

MR Brain Tissue Classification Using an Edge-Preserving Spatially Variant Bayesian Mixture Model

G. Sfikas^{1,3}, C. Nikou¹, N. Galatsanos², and C. Heinrich³

¹ University of Ioannina, Department of Computer Science, 45110 Ioannina, Greece

² University of Patras, Department of ECE, 26500 Rio, Greece

³ LSIIT-University of Strasbourg, 67412 Illkirch cedex, France

Abstract. In this paper, a spatially constrained mixture model for the segmentation of MR brain images is presented. The novelty of this work is an edge-preserving smoothness prior which is imposed on the probabilities of the voxel labels. This prior incorporates a line process, which is modeled as a Bernoulli random variable, in order to preserve edges between tissues. The main difference with other, state of the art methods imposing priors, is that the constraint is imposed on the probabilities of the voxel labels and not onto the labels themselves. Inference of the proposed Bayesian model is obtained using variational methodology and the model parameters are computed in closed form. Numerical experiments are presented where the proposed model is favorably compared to state of the art brain segmentation methods as well as to a spatially varying Gaussian mixture model.

1 Introduction

The segmentation of 3D brain magnetic resonance (MR) images into the three main tissues, namely, white matter (WM), gray matter (GM) and cerebro-spinal fluid (CSF) is of great importance in most neuroimaging applications. Although many research studies have been presented in this area, MRI brain segmentation still remains a challenging issue due to specific difficulties of MRI, such as intensity inhomogeneity, partial volume effect and acquisition noise. A first approach to the problem relied on the expectation-maximization (EM) algorithm [1,2] which led to an important category of methods resorting to Gaussian mixture models (GMM). Among them, many studies incorporate prior information (e.g. anatomical atlases) on tissue topology [3,4,5,6] or constrain the segmentation to be spatially smooth and take into account edge discontinuities (e.g. using Markov random field (MRF) priors) [7,8,9,10,11,12,13,14,15].

Modeling the probability density function (pdf) of pixel or voxel attributes with a GMM [16] is a natural way to cluster data because it automatically provides a grouping based on the components of the mixture that generated them. Furthermore, the likelihood of a GMM is a rigorous metric for clustering performance. The parameters of the GMM can be estimated very efficiently through maximum likelihood (ML) estimation using the Expectation-Maximization (EM) algorithm [16].

The prior knowledge that adjacent pixels most likely belong to the same cluster is not used in standard GMM. To overcome this shortcoming, spatial smoothness constraints may be imposed, generally applying an MRF prior, like the spatially varying Gaussian mixture model (SVGMM) in [17]. However, this model enforces smoothness between pixels belonging to different classes. Since the seminal work in [18], *line processes* were also introduced in several applications, other than brain tissue classification, to respond to this drawback, see for instance [19] and [20] for image restoration and superresolution respectively.

In this paper we propose a new, Bayesian, spatially varying Gaussian mixture model for the classification of brain images to the three tissue types (WM, GM, CSF). The main contribution of the model is that it takes into account not only that adjacent voxels are more probable to belong to the same class but it also prohibits smoothing across boundary voxels. Motivated by the studies in brain image segmentation incorporating MRF-based prior knowledge [5,7,8,9,10,11,12,13,14,15] we impose proper hyperpriors to simultaneously address local smoothing and edge preservation. The main difference with other, state of the art methods imposing MRF-type priors, is that the constraint is imposed on the probabilities of the voxel labels (generally known in mixture modeling as *contextual mixing proportions*) and not onto the labels themselves. By these means, closed form solutions are provided for the model parameters through variational inference.

2 The Bayesian Edge Preserving Spatially Varying GMM

The K-kernel spatially varying GMM differs from the standard GMM in the definition of the mixing proportions. More precisely, in the SVGMM, each voxel x^n , $n = 1, \dots, N$ has a distinct vector of mixing proportions denoted by π_j^n , $j = 1, \dots, K$. We call these parameters *contextual mixing proportions* to distinguish them from the mixing proportions of a standard GMM. Hence, the probability of a distinct voxel is expressed by:

$$f(x^n; \pi, \mu, \Lambda) = \sum_{j=1}^K \pi_j^n \mathcal{N}(x^n; \mu_j, \Lambda_j) \quad (1)$$

where $0 \leq \pi_j^n \leq 1$, $\sum_{j=1}^K \pi_j^n = 1$ for $j = 1, 2, \dots, K$ and $n = 1, 2, \dots, N$, μ_j are the Gaussian kernel means and Λ_j are the Gaussian kernel precision (inverse covariance) matrices.

We now assume that the voxels $X = \{x^1, x^2, \dots, x^N\}$ are independent and generated by the graphical model shown in figure 1.

