use the values of p_{ik} , μ_k , σ_k to initialise an 8 class model, that considers the existence of the 6 original classes (now considered “pure”) and 2 mixed classes {WM, cGM, dGM, eCSF, iCSF, DU, WM/GM, GM/CSF}. Even though every neighbouring class should have a mixed class in between, for the sake of computational complexity we limited the PV estimation to the cortical layer. Using the same framework, the 8 classes are modelled as Gaussian mixtures on the log transformed data. The prior probability, average and variance for the 8 classes model are denoted as π_{ik} , μ_k and σ_k , where the superscript * is used to indicate that they belong to the 8 class model. While the 6 pure classes maintain their previous parameters, the initial mean, standard deviation and priors for the 2 mixed classes have to be estimated from the data. Under the assumption of Gaussian distributed classes on log-transformed data, the initial mixed class Gaussian parameters can be approximated by a mixel distribution (Kitamoto and Takagi, 1999), with mean equal to the arithmetic

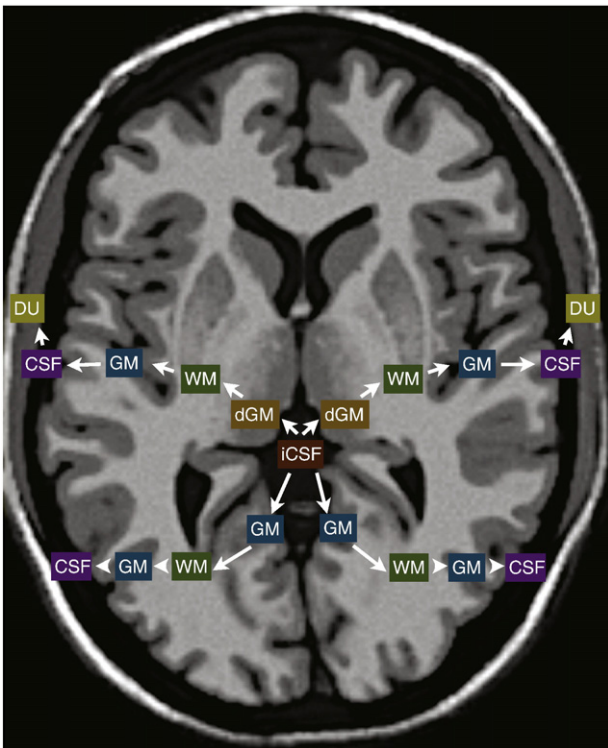


Fig. 2. MRF class connectivity network.

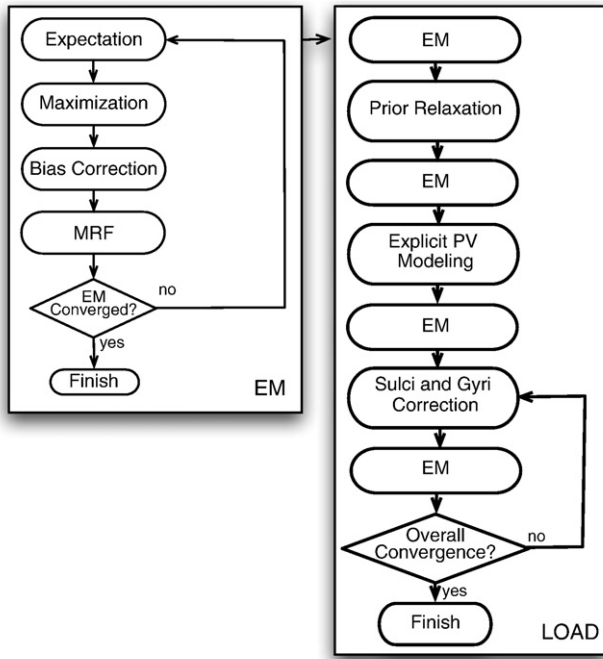


Fig. 3. Algorithm flowchart.

weighted average of its composing class parameters weighted by each class's average fractional content. Thus,

$$\mu_{j/k} = \Gamma_{j/k} \mu_j + (1 - \Gamma_{j/k}) \mu_k \quad (15)$$

where $\Gamma_{j/k}$ is the average of the fractional content (FC) between classes j and k for all voxels with $FC \in [0, 1]$. FC is defined as $FC = (\mu_j - y_i) / (\mu_j - \mu_k)$ and $y_i = y_i - \sum_j c_j \phi_j(x_i)$ is the image intensity corrected for INU. This is equivalent to calculating the average mixing vector $t = [\alpha, 1 - \alpha]$ in the model proposed by Van Leemput et al. (2003) for all the PV containing voxels and using that as a weighting factor. The initial value of the mixed class variance is estimated using the same mixel model. Assuming that the mixel variance is only dependent on his composing classes, the initial estimate of the mixed class variance can then be estimated by

$$(\sigma_{j/k}^2)^* = \Gamma_{j/k}^2 \sigma_j^2 + (1 - \Gamma_{j/k})^2 \sigma_k^2 \quad (16)$$

Van Leemput et al. (2003) observed that the extra parameters or extra Gaussians that have to be introduced into the PV model hamper the segmentation robustness because no prior is available for the PV location. Our approach obviates this problem using information from the 6 class model to generate a patient specific spatial atlas, used as an ad hoc prior for the mixed classes. Due to the multiplicative nature of the probabilities, the mixed class prior is generated from the normalised geometric mean of its composing tissue distributions p_{ij} and p_{ji} , normalised over all classes.

$$\pi_{i(j/k)}^* = \frac{\sqrt{p_{ij} p_{ik}}}{0.5 \Pi_i} \quad (17)$$

with Π_i as a normalisation constant over all classes (see Fig. 4). For the non-mixed classes $\pi_{ik} = p_{ik} / \Pi_i$. The normalised geometric mean reflects how close p_{ik} and p_{ij} are from the situation where both composing tissues have equal proportions, having the value of 1 where $p_{ik} = p_{ij} = 0.5$ and 0 where either p_{ik} or p_{ij} are 0. One should bear in mind though, that $\pi_{i(j/k)}$ is not an estimation of the amount of partial volume, but just a geometrical transformation necessary to

create priors for the mixed class. This new stage of the EM algorithm is initialised with $p_{ik} = \pi_{ik}$.

