

Adaptive Neonate Brain Segmentation

M. Jorge Cardoso¹, Andrew Melbourne¹, Giles S. Kendall²,
Marc Modat¹, Cornelia F. Haggmann², Nicola J. Robertson²,
Neil Marlow², and Sebastien Ourselin¹

¹ Centre for Medical Image Computing (CMIC), University College London, UK

² Academic Neonatology, EGA UCL Institute for Women's Health, London, UK

Abstract. Babies born prematurely are at increased risk of adverse neurodevelopmental outcomes. Recent advances suggest that measurement of brain volumes can help in defining biomarkers for neurodevelopmental outcome. These techniques rely on an accurate segmentation of the MRI data. However, due to lack of contrast, partial volume (PV) effect, the existence of both hypo- and hyper-intensities and significant natural and pathological anatomical variability, the segmentation of neonatal brain MRI is challenging. We propose a pipeline for image segmentation that uses a novel multi-model Maximum a posteriori Expectation Maximisation (MAP-EM) segmentation algorithm with a prior over both intensities and the tissue proportions, a B0 inhomogeneity correction, and a spatial homogeneity term through the use of a Markov Random Field. This robust and adaptive technique enables the segmentation of images with high anatomical disparity from a normal population. Furthermore, the proposed method implicitly models Partial Volume, mitigating the problem of neonatal white/grey matter intensity inversion. Experiments performed on a clinical cohort show expected statistically significant correlations with gestational age at birth and birthweight. Furthermore, the proposed method obtains statistically significant improvements in Dice scores when compared to the a Maximum Likelihood EM algorithm.

1 Introduction

Preterm birth (below 32 weeks gestation) is associated with significant cognitive and neuromotor impairments, the frequency and severity of which increase with decreasing gestational age [1]. Much research has been focused on the development of biomarkers of neurological injury in an attempt to identify babies at increased risk of adverse neurodevelopmental outcomes. Recently an approach combining magnetic resonance imaging and volumetric analysis has demonstrated a correlation with neurodevelopmental outcome at 2 years [2]. Fundamental to performing volumetric and morphometric studies is the ability to classify different brain tissues.

In contrast with adults, neonatal brain MRI is complex to segment due to lack of contrast, partial volume (PV) effect, the existence of hypo- and hyper-intensities and substantial natural and pathological anatomical variability. Dynamic natural developmental variability arises due to the receding germinal

matrix and proceeding myelination from posterior to anterior. Cortical complexity also increases as the cerebral grey matter surface becomes more convoluted. Signal intensities of grey and white matter (GM/WM) on neonatal T1 and T2 MRI are characteristically reversed and contrast between the two tissue types is progressively lost during the first few months of life as a result of the myelination process resulting in low contrast-to-noise (CNR) and signal-to-noise ratio (SNR) images. Recent developments in neonatal brain segmentation have used template atlases [3] to mitigate the low SNR and CNR of the images. These methods rely on the registration of templates in order to propagate knowledge of a particular population to a patient specific space. This is problematic in pathological cases, as their anatomy may be markedly different from the one derived from a normal population atlas (see figure 1). Zhuang et al. [4] proposed a simpler intensity based classification method based on K-means and augmented by a population atlas and Yu et al. [5] proposed a Parzen windows based Hidden Markov Random Field algorithm, optimised by an Expectation-Maximisation algorithm, again augmented by a population atlas. Both these methods will have problems in severely diseased patients as the propagation of the templates by image registration may not be able to cope with the variability. Furthermore, the contrast inversion causes a layer of PV corrupted voxels on the CSF/GM boundary to have similar intensities to pure WM. Xue et al.[6] tries to ameliorate this problem through morphological operations without implicitly modelling PV.