Note that a set $Z = \{z_j^n\}_{n=1..N, j=1..K}$ of $N \times K$ latent variables is introduced, in order to make inference tractable for the model. The Z variables are distributed multinomially:

$$p(Z|\Pi) = \prod_{j=1}^K \prod_{n=1}^N (\pi_j^n)^{z_j^n} \quad (2)$$

where each z^n is a binary vector, with $z_j^n = 1$ if datum n is generated by the j -th kernel and $z_j^n = 0$ otherwise.

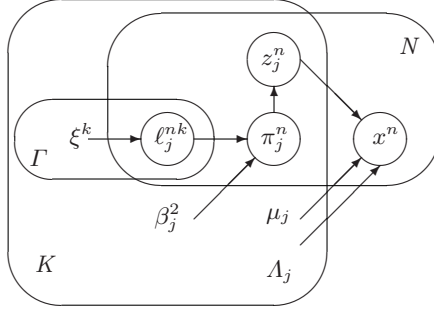


Fig. 1. Graphical model for the edge preserving spatially variant Gaussian mixture model. Superscript $n \in [1, N]$ denotes voxel index, subscript $j \in [1, K]$ denotes kernel (segment) index and $\Gamma \in [1, 26]$ describes the neighbourhood direction type.

Considering the set of *contextual mixing proportions* Π as random variables and assuming a proper prior, we can incorporate the intuitive fact that neighbouring voxels are more likely to share the same class label. We impose an edge preserving Gaussian prior on Π :

$$p(\Pi|\ell; \beta) = \prod_{j=1}^K \prod_{n=1}^N \prod_{k=1}^{\Gamma} \mathcal{N}(\pi_j^n | \pi_j^k; \beta_j^2)^{\ell^{nk}} \quad (3)$$

where ℓ^{nk} is a binary random variable we call *line-process*. If $\ell^{nk} = 1$, then there is a link on the random field between the voxel indexed n and its k -th *possible* neighbour (we denote by π_j^k the k -th neighbour of specific voxel n). Otherwise, if $\ell^{nk} = 0$ there is no link between them, signifying the presence of an edge. We assume that two voxels can be possible neighbours when they are vertically, horizontally or diagonally adjacent with regard to their spatial location in the three-dimensional mesh, implying $\Gamma = 26$ neighbours per voxel.

Parameters $\beta = \{\beta_1^2, \beta_2^2, \dots, \beta_K^2\}$ control the spatial smoothness of the *contextual mixing proportions*. The prior in eq. (3) implies that

$$\pi_j^n - \pi_j^k \sim \mathcal{N}(0, \beta_j^2), \quad \forall k \in [1, \Gamma] \mid \ell^{nk} = 1$$

reflecting the fact that the *contextual mixing proportions* which implicitly control voxel class membership, are similar for neighbouring voxels except in case there exists an edge.

We now regard the *line process* variables ℓ^{nk} as Bernoulli distributed random variables, governed by a parameter set $\xi = \{\xi^1, \xi^2, \dots, \xi^\Gamma\}$:

$$p(\ell|\xi) = \prod_{n=1}^N \prod_{k=1}^{\Gamma} p(\ell^{nk}|\xi^k) = \prod_{n=1}^N \prod_{k=1}^{\Gamma} \xi^k \ell^{nk} (1 - \xi^k)^{(1 - \ell^{nk})} \quad (4)$$

The Beta distribution is the conjugate for the Bernoulli pdf, therefore, we impose it on the ξ parameters:

$$p(\xi; \alpha_{\xi 0}, \beta_{\xi 0}) = \prod_{k=1}^{\Gamma} \frac{\Gamma(\alpha_{\xi k 0} + \beta_{\xi k 0})}{\Gamma(\alpha_{\xi k 0})\Gamma(\beta_{\xi k 0})} \xi^k (\alpha_{\xi k 0} - 1) (1 - \xi^k)^{(\beta_{\xi k 0} - 1)} \quad (5)$$

The main advantage of the model in fig. 1 is that (i) it takes into account that neighboring voxels are more probably generated by the same Gaussian pdf and (ii) it does not smooth adjacent voxels separated by an edge.

3 MAP Estimation Using Variational Inference

To perform segmentation, the evidence with respect to model parameters has to be optimized:

$$\operatorname{argmax}_{\mu, \Lambda, \Pi, \beta} \ln p(X, \Pi; \mu, \Lambda, \beta)$$

This MAP solution cannot be computed directly, or even estimated using the EM algorithm, due to the Π prior distribution complexity. Therefore, we resort to variational inference [16]. This leads to an iterative scheme with one step for the computation of the stochastic parameters Z , ℓ , ξ and Π and one step for the deterministic parameters μ , Λ and β . Due to lack of space we present here the final expressions.