Third level: MRF weighting for deep sulci and gyri delineation

As presented in Morris et al. (1996) and then discussed in Van Leemput et al. (2003) the MRF minimises the boundary length between tissues, discouraging classifications from accurately following the highly convoluted shape of the human cortex, resulting in incorrectly segmented structures such as deep sulci and gyri. Van Leemput et al. (2003) suggested that a nonstationary MRF model, with different parameters for uniform and convoluted regions, might be an interesting solution to the MRF problem. This is exactly the problem that we were trying to solve with the deep sulci and gyri delineation. Fischl et al. (2002) used a spatially varying MRF prior in order to increase the local label neighbourhood adaptiveness and robustness. Even with non empirical estimation of warp regularisation parameters (Yeo et al., 2008), the creation of sharp priors for this purpose is difficult due to the highly variable sulcal and gyral location. Thus, this method still does not optimally address the MRF bias-variance tradeoff. Instead, we propose to use a modified version of the current posterior estimates in order to generate a patient specific sulci and gyri atlas and use this information as an MRF strength weighting. Even though it is an ad hoc modification, it enables a robust and sharp localisation of these structures, improving the delineation of the cortical folds. In a similar way to Acosta et al. (2008) and Han et al. (2004), we use the information derived from a distance transform to estimate the location of deep sulci and gyri and change the priors and the strength of the MRF only in those locations. Cost functions based on the norm of the gradient of the Euclidean distance transform, like the one used in Acosta et al. (2008), have several drawbacks: Using a Euclidean based distance implicitly assumes that both banks of the sulci or gyri have the same thickness which is frequently not true; the metric is non informative with regards to the current PV estimates; it overlooks the fact that the norm of the gradient can be zero in both local maxima or minima, whereas the areas of interest should always be in local maxima. The cost function proposed by Han et al. (2004) uses the estimated segmentation to add information about the sulci position, however it still suffers from the same mathematical drawbacks as it is also only based on the gradient of the distance. In order to improve on these limitations, a previously published method (Cardoso et al., 2010) was used to detect the sulci and gyri location.

The assumption that both banks of the sulci and gyri have the same thickness can be removed by using the segmentation probabilities as a speed function for an evolving level set. Fig. 5(a) shows the current hard classification of GM, WM and CSF. In (b), the green area is the initial estimate of the level set, initialised from the current hard WM segmentation. This green surface evolves with a speed inversely proportional to the WM probability. Fig. 5(c) shows the resulting geodesic distance (time of arrival) for the evolving front. Both sides of the evolving front will stop as they meet, thereby defining the position of the sulci. These locations are then fed-back into the segmentation framework by locally weighting the MRF and changing the priors (Cardoso et al., 2010). The same process evolving from the eCSF towards the WM will detect the WM stalks.

The functions ω_i^{gyri} , ω_i^{sulci} , used to weight the MRF, are defined as follows:

$$\omega_i^{\text{gyri}} = \mathcal{H}\left(-\nabla \cdot \nabla \mathcal{G}_i\left(h_{\text{WM}}, \frac{\xi}{(\xi + p_{\text{CSF}})}\right)\right) \mathcal{H}\left(\left(1 - \|\nabla \mathcal{G}_i\left(h_{\text{WM}}, \frac{\xi}{(\xi + p_{\text{CSF}})}\right)\|\right)\right) \quad (18)$$

$$\omega_i^{\text{sulci}} = \mathcal{H}\left(-\nabla \cdot \nabla \mathcal{G}_i\left(h_{\text{CSF}}, \frac{\xi}{(\xi + p_{\text{WM}})}\right)\right) \mathcal{H}\left(\left(1 - \|\nabla \mathcal{G}_i\left(h_{\text{CSF}}, \frac{\xi}{(\xi + p_{\text{WM}})}\right)\|\right)\right) \quad (19)$$

where $\nabla \cdot \nabla$ is the Laplacian operator, $\mathcal{G}_i(h_k, s_j)$ is the geodesic distance from point i to the closest member of the hard segmentation

set $h_k = p_{ik} > 0.5$ given a speed function $s_j = \xi / (\xi + p_j)$ and \mathcal{H} is a limiting function defined as,

$$\mathcal{H}(x) = \begin{cases} 1 & x \geq 1 \\ x & 1 > x > 0 \\ 0 & x \leq 0 \end{cases} \quad (20)$$

The limiting function is necessary due to the behaviour of the first and second derivatives G_i in areas where the speed function is close to zero. It also clips the negative component of $\nabla \cdot \nabla \mathcal{G}$, removing the influence of the local minima in the overall cost function. Furthermore, the clipping effect leads to an ω function that is sharp (close to one voxel thick) enforcing a minimum opening. This was done by design since one would expect a sulci or gyri with more than two voxels thick to be already correctly classified. The constant ξ is set to 10^{-6} . An example of \mathcal{G} and ω is shown in Fig. 6. The main advantage of a divergence based metric is the ability to distinguish between local maxima and minima, improving the robustness of the sulci and gyri detection. In order to introduce local adaptivity of the MRF, a local weighting function is incorporated in Eq. (8) by making β_i a spatially varying value

$$\beta_i = (1 - \omega_i^{\text{sulci}})(1 - \omega_i^{\text{gyri}}) \quad (21)$$

Normally β_i corresponds to the overall MRF strength, however, in this case, the overall MRF strength is directly introduced into the α and γ penalty factors and β_i only acts as a local weighting. The values of ω^{sulci} and ω^{gyri} vary between [0,1] and have a value of 1 near the centre of the sulci and the centre of the gyri respectively. In a similar way, the value of β_i lies between [0,1] and has a value of 0 near the centre of the sulci and gyri.

The functions ω_i^{sulci} and ω_i^{gyri} are also used to iteratively change π_{ik} to give more prior probability to the respective classes in areas where deep sulci and gyri should exist.

For classes WM/GM, GM and GM/CSF, π_{ik} is updated as

$$\pi_{i(\text{WM/GM})}^* = p_{i(\text{WM/GM})} + (\omega_i^{\text{gyri}} p_{i\text{GM}}) \quad (22)$$

$$\pi_{i(\text{GM})}^* = p_{i\text{GM}} \beta_i \quad (23)$$

$$\pi_{i(\text{GM/CSF})}^* = p_{i(\text{GM/CSF})} + (\omega_i^{\text{sulci}} p_{i\text{GM}}) \quad (24)$$

The values of π_{ik} are then normalized in order to sum to one. Both the MRF's β_i and the priors π_i are updated every time the EM converges, and a new EM starts. The algorithm finishes when the ratio of likelihood change is less than a predefined ε , here set to 10^{-3} .

Experiments and results

In this section, the proposed cortical segmentation algorithm was evaluated in terms of its independent parts and its overall performance. The first two experiments are intended to show the contribution of each component to segmentation performance. The proposed method was then evaluated globally against synthetic and clinical data in order to access the accuracy of the PV estimation, segmentation overlap and group separation and additionally, the method was compared to three state of the art methods: FANTASM [Pham (2002)], SPM8 [Ashburner and Friston (2005)] and FAST [Zhang et al. (2001)]. The first method is a fuzzy c-means based segmentation with bias field optimisation and noise tolerance. The second method is an EM based iterative segmentation/registration algorithm with bias correction and the last method is an EM based segmentation, specifically chosen because it uses an MRF to add spatial consistency. In all statistical tests the significance level was set

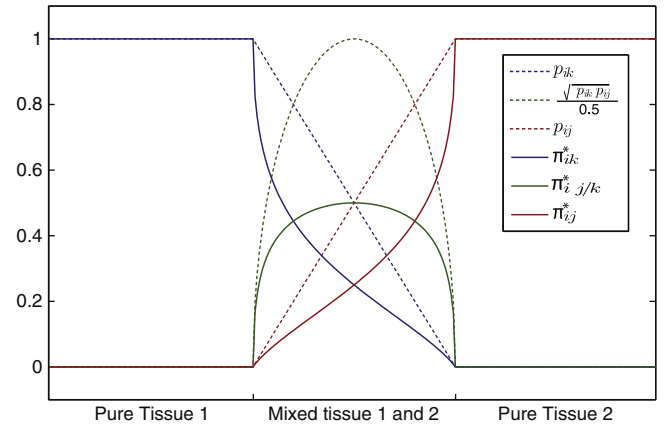


Fig. 4. The mixed class prior (dashed green) is the normalised geometric mean of p_{ik} and p_{ij} (dashed blue and red respectively). The continuous lines represent their value after normalisation over all classes.

to $p < 10^{-3}$. Unless mentioned otherwise, the relaxation fraction $R_f = 1$.