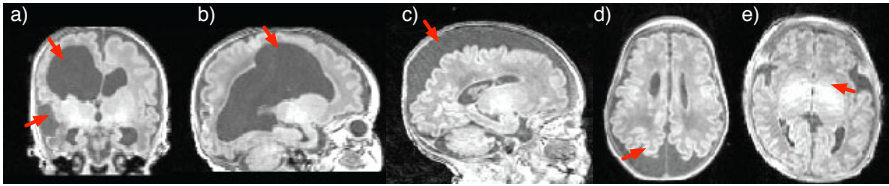


Fig. 1. Red arrows pointing to the problematic areas. Pathological areas that differ from the normal population in patients with a) and b) ventriculomegaly, c) and d) excessive sub-arachnoid CSF. e) Movement artefacts and hyper-intensities.

In this paper, we propose a new segmentation pipeline incorporating a novel *Maximum a Posteriori* Expectation-Maximization (MAP-EM) based probabilistic segmentation technique that includes intensity non-uniformity (INU) correction, spatial dependence via a Markov Random Field (MRF) and PV containing voxels. The pipeline also iteratively relaxes normal population priors, thus enabling their adaptation to pathological cases. The segmentation results are then used to iteratively update the registration from the template space to the subject, improving the prior alignment, the segmentation and the skull-stripping.

2 Methods

2.1 MAP Expectation-Maximization Segmentation

Assuming either single modality or coregistered multimodal datasets, let $i \in \{1, 2, \dots, n\}$ index the n voxels of an image domain, with its intensities forming

a feature vector $y_i \in \mathcal{R}^m$. For K tissue types, let z_i denote the tissue type of voxel i . Thus, $z_i = e_k$ 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. Let each tissue type be described as having multivariate normally distributed intensities with mean vector $\mu_k \in \mathcal{R}^m$ and respective covariance matrix $\Lambda_k \in \mathcal{R}^{m \times m}$ grouped in $\theta_k = \{\mu_k, \Lambda_k\}$, corrupted by a multiplicative bias field. In this, work, the bias field is modelled as a linear combination $c' \phi$ of J smoothly varying basis functions $\phi(x) = \{\phi(x)_1, \dots, \phi(x)_J\}$, where x denotes the spatial position and $c = \{c_1, \dots, c_m\}'$ with c_m as a vector of coefficients for channel m . For mathematical convenience and similarly to [7] and [8], we assume that the intensity of the voxels that belong to class k are normally distributed after log transformation. 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.

Let $\Phi_y = \{\theta_1, \theta_2, \dots, \theta_K, C\}$ represent the overall model parameters. Assuming an *a priori* distribution over these parameters, the problem can be formalised as a *Maximum a Posteriori* (MAP) estimation of Φ_y by

$$\hat{\Phi}_y = \arg \max_{\Phi} f(y | \Phi_y) g(\Phi_y) \tag{1}$$

with $f(y | \Phi_y) = \prod_i \prod_k f(y_i | z_i = e_k, \Phi_y) f(z_i = e_k | p_{\mathcal{N}_i} \Phi_z)$ and $g(\Phi_y)$ as a semi-conjugate gaussian prior over the mean. These priors are used to constrain the space of solutions of the mean, increasing the algorithm's robustness in complex segmentation cases. Here, the parameter μ_k , defined as

$$g(\mu_k) = \left((2\pi)^{\frac{m}{2}} |\Lambda_k^*|^{\frac{1}{2}} \right)^{-1} e^{-\frac{1}{2}(\mu_k - \mu_k^*)' (\Lambda_k^*)^{-1} (\mu_k - \mu_k^*)} \tag{2}$$

with $|\Lambda_k^*|$ as the determinant of Λ_k^* and μ_k^* and Λ_k^* as priors over the parameter μ_k . The need for this prior over the means will become apparent in Section 2.2.

