Bayesian Model Selection for Pathological Data

Carole H. Sudre[1,2], Manuel Jorge Cardoso[1,2], Willem Bouvy[3], Geert Jan Biessels[3], Josephine Barnes[2], and Sébastien Ourselin[1,2]

[1] Translational Imaging Group, CMIC, University College London, NW1 2HE, UK
[2] Dementia Research Centre, UCL Institute of Neurology, London, WC1N 3BG, UK
[3] Department of Neurology and Neurosurgery, UMC Utrecht, Netherlands

Abstract. The detection of abnormal intensities in brain images caused by the presence of pathologies is currently under great scrutiny. Selecting appropriate models for pathological data is of critical importance for an unbiased and biologically plausible model fit, which in itself enables a better understanding of the underlying data and biological processes. Besides, it impacts on one's ability to extract pathologically meaningful imaging biomarkers. With this aim in mind, this work proposes a fully unsupervised hierarchical model selection framework for neuroimaging data which permits the stratification of different types of abnormal image patterns without prior knowledge about the subject's pathological status.

1 Introduction

Measures of pathological load visible on MRI and comparison with healthy tissues can be used to ascertain clinical correlations and infer disease progression [6]. As the presence of pathology leads to unexpected observations (i.e. outlier intensities), from a modelling perspective, two main problems arise: first, bias is introduced in the model parameters' estimation by the pathological outliers when segmenting non-pathological tissues [3], [12]; second, there is a need for prior knowledge in order to design better pathology-specific segmentation algorithms. Due to pathology specific tuning and the reliance on knowledge-based heuristic rules [14], [7], these methods are not easily transposed from one pathology to another [1]. Furthermore, multiple types of outliers might be present [16].

Using a finite weighted sum of Gaussian components, also known as a Gaussian mixture model (GMM), is probably the most widespread way of modelling the distribution of observed intensities in neuroimaging data [2]. Optimising the model parameters is usually performed through the Expectation-Maximisation algorithm (EM) introduced in [4]. Due to the presence of pathology, an *a priori* fixed number of Gaussian components might not be appropriate to tackle the problems related to the presence of pathology-linked intensities [5]. Moreover, as underlined in [10] and [15], natural appearing features such as the cerebrospinal fluid (CSF) are not well explained by a single Gaussian component due to the high variability of their contents.

This work proposes a Bayesian inference criterion (BIC) regularised adaptive hierarchical Gaussian mixture model selection framework, enabling an automated selection of the number of components necessary to model both the

P. Golland et al. (Eds.): MICCAI 2014, Part I, LNCS 8673, pp. 323–330, 2014.

outlier (pathologies such as lesions, tumours or vasculature) and the inlier part (normal anatomical tissues such as white matter (WM), grey matter (GM), CSF or non-brain (NB)) of the observed data. The model selection framework exploits a split-and-merge strategy (SM) [9] for model search, which combined with the BIC, ensures a balance between computational burden, model fit and model complexity. Additional spatial knowledge is introduced through the use of probabilistic tissue atlases and a Markov Random Field (MRF) [14]. After describing the data generative model and the model selection and inference strategies, we present results illustrating the generality of the proposed method both on simulated and clinical data.