Atlas dependency study

A segmentation algorithm that is fully independent from the chosen atlas is expected to produce the same result when segmenting a dataset with two different atlases. However, the use of different atlases changes the prior probability and thus the final segmentation results. In order to evaluate the segmentation dependency on the atlases and the effect of the prior relaxation, a subset of 40 subjects, 20 patients diagnosed with AD and 20 age- and gender-matched controls were selected from the ADNI database. These datasets were segmented using two different anatomical atlases and 4 different relaxation factors R_f between 0 and 1, leading to 320 different segmentations. The two different atlases were the ICBM452 and the MNI305 Evans et al. (1993). The ICBM452 was created by non-rigidly registering and averaging 452 normal MRI scans while the MNI305 was created by affinely registering 305 normal MRI scans. Both atlases are representative of a normal population, with the main difference being the registration method used to create them (see Fig. 7).

For each dataset and relaxation factor, a fuzzy Dice score (Crum et al., 2006) was calculated between the cortical GM segmentations obtained using the two atlases. The fuzzy Dice score assesses the overlap and the PV differences between the segmentations without the need for a threshold value. The results are shown in Fig. 8. When the prior relaxation is applied to the control population there is almost zero difference in the Dice score average and just a small decrease in the standard deviation. However, when the prior relaxation is applied to an AD population, there is an upward trend in the median Dice score and a reduction in the interquartile difference when the relaxation factor is increased, with the median Dice score going from 0.906 to 0.924.

Thickness measurement evaluation

Voxel-based cortical thickness measurements are critically dependent on the quality of the segmentation and its topology. As there is no ground truth, a digital phantom was used in order to assess the accuracy and precision of thickness measurements.

A very high resolution digital phantom containing finger and sheet like collapsed sulci and gyri was created, simulating the complex and convoluted structure of the cortex. The phantom's white matter is homeotopic to a ball and the cortical layer has a Euclidean thickness of 8 mm between the inner and outer surface. The phantom was created on a 0.25 mm isotropic image resulting in $600 \times 600 \times 1000$ voxels.

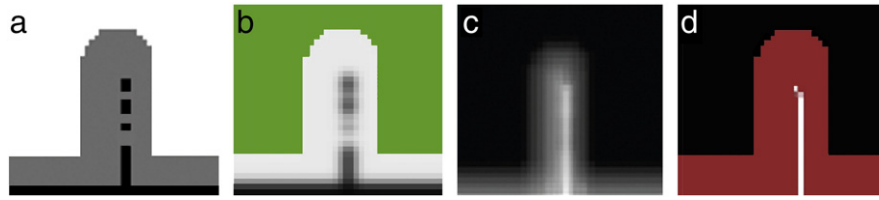


Fig. 5. Sulci localisation using the proposed metric. (a) Current binary segmentation, (b) hard segmented set in green with the respective speed function s_j in grey levels, (c) geodesic distance (time of arrival), (d) the phantom in red overlaid with the detected sulci location in white.

The thickness of the high resolution phantom was calculated using a Laplace equation based method (Acosta et al., 2009). Due to the curved nature of the Laplace equation's streamline, the resulting ground truth mean (standard deviation) thickness was 8.13 (0.15) mm. The phantom was then Fourier-resampled to reduce the resolution by a factor of 5 in each dimension. Two levels of complex Gaussian noise were also added in the Fourier domain, resulting in two low resolution Rician noise corrupted phantoms. To obtain artificial anatomical priors for the segmentation step, the ground truth segmented images were Gaussian filtered ($\sigma = 4$ mm) to simulate the anatomical variability. The thickness was then measured on the segmented low resolution phantoms using a Laplace equation based method with a Eulerian–Lagrangian approach as described in Acosta et al. (2009).

The results are shown in Fig. 9 and Table 1. When compared to the ground truth, the proposed method yields a difference in the average thickness of 0.14 mm and 0.48 mm for the low and high noise phantoms respectively. The standard ML approach with the MRF but without the proposed improvements yields a difference in average thickness of 4.74 mm and 4.36 mm for the low and high noise phantoms respectively. Finally, the standard ML approach without either the MRF or the proposed improvements yields a difference in average thickness of 3.98 mm and 1.22 mm for the low and high noise phantoms respectively.

Segmentation evaluation

20 datasets were downloaded from the BrainWeb (<http://www.bic.mni.mcgill.ca/brainweb>) MR image simulator. Each dataset contained a simulated T1-weighted image and a corresponding ground truth grey matter probabilistic atlas. The simulated data was generated using a spoiled FLASH sequence with $TR = 22$ ms, $TE = 9.2$ ms, $\alpha = 30^\circ$ and 1-mm isotropic voxel size with simulated 3% noise and 20% INU (Aubert-Broche et al., 2006). The 20 simulated images were segmented using the proposed method, SPM8, FAST and FANTASM, each one resulting in either a PV segmentation or its fuzzy c-means equivalent. We focused our analysis on the GM class as most of the differences between the methods will be in the cortical area.

For each segmentation, a normalised cumulative histogram of the absolute difference between the segmentation and the ground truth

was calculated. Fig. 10(a) shows the mean and standard deviation as error bars for the 20 datasets. The proposed method results in 94% of voxels having an absolute difference of less than 0.1 compared to 87% for FAST, 84% for SPM8 and 80% for FANTASM.

Fig. 10 also shows p-values calculated using a two-tailed unequal-variance two-group t-tests between the normalised absolute difference histogram values of our method and each of the other two methods. The proposed method achieves significantly better PV estimation than FAST, SPM8 and FANTASM for all threshold values.

To evaluate the quality of the binarised and PV segmentations, the binary and fuzzy Dice scores (Zijdenbos et al., 1994; Crum et al., 2006) were calculated between the segmentations and the ground truth. The binary Dice score was calculated in order to understand the behaviour of the overlap metric with regards to the threshold level. Here, the binary Dice score was estimated at several PV thresholds and two-tailed unequal-variance two-group t-tests were calculated between the proposed method, FAST, SPM and FANTASM. Fig. 10(b) shows the average Dice score and standard deviation as error bars for the 20 datasets and the results of the statistical test. For all threshold values, the proposed method achieved significantly higher average Dice scores than FAST, SPM and FANTASM. The fuzzy Dice score was calculated in order to give an overall measure of unthresholded segmentation alignment. Here, the average fuzzy Dice score for the 20 datasets was 0.959, 0.941, 0.929 and 0.927 for the proposed method, FAST, SPM and FANTASM respectively.