Here, $f(y | \Phi_y)$ can be seen as a mixture of multivariate normal distributions, thus $f(y_i | z_i = e_k, \Phi_y) = G_{\Lambda_k}(y_i - \mu_k - c' \phi(x_i))$, where $G_{\Lambda_k}(x)$ denotes a multivariate zero-mean normal distribution with standard deviation Λ_k , defined as $G_{\Lambda_k}(x) = ((2\pi)^{\frac{m}{2}} |\Lambda_k|^{\frac{1}{2}})^{-1} e^{-\frac{1}{2} x' (\Lambda_k)^{-1} x}$. Thus, the *Maximum a Posteriori* (MAP) parameters for Φ_y can be found by maximisation of Eq.1 giving the following update equations for the model parameters:

$$\begin{aligned} \mu_k^{(t+1)} &= \left(\mu_k^* \Lambda_k^{*-1} + (\sum_{i=1}^n p_{ik} \bar{y}_{ik}) \Lambda_k^{(t)-1} \right) \left(\Lambda_k^{*-1} + (\sum_{i=1}^n p_{ik}) \Lambda_k^{(t)-1} \right)^{-1} \\ \Lambda_k^{(t+1)} &= \sum_{i=1}^n p_{ik} \left(\bar{y}_{ik} - \mu_k^{(t+1)} \right)' \left(\bar{y}_{ik} - \mu_k^{(t+1)} \right) \left(\sum_{i=1}^n p_{ik} \right)^{-1} \end{aligned}$$

where $\bar{y}_{ik} = y_i - (c^{(t)})' \phi(x_i)$ and

$$p_{ik}^{(t+1)} = f(y_i | z_i = e_k, \Phi_y^{(t)}) f(z_i = e_k) \left(\sum_{j=1}^K f(y_i | z_i = e_j, \Phi_y^{(t)}) f(z_i = e_j) \right)^{-1}$$

is the responsibility at the index i and class k and t denotes the iteration number. The estimation of $c^{(t+1)}$ under a multimodel scheme is provided in [9].

Spatial Regularisation: Due to the low signal and contrast to noise ratio of the images under analysis, the assumption of statistical independence between neighbouring voxels makes the segmentation prone to noise and image artefacts. A spatial smoothness term by means of a Markov Random Field (MRF) is thus included. This MRF is a non-binary extension of a multiclass Potts model, and assumes that the probability that voxel i belongs to tissue k depends on its first-order neighbours \mathcal{N}_i .

Using the mean field approximation as described in [10] and [9], all previous equations still hold by setting $f(z_i = e_j) = f(z_i = e_k | p_{\mathcal{N}_i} \Phi_z, \pi_{ik})$ in Eq.2.1, where

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

Here, $U_{\text{MRF}}(z_i | p_{\mathcal{N}_i}, \Phi_z)$ is an energy function dependent on the parameters $\Phi_z = \{G_{kj}, \beta\}$, where β is the overall MRF strength. Under anisotropic voxel size, the interaction between neighbours is dependant on a connection strength factor s , 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. Under this framework,

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

Relaxation of the Anatomical Priors: Due to anatomical variability and similarly to [11], we assume that the proportions π_{ik} are not known *a priori*. Instead, we consider π_{ik} as a sample drawn from a distribution derived from the statistical atlas, i.e., we consider them as a posterior of a Dirichlet distribution. Here, π_{ik} is updated at each iteration by

$$\pi_{ik} = (1 - \alpha)\omega_{ik} + \alpha(\mathcal{G}_{\sigma_{Der}} * p_{ik}) \quad (3)$$

with $\mathcal{G}_{\sigma_{Der}}$ as a gaussian kernel with standard deviation σ_{Der} and $*$ as the convolution operation, weighted by α . This updating scheme iteratively spatially relaxes π_{ik} , enabling the segmentation of pathological cases. The amount of relaxation is dependent on the parameter α controlling the mixing proportions, and σ_{Der} controlling the amount of regularisation over the posterior. For $\alpha = 0$, only the anatomical priors are used. These anatomical priors ω that incorporate probabilistic information derived from a digital brain atlas are added to the model in order to condition the posterior probabilities and indirectly also condition the model parameters. These atlases are brought into correspondence using an affine registration [12] followed by a free-form non-rigid registration algorithm [13].

Assuming skull stripped images, we 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), cerebrospinal fluid (CSF), pons (Pon) and cerebellum (Cer) respectively at every voxel position. The anatomical priors provided by [14] are used here. The time point in the 4D atlas is chosen depending on the gestational age of the neonate under analysis.