2 Methods

2.1 General Modelling and EM Algorithm

The proposed framework models the intensity observations as a hierarchical mixture of Gaussian and uniform components. In the following, $\mathcal{Y}$ denotes the set of N log-transformed intensity vectors of dimension d, $\{\mathbf{y}_1, \cdots, \mathbf{y}_n, \cdots \mathbf{y}_N\}$, where N and d are respectively the number of voxels and the number of available modalities (T1w, T2w, PD, etc). The data likelihood for model $\mathbf{K}$ is defined as $F(\mathcal{Y}|\boldsymbol{\Theta}_{\boldsymbol{K}})$, where $\boldsymbol{\Theta}_{\mathbf{K}}$ represents the model parameters. More specifically, we propose a three level hierarchical model: a first level (l) characterising if an observation is an outlier or an inlier, a second level (j) characterising its tissue class (i.e. if an inlier/outlier voxel belongs to WM, GM, CSF or NB) and a third level (k) characterising the multiple intensity clusters/distributions of each inlier or outlier tissue class. In order to robustly model the data, one can separate the first level of the model into two density functions I and O, that correspond respectively to the inlier part, modelling the healthy tissues, and to the outlier part, related to the unexpected tissues, such that $F(\mathbf{y}_n|\boldsymbol{\Theta}_{\boldsymbol{K}}) = b_I \cdot I(\mathbf{y}_n|\boldsymbol{\Theta}_{\mathbf{K}}) + b_O \cdot O(\mathbf{y}_n|\boldsymbol{\Theta}_{\mathbf{K}})$ with $b_I + b_O = 1$. In this scenario, previously proposed models have classically assumed a uniform distribution for O [17]. At the second level of the hierarchical model, the distribution is $F(\mathbf{y}_n|\boldsymbol{\Theta}_{\mathbf{K}}) = \sum_{l\in I,O} b_l \sum_{j=1}^{J} a_{l_j} \Phi\left(\mathbf{y}_n\middle|\boldsymbol{\Theta}_{l_j}\right)$ where b_l, a_{l_j} and $\Phi\left(y_n\middle|\boldsymbol{\Theta}_{l_j}\right)$ are respectively the mixing weight for l, the class weight for l_j and the likelihood of the data at voxel n for the tissue class l_j. In the third level of this multi-layered model, indexed by k, each anatomical class density distribution is finally modelled by a mixture of Gaussian and/or uniform components such that $\Phi\left(\mathbf{y}_n\middle|\Theta_{l_j}\right) = \sum_{k=1}^{K_{l_j}} w_{l_{j_k}} \mathcal{G}\left(\mathbf{y}_n\middle|\theta_{l_{j_k}}\right) + w_{l_{j_{\text{unif}}}} \mathcal{U}_{l_j}$ where K_{l_j} is the number of Gaussian components in class l_j, $w_{l_{j_k}}$ is the mixing proportion of class l_{j_k}, $\mathcal{G}$ is a Gaussian density distribution and $\mathcal{U}$ is a uniform distribution. The parameters of the Gaussian distribution are $\theta_{l_{j_k}} = \left\{\boldsymbol{\mu}_{l_{j_k}}, \Lambda_{l_{j_k}}\right\}$, where $\boldsymbol{\mu}$ and Λ are respectively the multivariate mean and covariance matrix. Assuming *iid* observations, the multi-layered mixture model can finally be written as follows:

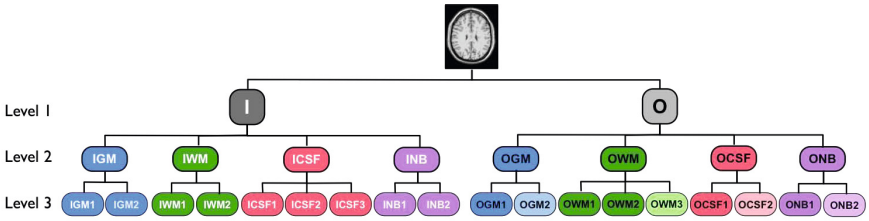

Fig. 1. Example of hierarchical model with 4 main tissue classes (GM, WM, CSF and NB). The lighter colored components follow a uniform distribution.

$$F\left(\mathcal{Y}|\boldsymbol{\Theta}_{\mathbf{K}}\right)=\prod_{i=1}^{N}\sum_{l\in I,O}\sum_{j=1}^{J}\left[\sum_{k=1}^{K_{l_j}}b_l a_{l_j} w_{l_{j_k}}\mathcal{G}\left(\mathbf{y}_n\middle|\theta_{l_{j_k}}\right)+b_l a_{l_j} w_{l_{j_{\text{unif}}}}\right].$$

The final model structure is depicted in Figure 1 where both the inlier and outlier versions of the GM, WM, CSF and NB classes are separated into their sub-clusters. The EM algorithm is used within this setting to optimise the model parameters. Introducing the labelling configuration set $\mathcal{Z}=\{\mathbf{z}_1,\cdots,\mathbf{z}_n,\cdots,\mathbf{z}_N\}$, where $\mathbf{z}_n$ is a unity vector of dimension $\sum_l\sum_j K_{l_j}$, with one component being equal to 1 and all the others to 0, the EM algorithm alternates between the expectation of the complete data log-likelihood $\mathbb{E}_{\boldsymbol{\Theta}_{\mathbf{K}}^{(t)}}\left[\log\left(F\left(\mathcal{Y}|\mathcal{Z},\boldsymbol{\Theta}_{\mathbf{K}},\boldsymbol{\pi}\right)F\left(\mathcal{Z}|\boldsymbol{\pi}\right)\right)\right]$ at iteration t, also known as the E-step, and the maximisation of this function with respect to the different parameters, also known as the M-step, where $\boldsymbol{\pi}=\{\boldsymbol{\mathcal{B}},\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{W}}\}$ are the sets of weights attributed to the different components of the mixture levels. At this point, three main types of parameters have to be optimised: first, the parameters of each Gaussian distribution $\theta_{l_{j_k}}$; second, the contribution $w_{l_{j_k}}$ of each distribution to the overall observation model; and third and most importantly, the number of Gaussian components K_{l_j} necessary to describe the underlying distribution of tissue class l_j.