ADNI data study

To further investigate if the proposed refinements are useful when extracting measurements from the segmentation, cortical thickness was calculated on ADNI data in order to evaluate group separation between controls and Alzheimer's Disease (AD) diagnosed patients. This metric was chosen because it is dependent on both the accuracy and the topology of the segmentation. A subset of the ADNI database was used in this study. From the full database, 88 AD diagnosed patients and 82 age- and gender-matched controls were selected, with T1-weighted volumetric images acquired on 1.5 T units using 3D MPRAGE or equivalent protocols with varying resolutions (typically $1.25 \times 1.25 \times 1.2$ mm).

All 170 datasets were segmented using the proposed method and the two best methods with regards to the fuzzy Dice score from the previous section – SPM8's standard unified segmentation and FAST. Cortical thickness was then calculated using a Laplace equation based algorithm (Acosta et al., 2009). This method requires the user to select a threshold to classify a voxel as pure (normally 0.95) in order to solve the Laplace equation. This threshold is normally set high and not at the optimum Dice score in order to increase the level of detail in the obscured sulci and gyri area, resulting in less biased thickness measurements. As both FAST and the proposed method use an MRF to add spatial consistency, a voxel was considered pure when $p_{GM} = 1$. However, for SPM8, a voxel was considered pure for $p_{GM} > 0.95$ to compensate for the lack of MRF. The same transformation used to map the priors to the individual subjects was used to propagate the AAL template (Tzourio-Mazoyer et al., 2002), and the average thickness at

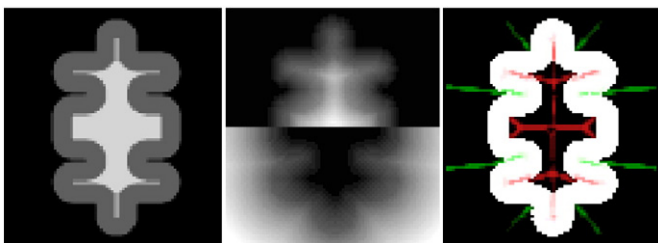


Fig. 6. Sulci and gyri enhancement: (left) expected segmentation; (centre) $G(h_{CSF}, S_{WM})$ and $G(h_{WM}, S_{CSF})$ on the top and bottom respectively; (right) ω_i^{sulci} and d_i^{gyri} in green and red respectively.

the central Laplacian isoline was calculated for 52 AAL cortical regions. Two-tailed unequal-variance two-group t-tests between patients and controls over each AAL region were calculated. In order to visualise the results (Fig. 11), log transformed p-values were propagated back to the corresponding areas on a normal population smoothed 3D model with positive and negative values associated with thinning and thickening respectively. The p-values were thresholded at $p = 10^{-3}$. The expected areas affected in AD patients are the middle and inferior temporal, superior and inferior parietal and middle frontal gyrus bilaterally. Using the proposed method as segmentation, all of these areas show statistically significant differences in thickness with $p < 10^{-5}$ in the middle temporal and parietal regions and $p < 10^{-3}$ in the middle frontal gyrus region. Although most of the same expected areas are statistically significant when using FAST's segmentation, the middle frontal gyrus area does not show statistically significant differences. Also, only the left middle and inferior temporal regions and right parietal region show statistically significant differences in thickness with $p < 10^{-5}$ leading to a noticeable lack of symmetry between hemispheres. Using SPM, there is an overall decrease of statistical significance throughout the brain, with only the middle and inferior temporal areas above the $p < 10^{-3}$ threshold.

Computation time

The total computation time is in line with current state of the art segmentation methods. The segmentation step takes on average less than 2 min, with an overhead of less than 3 min for the registration of the priors since the registration is fairly broad, resulting in an average total time below 6 min per dataset.

Discussion

In this work we have developed a segmentation method specifically designed for the cerebral cortex. We evaluated the robustness and accuracy of the segmentation and PV estimation and the ability to directly use the segmentation for cortical thickness estimation on synthetic and real data.

In Atlas dependency study section, a study testing for atlas independence was performed on real data from the ADNI database in order to evaluate the efficacy of the prior relaxation. When segmenting the datasets using two normal population atlases, an algorithm that is less dependent on the prior probability would produce two closely matching segmentations. As expected, the results show that when priors derived from a control population are applied to a control group, there is no change in the average dice score, since the atlas is representative of that specific population. However, when a control population atlas is applied to an AD population, an increase of the relaxation factor has a positive effect on the segmentation overlap. Although the difference is not significant, there is an upward trend on the average and a decrease on the standard deviation of the Dice score distributions. This shows that after prior relaxation, the

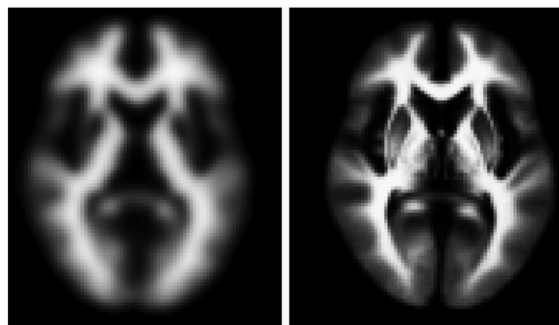


Fig. 7. (Left) The MNI305 atlas and (right) the ICBM452.

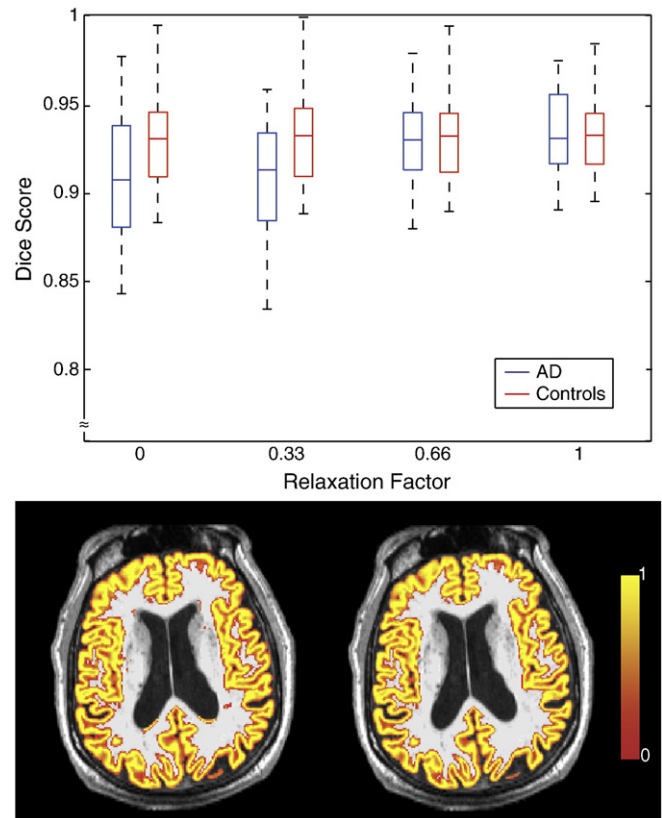


Fig. 8. (Top) The fuzzy Dice scores between the cortical GM segmentations using different atlas and relaxation factors. Segmentation example with RelaxationFactor = 0 (bottom left) and Relaxation Factor = 1 (bottom right). Notice the improved segmentation results in the ventricle area.