2.2 Segmentation Pipeline

The proposed pipeline can be described as a sequence of registrations and segmentations, that acts to improve the atlas alignment, the propagation of the brain mask and consequently the final segmentation. First, the population template is registered to the patient space and both the mask and the anatomical priors are propagated using the same deformation field. Due to the low signal to noise ratio and architectural variability of the images (shown in Figure 1) the registration is not able to capture the correct deformation field on the first pass. However, at this stage, the registration is able to align the more anatomically consistent structures (skull, scalp) despite performing quite badly in situations with excessive sub-arachnoid CSF or with marked ventricular expansion.

After the first registration step, the image is segmented with the above described MAP-EM segmentation algorithm with very broad parameters for the prior relaxation. At this stage, we use $\sigma_{Der} = 6$ and $\alpha = 0.5$, enabling a great adaptation of the priors. Because the prior relaxation is an iterative process, in patients with marked anatomical disparity the tissue classes might not converge to the desired solution at this stage. For example, in Fig.2(left), the WM class converges to a gaussian with large standard deviation including both CSF and WM and the INU correction tries to compensate the intensity difference. In order to ameliorate this problem, the images are intensity normalised using a robust max-min (assuming 2 % outliers) after masking. The expected distribution of the parameter μ for each class was then obtained by manually sampling several 3x3x3 voxel cubic patches over a set of intensity normalised images. From this sampling process, the mean and standard deviation of the mean intensity over each patch was calculated and assigned to the prior μ_k^* and Λ_k^* . On multimodal images, the process is done separately for each modality as Λ_k^* is assumed to be diagonal. The inclusion of this extra term adds some *a priori* knowledge about the class specific distribution of the parameter μ

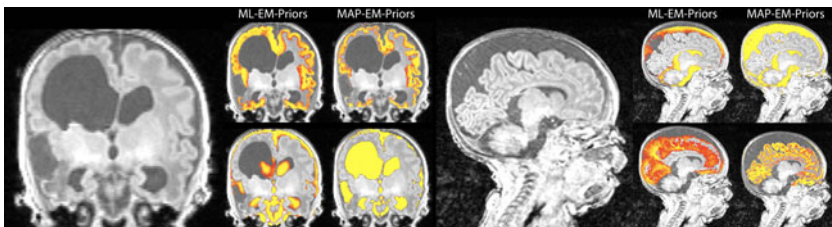


Fig. 2. Patients with either marked ventricular expansion (left) or excessive sub-arachnoid CSF (right) and their corresponding cortical GM and CSF segmentations, with and without the prior over the mean intensities.

increasing the robustness of its estimation. After the segmentation, the registration is updated by matching the brain prior (sum of all the brain structures excluding CSF) with the current brain segmentation. This allows a much less noisy and simplified registration process. Both the brain mask and the anatomical priors are re-propagated using the new deformation field. The images are segmented again using the new priors and mask, and the values of σ_{Der} and α are reduced to 2 and 0.3 respectively, resulting in a more prior driven and less adaptive segmentation.

Explicit PV modelling: At this stage, the segmentation is generally good and the priors are now correctly aligned. However, there is still a layer of WM classified voxels on the GM/CSF, Cerebellum/CSF and Pons/CSF interfaces (see Fig.3). 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 [15]. Ruan *et al.*[16] showed that, for brain imaging and for the signal-to-noise ratio levels of 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, and similarly to [17], we use the previously estimated values of p_{ik} , μ_k and σ_k to initialise the mixed classes.