2.2 Spatial and Smoothness Priors

Probabilistic tissue atlases are used to introduce knowledge in the generative model about the location of each tissue class l_j. As b_l is initialised to a global value, $\boldsymbol{\pi}$ has to be normalised to $\boldsymbol{\pi}=\{\tilde{\boldsymbol{\mathcal{B}}},\tilde{\boldsymbol{\mathcal{A}}},\boldsymbol{\mathcal{W}}\}$, so that $\forall n,\sum_{j=1}^{J}\tilde{a}_{nj}=1$ and $\sum_{l\in\{I,O\}}\tilde{b}_{nl}=\sum_{l\in\{I,O\}}\tilde{b}_l=1$. The priors $\tilde{\boldsymbol{\mathcal{B}}}$ and $\tilde{\boldsymbol{\mathcal{A}}}$ follow a Dirichlet distribution as in [11].

In addition, a MRF is used to spatially regularise the labelling l_{j_k} as in [11,14]. This work defines a symmetric MRF energy matrix H of size $K_{l_j}\times K_{l_j}$ containing the neighborhood clique energies, such that

$$H(l'_{j'k'},l_{jk})=H(l_{jk},l'_{j'k'})=\begin{cases}0 & \text{if } j=j'\\ 0.2 & \text{if } (j,j')=(\text{WM},\text{CSF}) \text{ or } (\text{G/WM},\text{NB})\\ 0.1 & \text{otherwise}\end{cases}$$

Let $\mathbf{p}_m$ be the set of responsabilities for voxel m, updated with Bayes' Rule during the E-step. Under a mean field approximation, with $\mathcal{N}_n$ the set of neighbours of voxel n, the labelling $\mathcal{Z}$ can be defined as:

$$F\left(\mathcal{Z}|H,\boldsymbol{\pi}\right) \propto \prod_{n=1}^{N} \prod_{l\in I,O} \prod_{j=1}^{J} \prod_{k=1}^{K_{l_j}} \left[\tilde{b}_{nl}\tilde{a}_{nl_j} w_{l_{j_k}}\right]^{z_{nl_{j_k}}} \exp\left(-\mathbf{z}_n^t H \sum_{m\in\mathcal{N}_n} \mathbf{p}_m\right).$$

2.3 Model Selection

The flexibility of the proposed model lies in the automatic selection of the appropriate number of components K_{l_j} needed to model each l_j class. Ideally, this parameter could be optimised using a Markov Chain Monte Carlo algorithm. However, due to the computational complexity of such an approach, here, a split-and-merge (SM) algorithm is used for model selection [9,13]. In this framework, a merge operation consists of transforming two close enough Gaussian distributions of the same l_j class, $l_{j_{k_1}}$ and $l_{j_{k_2}}$, into a single component l_{j_k}. A split operation is the transformation of a single distribution, Gaussian or Uniform, into two subcomponents. The symmetric Kullback-Leibler Divergence (KLD) is used to define which component(s) should be split (largest KLD compared to the observations) or merged (smallest KLD between the components' distributions). The SM algorithm alternates between a split and a merge operation given the current model estimates. Parameters' initialisation when splitting or merging Gaussian distribution(s) follows the strategy used in [9]. As no closed-form solution exists when splitting a uniform distribution, a 2-class k-means algorithm is used to estimate the 2 sub-clusters of the samples $\mathcal{U}_{l_j}$. The mean and covariance of the cluster with the smallest variance is used to initialise a new Gaussian class. Finally, the proposed method is optimised using an iterative conditional modes (ICM) approach, where it switches between the optimisation of the model parameters and the model selection. In order to provide a bias-variance trade-off between accuracy and complexity of the model, the Bayesian inference criterion for model $\boldsymbol{K}$, expressed as $\text{BIC}(\boldsymbol{K}) = \kappa \log\left(F\left(\mathcal{Y}|\boldsymbol{\Theta}_{\boldsymbol{K}}\right)\right) - C(\boldsymbol{K})$ is used as an objective function. It penalises the log-likelihood of the model according to the cost function $C(\boldsymbol{K}) = \left[\sum_l \sum_j K_{l_j}\left(\frac{(d+1)d}{2}+1\right) - j\right] \cdot \log(N\cdot\kappa)$ that depends on the number of free parameters to optimise. Here, κ is a correction factor accounting for the proportion of voxels considered as independent [8]. For each ICM iteration, the model evolves given the most probable model. If the selected model fails to increase the objective function $\text{BIC}(\boldsymbol{K})$ after convergence of the EM, the next most probable model is tested. The SM search stops when all models have been tested.