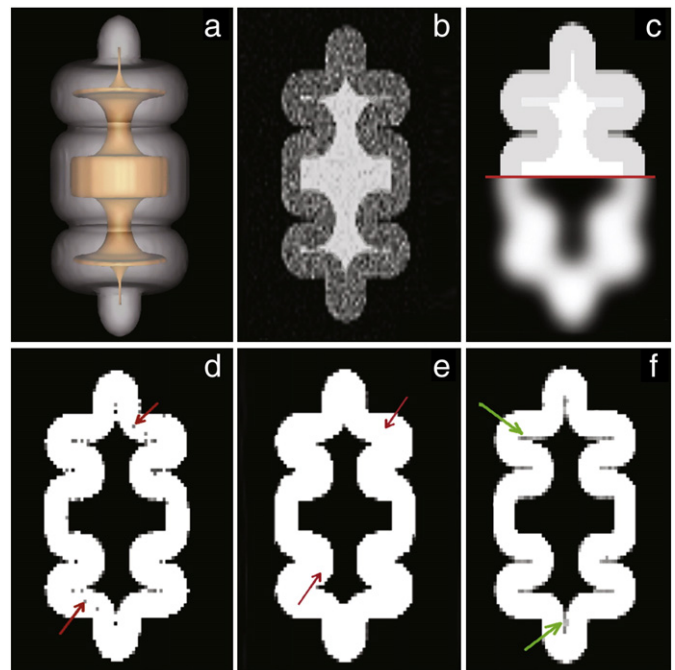


Fig. 9. Phantom segmentation and thickness results: a) 3D model of the phantom, b) high noise phantom, c) true labels and GM prior used, d) ML without MRF, e) ML with MRF, f) proposed method. The red arrows point to the presence of noise and lack of detail causing wrong thickness measurements. The green arrows point to the detected deep gyri.

segmentations become more similar, and thus, less dependent on the priors. Visual assessment shows a noticeably better segmentation in the ventricle area of the AD patients, mainly when the ventricles are expanded (see Fig. 8). This is caused by the spatial ambiguity when the ventricle edge is close to the cortical GM. A higher relaxation factor also produces a visually better temporal lobe segmentation when these are highly atrophied. Overall, the extra knowledge introduced in the prior relaxation step by the neighbouring tissue structure leads to reduced bias, resulting in less ambiguity regarding miss-segmented areas due to different anatomy.

A second experiment showed that the proposed improvements can help to accurately extract meaningful thickness measurements from the segmentation. Using a digital phantom created specifically for this purpose, the average thickness was measured with the proposed method, without the refinement steps (MAP with MRF), and just using the intensity component of the model (MAP without MRF). The results are displayed in Table 1. Consistent results were found for both low and high noise cases. An unweighted MRF causes an overestimation of the thickness and standard deviation due to the lack of detail in highly convoluted and PV corrupted areas. Without any type of MRF, the thickness measurements are much more prone to noise, leading to a number of short paths to mis-segmented voxels and consequently an artificial increase of the standard deviation of the measurement. Oddly, when the noise level is high, the presence of short paths combined with the lack of detail leads to a more accurate estimate of the average thickness. However, because the standard deviation is much higher than expected, this measurement lacks precision.

In Segmentation evaluation section, the Dice score and PV estimation accuracy were evaluated using BrainWeb data. The proposed method and FAST both showed higher PV estimation accuracy than SPM8 and FANTASM. This is most probably due to the MRF smoothing properties that make the PV estimation more robust. Also, the MRF will ensure a more robust assignment of voxels surrounded by only one tissue class, thus making the posterior probabilities more closely resemble partial volume fractions. The small Dice score improvement of the proposed method can be explained by the adaptive nature of the MRF in areas close to sulci and gyri, increasing the level of detail whilst maintaining robustness to noise. On the other hand, due to the lack of adaptivity in FAST's MRF, some of the details are lost, leading to worse PV estimation when compared to the proposed method. SPM8 underperforms both FAST and the proposed method with regards to PV estimation accuracy. We speculate that for cortical segmentation specifically, the advantages of having an iterative segmentation/registration procedure may not compensate for the lack of MRF. Finally, even though FANTASM is tolerant to noise, it does not model noise implicitly. This might explain the small underperformance with regards to Dice score of FANTASM over the other methods for low PV differences. The difference between FANTASM and the proposed method becomes smaller for difference values above 0.3.

The proposed method achieved significantly higher Dice scores when compared to FAST, SPM and FANTASM. We hypothesise that the improved overlap between structures is probably due to the enhanced delineation of the sulci and gyri and implicit PV modelling. Also

because these improvements are iteratively fed back into the segmentation, there is a gradual reduction of the PV related parameter bias. One might also conclude that SPM outperforms FAST in terms of Dice score due to the iterative segmentation/registration procedure, improving the overlap of the segmented structures. Another explanation might be the lack of spatial adaptiveness in FAST's MRF, as the MRF tends to minimize the boundary length between tissues which discourages classifications from accurately following the highly convoluted shape of the human cortex. For the proposed method, this problem is reduced as the MRF is spatially adaptive.

In the fourth experiment, using ADNI data, we compared three segmentation methods in terms of group separation between control subjects and Alzheimer's Disease (AD) diagnosed patients. Using the proposed segmentation we see a statistically significant, clinically-expected pattern of difference in cortical thickness between AD patients and controls. Although most of the same expected areas are also statistically significant when using FAST's segmentation, there is a less symmetric pattern of atrophy and some of the expected areas (i.e. right and left middle frontal gyrus) do not achieve statistical significance. This is probably caused by the lack of detail due to the use of a stationary MRF. When using SPM, there is a noticeable overall decrease of statistical significance throughout the brain, with only the middle and inferior temporal areas achieving statistical significance. This is again caused by the lack of detail, mostly due to the need for an artificial threshold to separate pure from non-pure voxels. This shows how important the presence of an MRF is when segmenting the cortex. Throughout the literature, the vast majority of clinical studies have been carried out using surface-based cortical thickness techniques (Lerch et al., 2005; Du et al., 2007; Lehmann et al., in press; Rosas et al., 2008; Nesvåg et al., 2008; Salat et al., 2004) with a few using voxel-based techniques (Querbes et al., 2009). Both methods depend on the segmentation step; however, for surface-based techniques, the segmentation is only used as an initialisation for a surface mesh. The mesh is typically deformed to fit the cortical GM/WM boundary and expanded outwards to the GM/CSF boundary. This gives surface-based methods sub-voxel accuracy and robustness to noise. However, due to smoothness and topology constraints, it is difficult to correctly fit the surface to very complex shapes thus requiring laborious manual corrections. Additionally, the implicit surface modelling can lead to bias in the thickness measurements (MacDonald et al., 2000; Kim et al., 2005). Conversely, voxel-based techniques can potentially cope with any topology or shape because they work on the 3D voxel grid. However, these techniques were never specifically tailored for the highly convoluted shape of the cortex. The proposed segmentation method improves the quality and topology of the cortical segmentation and enhances the detection of PV corrupted sulci and gyri, enabling the direct use of the segmentation for cortical thickness as opposed to requiring post-processing techniques (Hutton et al., 2008; Lohmann et al., 2003; Acosta et al., 2009).