Under the assumption of Gaussian distributed classes on log-transformed data, the initial PV class Gaussian parameters can be approximated by a mixel distribution [18], with mean equal to the arithmetic weighted average of its composing class parameters weighted by the determinant of the covariance matrix of each class. Thus, $\mu_{j/k}^* = |\Lambda_j| \mu_j + |\Lambda_k| \mu_k$. Due to the multiplicative nature of the probabilities, the mixed class prior is generated as the normalised geometric mean of its composing tissue distributions p_{ij} and p_{ik} .

$$\omega_{i(j/k)}^* = \sqrt{p_{ij} p_{ik}} \quad (4)$$

and

$$\omega_{i(WM)}^* = \omega_{i(WM)} \left(1 - \sqrt{\omega_{i(GM)} \omega_{i(CSF)}}\right) \quad (5)$$

This transformation will reduce the *a priori* probability for PV containing voxels to belong to WM. The priors are then normalised in order to sum to one at each voxel position. Even though these new priors cannot be considered as priors in a strict sense as they are derived from patient specific data, they behave as such in the model. After these patient specific PV priors are created, the image is segmented again using the above described MAP-EM algorithm with the 3 extra PV containing classes (GM/CSF, Cerebellum/CSF and Pons/CSF interfaces), resulting in a much reduced amount of PV containing voxels classified as WM. Please refer to Fig. 3 for a visual assessment of the segmentation steps and their improvements.

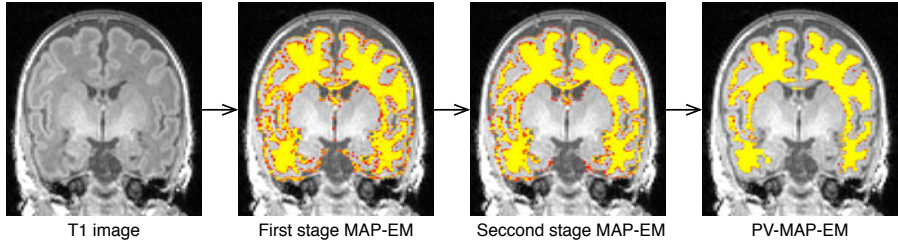


Fig. 3. From left to right: Segmentation pipeline showing the original image, the first segmentation step, the second segmentation step with the improved prior alignment and finally the PV corrected segmentation result .

3 Validation

Data: The data were acquired on a 1.5T Siemens’s Avanto using $TR = 17ms$, $TE = 6ms$ and flip angle of 21° . In total 43 T1-weighted volumes are analysed with resolution $0.39 \times 0.39 \times 1mm$. The mean gestational age is 27.1 ± 2.7 weeks (range 23.1 – 32.3), mean birthweight $970 \pm 373g$ (range 540 – 2470g) and mean age at scan 40.4 ± 1.74 (range 35.7 – 44.3). The male to female ratio is 37/36.

Clinical correlation: For each segmentation we analyse correlations with GA at birth and birthweight. We additionally correlate the grey/white matter ratio. The correlation coefficient required to achieve a p-value ≤ 0.01 for this sample size (43) is $r = 0.36$. With the exception of the CSF volume, each of the segmented components correlates significantly ($r > 0.5$) with both gestational age at birth and birthweight. The absence of correlation with CSF is expected due to significant natural and pathological variation in ventricle size. There are no correlations with gender.

Quantitative and qualitative analysis: In order to assess the quality of the segmentation, the proposed segmentation was compared to manually segmented ventricles. Two subsets were chosen from the full database: one containing 15 neonates with normal anatomy and another containing 4 patients with noticeable anatomical differences (2 cases with ventriculomegaly and 2 with excessive sub-arachnoid CSF). For each dataset, a Dice score was calculated between the proposed method and the manual segmentation. The proposed algorithm was also compared to the Maximum Likelihood Expectation Maximisation (ML-EM) algorithm [9]. On the subset containing the 15 normal subjects, the the proposed method obtains a dice score of 0.908 ± 0.034 when compared to 0.725 ± 0.217 for the ML-EM, representing a statistically significant improvement ($p < 0.01$) in the accuracy of the segmentation. On the second subset, the dice scores for the proposed method are 0.92, 0.86, 0.93 and 0.94, and respectively 0.22, 0.15, 0.35, and 0.55 for the ML-EM method, showing a marked improvement in the accuracy of the segmentation in pathological cases.