Implementation Details. In this implementation, the weight for the inlier uniform class is set to 0, as inlier classes are expected to follow a GMM. Also, in order to avoid instability in the inference strategy, the Dirichlet priors are only updated when the initial model converges. The code and further implementation details will be made publicly available.

Table 1. DSC and KLD when using EMS and BaMoS for different modality combinations and lesion loads on the Brainweb MS model

		Mild		Moderate		Severe	
	Modalities	EMS	BaMoS	EMS	BaMoS	EMS	BaMoS
DSC (%)	T1T2PD	4.13	53.31	28.77	54.8	53.78	65.45
	T1T2	2.43	50.94	16.57	52.08	35.28	63.82
	T1PD	13.93	24.63	50.44	59.67	69.24	78.77
	T2PD	3.74	11.58	27.42	39.50	53.43	64.42
KLD(×10)	T1T2PD	6.10	4.21	6.16	4.12	6.13	4.08
	T1T2	0.95	0.29	0.95	0.28	0.93	0.28
	T1PD	1.98	0.46	1.99	0.46	1.99	0.45
	T2PD	3.03	1.01	3.05	1.00	3.05	0.99

3 Validation

3.1 Simulated Images - Brainweb

The simulated multiple sclerosis (MS) lesion model provided by Brainweb `http://brainweb.bic.mni.mcgill.ca`, with ground truth lesion segmentations, was used to evaluate the performance of the proposed model, here denoted as BaMoS (Bayesian Model Selection). All combinations of the 3 available modalities (T1w, T2w, PD) on the 3 lesion loads (mild, moderate, severe), at 3% of noise and without magnetic field inhomogeneity were assessed. In this experiment, the parameter b_O was initialised to 0.01. The proposed method was compared with the classical lesion segmentation method by van Leemput *et al.* [14], here denoted as EMS, with the parameter 3 for the Mahalanobis distance threshold and MRF parameters as defined in Sec 2.2. The outlier components O_{j_k} of the proposed model with $\mu_{O_{j_k}} > \mu_{I_{WM}}$ and $j = \text{WM}$, for the T2 and PD modalities, were automatically selected as lesion-related intensity clusters. These MS lesion clusters are then added together to form the total lesion segmentation. The Dice similarity coefficient (DSC) and the KLD between the modelled distribution and the observations, here used to compare respectively the binary lesion segmentation overlap and the quality of the model fit, are gathered in Table 1, showing the general improvement brought by BaMoS. Figure 2 depicts the different intensity clusters associated with MS lesions. Note that each lesion component is related to the lesion density in the ground truth.

3.2 Clinical Data - Type 2 Diabetes

Type 2 Diabetes (T2DB) patients typically present white matter hyperintensities (WMH) that correlate with cognitive decline. Segmentation of such variable outliers is needed to study their evolution and their clinical correlates. The behavior of BaMoS on clinical data with multiple types of outliers was evaluated using data with WMH from the MICCAI MRBrainS2013 challenge. For this study, brain images from T2DB patients and controls (age > 50) were acquired on a