The expected values of the stochastic parameter are

$$\begin{aligned} \langle z_j^n \rangle &= \tilde{\pi}_j^n, \quad \langle l^{nk} \rangle = \tilde{\xi}^{nk}, \quad \langle \ln \xi_k \rangle = \psi(\alpha_{\xi k}) - \psi(\alpha_{\xi k} + \beta_{\xi k}), \\ \langle \ln(1 - \xi_k) \rangle &= \psi(\beta_{\xi k}) - \psi(\alpha_{\xi k} + \beta_{\xi k}), \end{aligned}$$

where $\psi(\cdot)$ is the digamma function and the expectations (denoted by a tilde) being as follows, with $\operatorname{sig}(x) = (1 + e^{-x})^{-1}$:

$$\begin{aligned} \tilde{\pi}_j^n &= \frac{\pi_j^n \mathcal{N}(x^n; \mu_j, \Lambda_j)}{\sum_{l=1}^K \pi_l^n \mathcal{N}(x^n; \mu_l, \Lambda_l)}, \\ \tilde{\xi}^{nk} &= \operatorname{sig} \left(\sum_{j=1}^K \ln \mathcal{N}(\pi_j^k | \pi_j^n; \beta_j^2) + \langle \ln \xi_k \rangle - \langle \ln(1 - \xi_k) \rangle \right), \\ \tilde{\alpha}_{\xi k} &= \alpha_{\xi 0} + \sum_{n=1}^N \langle l^{nk} \rangle, \quad \tilde{\beta}_{\xi k} = \beta_{\xi 0} + \sum_{n=1}^N \langle 1 - l^{nk} \rangle. \end{aligned}$$

The *contextual mixing proportions* π_j^n are computed as the roots of a quadratic equation:

$$a_j^n (\pi_j^n)^2 + b_j^n (\pi_j^n) + c_j^n = 0 \quad (6)$$

with coefficients:

$$a_j^n = - \sum_{k=1}^{\Gamma} \langle l^{nk} \rangle, \quad b_j^n = \sum_{k=1}^{\Gamma} \langle l^{nk} \rangle \pi_j^k, \quad c_j^n = \frac{\langle z_j^n \rangle \beta_j^2}{2}.$$

The form of the coefficients guarantees that there is always a non negative solution. The solutions of eq. (6) however will not in general satisfy the constraint $\sum_{j=1}^K \pi_j^n = 1, \pi_j \geq 0, \forall j \in [1..K]$ so we project the corresponding π^n vectors $\forall n \in [1..N]$ onto the constraints subspace; this is done using the quadratic programming algorithm described in [21].

Furthermore, the deterministic parameters are also obtained in closed form:

$$\tilde{\mu}_j = \frac{\sum_{n=1}^N \langle z_j^n \rangle x^n}{\sum_{n=1}^N \langle z_j^n \rangle}, \quad \tilde{\Lambda}_j^{-1} = \frac{\sum_{n=1}^N \langle z_j^n \rangle (x^n - \mu_j)(x^n - \mu_j)^T}{\sum_{n=1}^N \langle z_j^n \rangle} \quad (7)$$

$$\tilde{\beta}_j^2 = \frac{\sum_{n=1}^N \sum_{k=1}^{\Gamma} \langle l^{nk} \rangle (\pi_j^n - \pi_j^k)^2}{\sum_{n=1}^N \sum_{k=1}^{\Gamma} \langle l^{nk} \rangle} \quad (8)$$

The above update equations, for both the stochastic and deterministic parameters, are considered for the full range of each of the indices, namely n, j and k and are computed iteratively until convergence [22] of the variational lower bound.

4 Experimental Results

We have evaluated the proposed model on simulated images with known ground truth from the BrainWeb database [23],[24] using the voxel intensities as features.

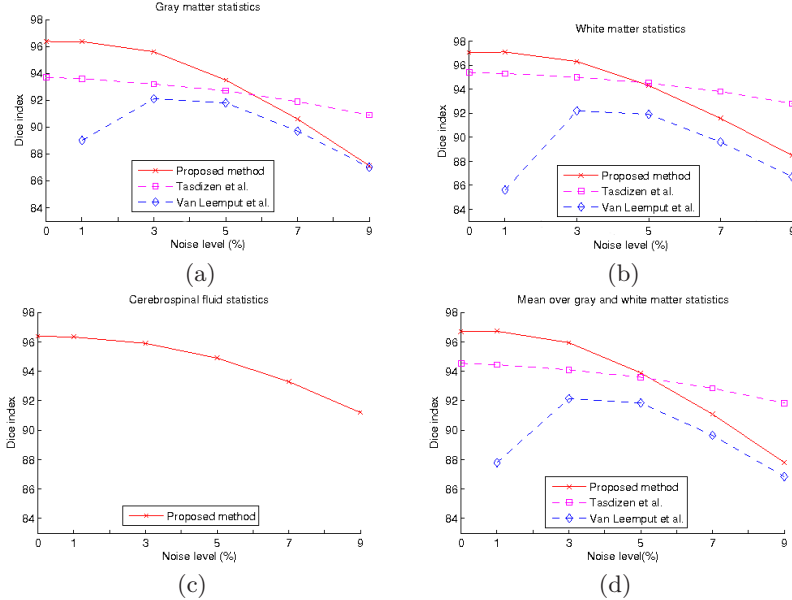


Fig. 2. Dice metric as a function of noise level for data without bias field. (a) Gray matter, (b) white matter, (c) CSF, (d) mean over gray and white matter. The dashed lines plot the results for the same images as presented in [5] and [10]. Results are not provided by the respective publications for the case in (c).