4 Conclusions

We have presented a segmentation algorithm tailored specifically for neonatal T1 segmentation, using a MAP-EM algorithm with a new prior relaxation strategy combined with a semi-conjugate prior over the intensities and an implicit PV model in order to mitigate the problematic misclassified layer of PV voxels. Statistically significant correlations are shown between structural volumes and both weight at birth and gestational age. Furthermore, experiments performed on a clinical cohort also show significant improvements in segmentation accuracy, mainly in pathological cases, when compared to a ML-EM algorithm.

References

1. Marlow, N., Wolke, D., Bracewell, M.A., Samara, M.: Neurologic and developmental disability at six years of age after extremely preterm birth. *New England Journal of Medicine* 352(1), 9–19 (2005)
2. Boardman, J., Craven, C., Valappil, S., Counsell, S., Dyet, L., Rueckert, D., Aljabar, P., Rutherford, M., Chew, A., Allsop, J., Cowan, F., Edwards, A.: A common neonatal image phenotype predicts adverse neurodevelopmental outcome in children born preterm. *NeuroImage* 52(2), 409–414 (2010)
3. Weisenfeld, N.I., Warfield, S.K.: Automatic segmentation of newborn brain mri. *NeuroImage* 47(2), 564–572 (2009)
4. Song, Z., Awate, S., Licht, D., Gee, J.: Clinical neonatal brain mri segmentation using adaptive nonparametric data models. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I. LNCS*, vol. 4791, pp. 883–890. Springer, Heidelberg (2007)
5. Yu, X., Zhang, Y., Lasky, R.E., Parikh, N.A., Narayana, P.A.: Comprehensive brain mri segmentation in high risk preterm newborns. *PLoS ONE* 5(11) (2010)
6. Xue, H., Srinivasan, L., Jiang, S., Rutherford, M., Edwards, A.D., Rueckert, D., Hahn, J.V.: Automatic segmentation and reconstruction of the cortex from neonatal mri. *NeuroImage* 38(3), 461–477 (2007)
7. Wells, W., Grimson, W.E., Kikinis, R., Jolesz, F.A.: Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging* 15(4), 429–442 (1996)
8. Zhang, Y., Brady, M., Smith, S.M.: Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging* 20(1), 45–57 (2001)
9. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. *IEEE TMI* 18(10) (1999)
10. Zhang, J.: The mean field theory in em procedures for markov random fields. *IEEE Transactions on Signal Processing* 40(10), 2570–2583 (1992)
11. Shiee, N., Bazin, P.L., Cuzzocreo, J.L., Blitz, A., Pham, D.L.: Segmentation of brain images using adaptive atlases with application to ventriculomegaly
12. Ourselin, S., Roche, A., Prima, S., Ayache, N.: Block matching: A general framework to improve robustness of rigid registration of medical images. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) *MICCAI 2000. LNCS*, vol. 1935, pp. 557–566. Springer, Heidelberg (2000)
13. Modat, M., Ridgway, G., Taylor, Z., Lehmann, M., Barnes, J., Hawkes, D., Fox, N., Ourselin, S.: Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine* (October 2009)

14. Kuklisova-Murgasova, M., Aljabar, P., Srinivasan, L., Counsell, S.J., Gousias, I.S., Boardman, J.P., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D.: A dynamic 4d probabilistic atlas of the developing brain. *NeuroImage* 54(4) (2011)
15. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: A unifying framework for partial volume segmentation of brain MR images. *IEEE Transactions on Medical Imaging* 22(1), 105–119 (2003)
16. Ruan, S., Jaggi, C., Fadili, J., Bloyet, D.: Brain tissue classification of magnetic resonance images using partial volume modeling. *IEEE TMI* 19(12) (December 2000)
17. Cardoso, M.J., Clarkson, M.J., Ridgway, G.R., Modat, M., Fox, N.C., Ourselin, S., The Alzheimer's Disease Neuroimaging Initiative: LoAd: A locally adaptive cortical segmentation algorithm. *NeuroImage* 56(3), 1386–1397 (2011)
18. Kitamoto, A., Takagi, M.: Image classification using probabilistic models that reflect the internal structure of mixels. *Pattern Analysis and Applications* 2 (1999)