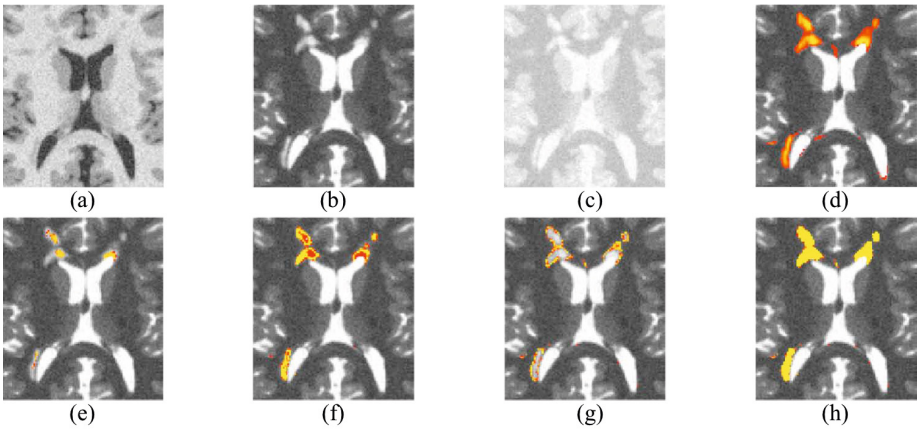

Fig. 2. Top) A zoomed-in section of the T1w (a), T2w (b) and PD (c) simulated Brainweb images with severe MS lesion load followed by the ground truth lesion segmentation (d). Bottom) The three automatically extracted probabilistic lesion maps (e-g), followed by their sum (h) overlayed on the T2 image.

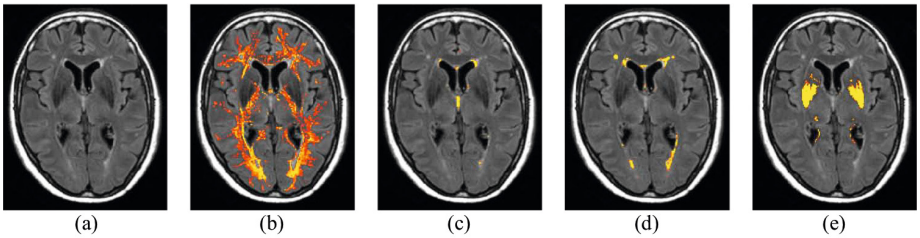

Fig. 3. Components of interest obtained using BaMoS. From left to right) FLAIR image (a), inlier component of WM related to penumbra (b), lesion-related WM outliers (c - d) and WM outlier component localising cysteine-iron complex deposition in the globus pallidus (e).

3T Philips scanner. Multi-slice T2-FLAIR images (0.958 × 0.958 × 3 mm) and T1w 3D registered images were used. Further details about the acquisition and preprocessing can be found at `http://mrbrains13.isi.uu.nl`. Manual WMH segmentation was performed on 16 FLAIR images. BaMoS was run using the parameters detailed in Sec. 3.1 after extraction of the brain tissues and registration of the statistical atlases to the T1w image. The results obtained for BaMoS and EMS are gathered in Table 2. A 1-tailed t-test showed a significant improvement in terms of both model fit (p-value<0.01) and DSC (p-value<0.05) for BaMoS compared to EMS. The R^2 measure of correlation with the manually segmented WMH's volumes was improved with BaMoS (0.938) compared to EMS (0.887). Figure 3 presents an example of different components of interest separated by BaMoS. BaMoS automatically found separate clusters for both

Table 2. DSC and KLD when comparing EMS and BaMoS performance on WMH segmentation for the studied T2DB and control cases

Case	1	2	4	5	6	7	8	9	10	11	13	15	17	18	19	20
DSC (%)																
EMS	24.2	5.4	40.9	47.4	69.2	5.9	45.7	7.3	20.0	55.5	33.3	24.7	41.7	52.9	3.7	9.3
BaMoS	57.6	8.0	67.8	64.2	80.3	0	69.1	4.0	19.4	49.1	32.0	0	56.4	75.3	4.3	15.3
KLD ($\times 10$)																
EMS	1.23	1.06	1.59	1.36	1.22	1.34	1.17	1.08	1.57	1.30	1.11	1.27	1.23	1.39	1.38	1.25
BaMoS	0.67	0.84	0.69	0.86	0.73	0.71	0.88	0.65	0.76	0.72	0.69	0.84	0.83	2.21	0.76	0.74

WM lesion-related hyperintensities and WM hypointensities, strengthening the generalisability argument of the proposed method. The lesion-penumbra related subclass of the WM inlier class might be of interest to study WMH evolution and apparition. In low performance cases, corresponding to milder hyperintensities, the information in the data and model complexity constraints did not support the existence of an extra outlier lesion-related subclass. Finally, we noted the stability in the number of components for each inlier class of the selected model across all subjects.