In this paper, the focus has been on accurate segmentation specifically for the cortex and how can that directly influence the thickness measurements. We have not compared cortical thickness results with other cortical thickness algorithms. We consider that the comparison with other cortical thickness estimation methods is necessary in order to validate the segmentation method for cortical thickness estimation. However, such a comparison requires voxel-based and surface-based measurements to be brought together in a common space, which is difficult to achieve without bias towards either approach. For this reason, we believe that comparison to surface-based methods is out of the scope of the paper. Future work will compare voxel-, registration- and surface-based cortical thickness estimation techniques.

On a methodological side, future work will investigate the use of Variational Bayes inference and hyperparameter optimisation in a similar way to (Woolrich and Behrens, 2006), enabling an unification of the segmentation framework. Furthermore, we would also like to

Table 1

Table contains the thickness average and standard deviation for the three methods and two levels of noise.

	Low noise	High noise
	Mean (std) mm	Mean (std) mm
ML without MRF	12.11(2.55)	9.35(3.10)
ML with MRF	12.87(2.98)	12.48(2.82)
Proposed method	8.27(0.32)	8.61(0.91)

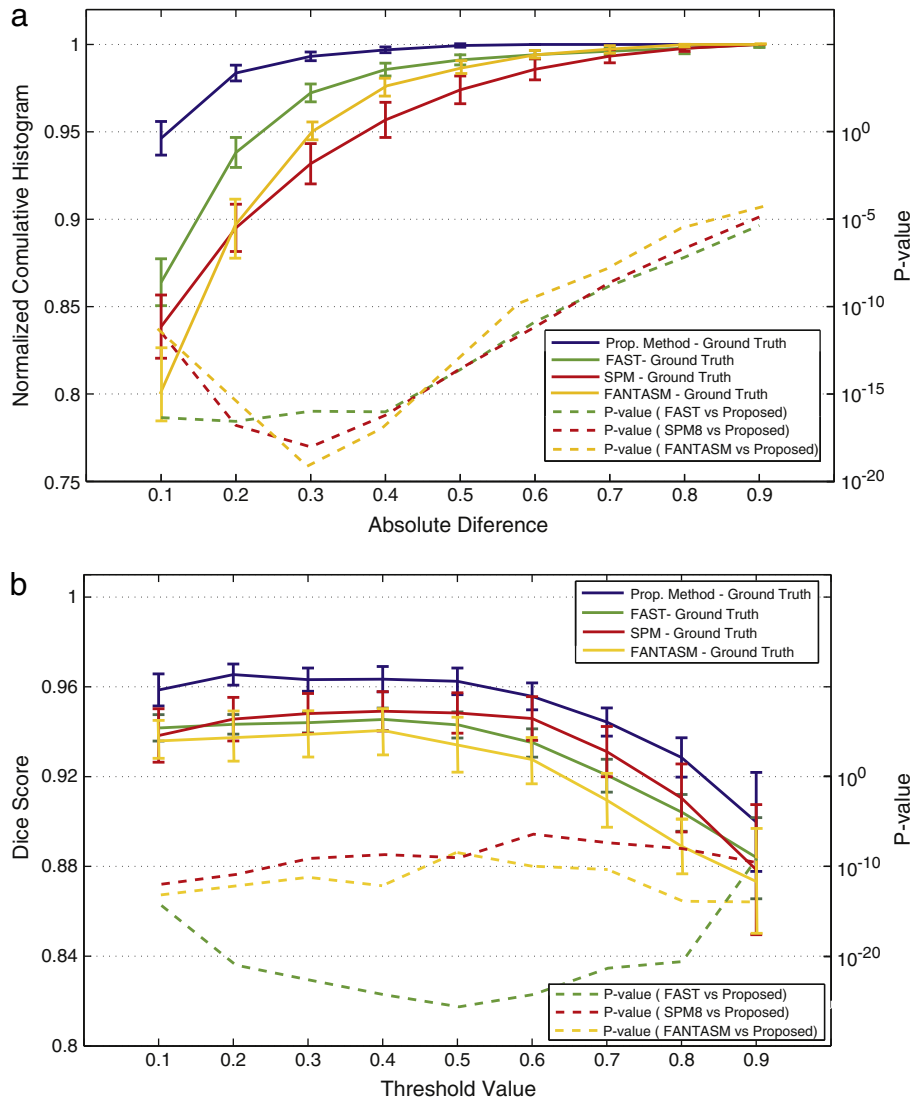


Fig. 10. (a) Normalised cumulative histogram of the absolute difference between the segmentation and the ground truth; (b) Dice score between the segmentation and the ground truth at several threshold values.

explore the use of topological constrains on the space of solutions in order to obtain a topologically correct segmentation of each structure.

Conclusions

We have presented a segmentation algorithm tailored for applications such as cortical thickness estimation. The main contributions of this work lie in three refinement steps. First we developed a method that iteratively relaxes and modifies the prior information in an anatomically coherent way to reduce the bias towards the priors. We then modelled the PV effect explicitly and adapted an MRF energy to reflect the inclusion of these new classes. Finally, we introduced a new distance based cost function to add information about the location of PV corrupted grey matter folds and integrated that information into the segmentation framework.

The method achieves more accurate and precise delineation of collapsed grey matter folds without losing robustness to noise and intensity inhomogeneity. Even though some of these refinement steps can be considered as *ad-hoc*, they are integrated within a single framework. Quantitative analysis on 20 BrainWeb datasets showed statistically significant improvements in the accuracy of the PV estimation and in the Dice score when compared to three state of the art techniques. Cortical thickness measurements on a new digital

phantom demonstrated improvements in the accuracy and robustness of the thickness calculation using the proposed method, when compared to other methods. Results on ADNI data showed clinically-expected patterns of cortical thinning between AD patients and controls using the proposed method, with highly significant group differences in several expected regions.