4 Discussion and Conclusion

This work presents an automated model selection strategy that differentiates between different types of outliers. The main interest of the developed outliers modelling strategy resides in its generality. Indeed, its performance in terms of detection of specific types of outliers, such as MS lesions or white matter hyperintensities, is comparable to other methods specifically tuned towards this goal. However, the proposed method can also simultaneously model other types of outliers (e.g. vessels, iron deposition, etc). This property is of major interest in neurodegenerative disease studies, since different pathological changes might be present concomitantly. Furthermore, the ability to distinguish different lesion densities and their spatial location might be of further interest to help characterise the underlying pathophysiological process. Further work will investigate the balance between the different model parameters, their relationship with the image characteristics and the contribution of the different used modalities.

Acknowledgements. This work was supported by the Wolfson Foundation, the Faculty of Engineering Science UCL, EPSRC (EP/W046410/1, EP/H046410/1, EP/J020990/1, EP/K005278), the MRC (MR/J01107X/1), the EU-FP7 project VPH-DARE@ IT (FP7-ICT-2011-9-601055), NIHR Queen Square Dementia Biomedical Research UK, the NIHR UCLH/UCL Biomedical Research Centre (High Impact Initiative) and Alzheimer's Research UK (ARUK). The Dementia Research Centre is supported by ARUK, Brain Research Trust and the Wolfson

Foundation. The authors would like to thank the staff and patients of the Vascular Cognitive Impairment Group at UMC Utrecht without whom collection and generation of the diabetes-related data would not be possible.

References

1. Admiraal-Behloul, F., van den Heuvel, D., Olofsen, H., van Osch, M., van der Grond, J., van Buchem, M., Reiber, J.: Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. NeuroImage 28 (2005)
2. Balafar, M.A.: Gaussian mixture model based segmentation methods for brain MRI images. Artificial Intelligence Review (2012)
3. Battaglini, M., Jenkinson, M., De Stefano, N.: Evaluating and reducing the impact of white matter lesions on brain volume measurements. HBM 33, 2062–2071 (2012)
4. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM Algorithm. J. R. Stat. Soc. B 39(1), 1–38 (1977)
5. Fraley, C., Raftery, A.E.: Model-based clustering, discriminant analysis, and density estimation. J. Am. Statist. Assoc. 97(458), 611–631 (2002)
6. García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D.L., Collins, D.L.: Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Medical Image Analysis 17, 1–18 (2013)
7. García-Lorenzo, D., Prima, S., Douglas, A.L., Collins, D.L., Barillot, C.: Trimmed-likelihood estimation for focal lesions and tissue segmentation in multisequence mri for multiple sclerosis. IEEE TMI 30(8), 1455–1467 (2011)
8. Groves, A.R., Beckmann, C.F., Smith, S.M., Woolrich, M.W.: Linked independent component analysis for multimodal data fusion. NeuroImage 54(3) (2011)
9. Li, Y., Li, L.: A split and merge EM algorithm for color image segmentation. In: IEEE ICIS 2009, vol. 4, pp. 395–399 (2009)
10. Sajja, B.R., Datta, S., He, R., Mehta, M., Gupta, R.K., Wolinsky, J.S., Narayana, P.A.: Unified approach for multiple sclerosis lesion segmentation on brain MRI. Ann. Biomed. Eng. 34(1), 142–151 (2006)
11. Shiee, N., Bazin, P.-L., Cuzzocreo, J., Blitz, A., Pham, D.L.: Segmentation of brain images using adaptive atlases with application to ventriculomegaly. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 1–12. Springer, Heidelberg (2011)
12. Shroeter, P., Vesin, J.M., Langenberger, T., Meuli, R.: Robust parameter estimation of intensity distributions for brain magnetic resonance images. IEEE TMI 17(2), 172–186 (1998)
13. Ueda, N., Nakano, R., Ghahramani, Z., Hinton, G.E.: SMEM Algorithm for Mixture Models. Neural Comput. 12(9), 2109–2128 (2000)
14. Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE TMI 20(8), 677–688 (2001)
15. Wolff, Y., Miron, S., Achiron, A., Greespan, H.: Improved CSF classification and lesion detection in MR brain images with multiple sclerosis. In: SPIE, vol. 6512 (2007)
16. Wu, Y., Warfield, S.K., Tan, I.L., Wells III, W.M., Meier, D.S., van Schijndel, R.A., Barkhof, F., Guttmann, C.R.: Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. NeuroImage 32, 1205–1215 (2006)
17. Zhuang, X., Huang, Y., Palaniappan, K., Zhao, Y.: Gaussian mixture density modeling, decomposition, and applications. IEEE TIP 5(9), 1293–1302 (1996)