Acknowledgments

This study was supported by a scholarship from the Fundação para a Ciência e a Tecnologia, Portugal (Scholarship number SFRH/BD/43894/2008). Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer’s Association and Alzheimer’s Drug Discovery Foundation, with participation from the U.S. Food and Drug

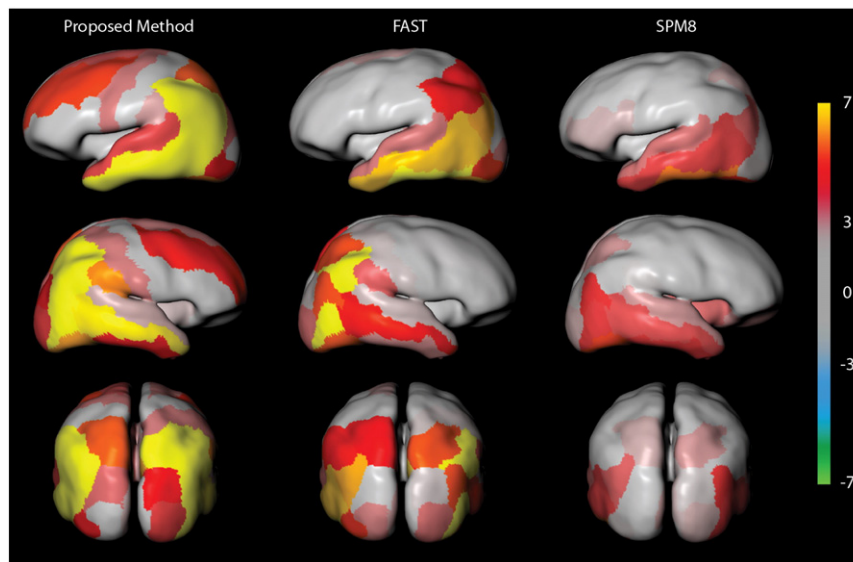


Fig. 11. Statistical significance of cortical thickness between AD patients and controls: colour coded p-values are represented in logarithmic scale with positive and negative values associated with thinning and thickening respectively.

Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organisation is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation.

This work was undertaken at UCL/UCLH which received a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. The Dementia Research Centre is an Alzheimer's Research Trust Co-ordinating centre and has also received equipment funded by the Alzheimers Research Trust. Matthew J. Clarkson is supported by the UCLH/UCL Comprehensive Biomedical Research Centre grant 168 and previously by the TSB grant M1638A, Nick C. Fox is funded by the Medical Research Council (UK). The authors would like to thank the ADNI study subjects and investigators for their participation.

Appendix A. Clinical data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.⁴ The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organisations, as a \$60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California-San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions

and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research – approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years. For up-to-date information see <http://www.adni-info.org>.

References

- Acosta, O., Bourgeat, P., Frapp, J., Bonner, E., Ourselin, S., Salvado, O., 2008. Automatic delineation of sulci and improved partial volume classification for accurate 3D voxel-based cortical thickness estimation from MR. *Lecture Notes in Computer Science – MICCAI*, pp. 253–261.
- Acosta, O., Bourgeat, P., Zuluaga, M.A., Frapp, J., Salvado, O., Ourselin, S., Alzheimer's Disease Neuroimaging Initiative, 2009. Automated voxel-based 3D cortical thickness measurement in a combined Lagrangian–Eulerian PDE approach using partial volume maps. *Med. Image Anal.* 13 (5), 730–743 (Oct).
- Ashburner, J., Friston, K.J., 2005. Unified segmentation. *Neuroimage* 26 (3), 839–851 (Jan).
- Ashburner, J., Friston, K.J., 2009. Computing average shaped tissue probability templates. *Neuroimage* 45 (2), 333–341 (Jan).
- Aubert-Broche, B., Griffin, M., Pike, G.B., Evans, A.C., Collins, D.L., 2006. Twenty new digital brain phantoms for creation of validation image data bases. *IEEE Trans. Med. Imaging* 25 (11), 1410–1416 (Nov).
- Cardoso, M.J., Clarkson, M.J., Modat, M., Ridgway, G.R., Ourselin, S., 2010. Locally weighted Markov random fields for cortical segmentation. 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 956–959 (14–17 2010).
- Crum, W., Camara, O., Hill, D., 2006. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans. Med. Imaging* 25 (11), 1451–1461.
- Desikan, R.S., Cabral, H.J., Hess, C.P., Dillon, W.P., Glastonbury, C.M., Weiner, M.W., Schmansky, N.J., Greve, D.N., Salat, D.H., Buckner, R.L., Fischl, B., Alzheimer's Disease Neuroimaging Initiative, 2009. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. *Brain* 132 (Pt 8), 2048–2057 (Aug).
- Du, A.-T., Schuff, N., Kramer, J.H., Rosen, H.J., Gorno-Tempini, M.L., Rankin, K., Miller, B.L., Weiner, M.W., 2007. Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. *Brain* 130 (Pt 4), 1159–1166 (Apr).
- Evans, A., Collins, D., Mills, S., Brown, E., Kelly, R., Peters, T., 1993. 3D statistical neuroanatomical models from 305 MRI volumes. *Nucl. Sci. Symp. Med. Imaging Conf.* 3, 1813–1817.
- Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. USA* 97 (20), 11050–11055 (Jan).
- Fischl, B., Salat, D.H., Busa, E., Albert, M.S., Dieterich, M., Haselgrove, C., Kouwe, A.V.D., Killiany, R.J., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33 (3), 341–355 (Jan).
- Freeborough, P.A., Fox, N.C., Kitney, R.L., 1997. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. *Comput. Meth. Programs Biomed.* 53 (1), 15–25.

⁴ <http://www.loni.ucla.edu/ADNI>.

- Garza-Jinich, M., Medina, V., Yañez..., O., 1999. 10th International Conference on Image Analysis and Processing (ICIAP'99) (Jan) (computer.org).
- Gelman, N., Gorell, J., Barker, P., Savage, R.M., Spickler, E.M., Windham, J.P., Knight, R.A., 1999. Mr imaging of human brain at 3.0 t: preliminary report on transverse relaxation rates and relation to estimated iron content. *Radiology* 1 (210), 759–767 (Jan).
- Han, X., Pham, D.L., Tosun, D., Rettmann, M.E., Xu, C., Prince, J.L., 2004. CRUISE: cortical reconstruction using implicit surface evolution. *Neuroimage* 23 (3), 997–1012 (Nov).
- Hutton, C., Vita, E.D., Ashburner, J., Deichmann, R., Turner, R., 2008. Voxel-based cortical thickness measurements in MRI. *Neuroimage* 40 (4), 1701–1710 (Jan).
- Jones, S.E., Buchbinder, B.R., Aharon, I., 2000. Three-dimensional mapping of cortical thickness using Laplace's equation. *Hum. Brain Mapp.* 11 (1), 12–32 (Jan).
- Kim, J.S., Singh, V., Lee, J.K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., Lee, J.M., Kim, S.I., Evans, A.C., 2005. Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *Neuroimage* 27 (1), 210–221 (Aug).
- Kitamoto, A., Takagi, M., 1999. Image classification using probabilistic models that reflect the internal structure of mixels. *Pattern Anal. Appl.* 2, 31–43.
- Klauschen, F., Goldman, A., Barra, V., Meyer-Lindenberg, A., Lundervold, A., 2009. Evaluation of automated brain MR image segmentation and volumetry methods. *Hum. Brain Mapp.* 30 (4), 1310–1327 (Apr).
- Lehmann, M., Crutch, S.J., Ridgway, G.R., Ridha, B.H., Barnes, J., Warrington, E.K., Rossor, M.N., Fox, N.C., in press. Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease. *Neurobiol. Aging*. doi:10.1016/j.neurobiolaging.2009.08.017.
- Lerch, J., Pruessner, J., Zijdenbos, A.P., Hampel, H., Teipel, S., Evans, A.C., 2005. Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. *Cereb. Cortex* 15 (7), 995 (Jul).
- Lohmann, G., Preul, C., Hund-Georgiadis, M., 2003. Morphology-based cortical thickness estimation. *Lecture Notes in Computer Science – IPMI*, pp. 89–100.
- MacDonald, D., Kabani, N., Avis, D., Evans, A.C., 2000. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *Neuroimage* 12 (3), 340.
- Mangin, J., Frouin, V., Bloch, I., Régis, J., 1995. From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations. *J. Math. Imaging Vis.* 5 (4), 297–318 (Jan).
- Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S., 2010. Fast free-form deformation using graphics processing units. *Comput. Meth. Programs Biomed.* 98 (3), 278–284 (Jun).
- Morris, R.D., Descombes, X., Zerubia, J., 1996. The Ising/Potts model is not well suited to segmentation tasks. *IEEE Digital Signal Processing Workshop* (Sep).
- Nesvåg, R., Lawyer, G., Varnås, K., Fjell, A.M., Walhovd, K.B., Frigessi, A., Jönsson, E.G., Agartz, I., 2008. Regional thinning of the cerebral cortex in schizophrenia: effects of diagnosis, age and antipsychotic medication. *Schizophr. Res.* 98 (1–3), 16–28 (Jan).
- Ourselin, S., Roche, A., Prima, S., Ayache, N., 2000. Block matching: a general framework to improve robustness of rigid registration of medical images. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2000*.
- Pham, D.L., 2002. Robust fuzzy segmentation of magnetic resonance images. *Computer-Based Medical Systems*, pp. 127–131 (Jan).
- Querbes, O., Aubry, F., Pariente, J., Lotterie, J.-A., Démonet, J.-F., Duret, V., Puel, M., Berry, I., Fort, J.-C., Celsis, P., Alzheimer's Disease Neuroimaging Initiative, T., 2009. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain* 8 (132), 2036–2047.
- Rosas, H.D., Salat, D.H., Lee, S.Y., Zaleta, A.K., Pappu, V., Fischl, B., Greve, D.N., Hevelone, N., Hersch, S.M., 2008. Cerebral cortex and the clinical expression of Huntington's disease: complexity and heterogeneity. *Brain* 131 (Pt 4), 1057–1068 (Apr).
- Ruan, S., Jaggi, C., Xue, J., Fadili, J., Bloyet, D., 2000. Brain tissue classification of magnetic resonance images using partial volume modeling. *IEEE Trans. Med. Imaging* 19 (12), 1179–1187 (Dec).
- Salat, D.H., Buckner, R.L., Snyder, A.Z., Greve, D.N., Desikan, R.S.R., Busa, E., Morris, J.C., Dale, A.M., Fischl, B., 2004. Thinning of the cerebral cortex in aging. *Cereb. Cortex* 14 (7), 721–730 (Jul).
- Scott, M.L.J., Bromiley, P.A., Thacker, N., Hutchinson, C.E., Jackson, A., 2009. A fast, model-independent method for cerebral cortical thickness estimation using MRI. *Med. Image Anal.* 13 (2), 269–285 (Apr).
- Seghier, M.L., Ramakrishnan, A., Crinion, J., Jeff, A.P., Price, C.J., 2008. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *Neuroimage* 41 (4), 1253–1266.
- Shefer, V.F., 1973. Absolute number of neurons and thickness of the cerebral cortex during aging, senile and vascular dementia, and Pick's and Alzheimer's diseases. *Neurosci. Behav. Physiol.* 6 (4), 319–324 (Jan).
- Srivastava, S., Maes, F., Vandermeulen, D., Dupont, P., Paesschen, W., Suetens, P., 2003. An automated 3D algorithm for neo-cortical thickness measurement. *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, pp. 488–495.
- Tang, H., Wu, E., Ma, Q., Gallagher, D., Perera, G., Zhuang, T., 2000. MRI brain image segmentation by multi-resolution edge detection and region selection. *Comput. Med. Imaging Graph.* 24 (6), 349–357.
- Thambisetty, M., Wan, J., Carass, A., An, Y., Prince, J., Resnick, S.M., 2010. Longitudinal changes in cortical thickness associated with normal aging. *NeuroImage* 52 (4), 1215–1223.
- Thompson, P.M., Lee, A.D., Dutton, R.A., Geaga, J.A., Hayashi, K.M., Eckert, M.A., Bellugi, U., Galaburda, A.M., Korenberg, J.R., Mills, D., 2005. Abnormal cortical complexity and thickness profiles mapped in Williams syndrome. *J. Neurosci.* 25 (16), 4146–4158.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15 (1), 273–289.
- Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999a. Automated model-based bias field correction of MR images of the brain. *IEEE Trans. Med. Imaging* 18 (10), 885–896.
- Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999b. Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imaging* 18 (10), 897–908.
- Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 2003. A unifying framework for partial volume segmentation of brain MR images. *IEEE Trans. Med. Imaging* 22 (1), 105–119.
- von Economo, C., 1929. *The Cytoarchitectonics of the Human Cerebral Cortex*. Oxford University Press, H. Milford. (Jan).
- Wang, H., Fei, B., 2009. A modified fuzzy c-means classification method using a multiscale diffusion filtering scheme. *Med. Image Anal.* 13 (2), 193–202.
- Wells, W.M., Grimson, W.E.L., Kikinis, R., Jolesz, F.A., 1996. Adaptive segmentation of MRI data. *IEEE Trans. Med. Imaging* 15 (4), 429–442.
- Woolrich, M., Behrens, T., 2006. Variational Bayes inference of spatial mixture models for segmentation. *IEEE Trans. Med. Imaging* 25 (10), 1380–1391 (Oct).
- Yeo, B.T.T., Sabuncu, M.R., Desikan, R., Fischl, B., Golland, P., 2008. Effects of registration regularization and atlas sharpness on segmentation accuracy. *Med. Image Anal.* 12 (5), 603–615 (Oct).
- Zhang, J., 1992. The mean field theory in em procedures for Markov random fields. *IEEE Trans. Signal Process.* 40 (10), 2570–2583.
- Zhang, Y., Brady, M., Smith, S.M., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20 (1), 45–57.
- Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palmer, A.C., 1994. Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE Trans. Med. Imaging* 13 (4), 716–724 (Jan).