Table 1. Mean values for the Dice metric as a function of noise level over the three tissue types. The compared methods are the proposed method (Bayes-SVGMM), the non edge preserving spatially varying GMM (SVGMM) proposed in [17] and a standard GMM.

Noise (%)	<i>Bayes-SVGMM</i>	<i>SVGMM</i>	<i>GMM</i>
0	96.6	96.5	77.2
1	96.6	96.5	89.8
3	96.0	95.9	95.3
5	94.2	94.2	94.1
7	91.8	91.7	90.8
9	88.9	88.8	86.8

Prior to segmentation, we have preprocessed each volume so that only WM, GM and CSF are included (fig. 2). Hence, we set the number of kernels in our model to $K = 3$. The hyperparameter values of the Beta prior distribution were set to $\alpha_{\xi k 0} = \beta_{\xi k 0} = 1, \forall k$, making the prior uninformative as the data size $N \gg 1$.

The algorithm was applied to a simulated T1-weighted data without any bias field and with intensity noise levels between 0% and 9%. The noise percentages were defined with respect to the mean intensity of each tissue class. We have compared our segmentation results with two of the state of the art methods of Van Leemput *et al.* [10] and Tasdizen *et al.* [5]. In both of these studies, the Dice metric was used for evaluation. Therefore we present our results using this performance measure. Figure 2 summarizes the Dice metrics for the compared methods. In that figure, the curves for the state-of-the-art methods are reproduced from the respective publications [10,5]. As it can be observed, in all cases, our method provides better segmentations with respect to the method in [10]. Also, for low level noise the Dice metric of the proposed method is higher with respect to the method proposed in [5]. On the other hand, the method of Tasdizen *et al.* [5] performs better for noise levels of 7% and 9%. However, our method takes no more than 50 minutes to run on a 2.7 GHz standard PC whereas the method in [5] requires at least *six* hours runtime for convergence.

We have also compared our Bayesian model to a standard GMM as well as to the spatially varying GMM (SVGMM) proposed in [17]. In all cases, our model performed better than both methods. Both of the spatially varying models

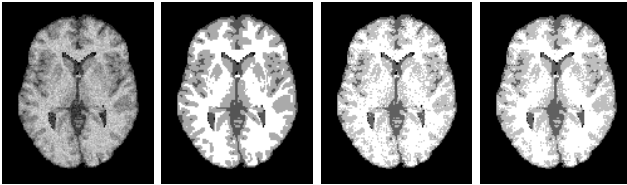


Fig. 3. Axial slices of a 3D segmentation example. From left to right: original MR slice with (9%) noise, ground truth, segmentation using a GMM, and the proposed method.

provided Dice metrics significantly better than the standard GMM. Moreover, the difference between our model and the SVGMM [17] is small but consistently in favor of our method. These differences are underpinned in table 1. A representative segmentation example is presented in figure 3.

5 Conclusion

We have presented a framework for segmenting the brain anatomy from 3D MRI. The proposed method relies on a Bayesian finite mixture model with a Gauss-Markov random field prior on the probabilities of the pixel labels. Also, the model incorporates a probabilistic line process for edge preservation. The quantitative evaluation reveals that the method not only improves the standard GMM and refines the SVGMM [17] but also performs at least at the same level as other state-of-the-art methods. A perspective of this study is the extension of the model to include more brain tissues and to integrate bias field correction into the segmentation procedure.

References

1. Kapur, T., Grimson, W.E.L., Wells III, W.M., Kikinis, R.: Segmentation of brain tissue from magnetic resonance images. *Medical Image Analysis* 1, 109–127 (1996)
2. Wells III, W., Grimson, W., Kikinis, R., Jolesz, F.: Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging* 15, 429–442 (1996)
3. Marroquin, J., Vemuri, B., Botello, S., Calderon, F., Fernandez-Bouzas, A.: An accurate and efficient Bayesian method for automatic segmentation of brain MRI. *IEEE Transactions on Medical Imaging* 21, 934–945 (2002)
4. Prastawa, M., Gilmore, J., Lin, W., Gerig, G.: Automatic segmentation of neonatal brain MRI. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) *MICCAI 2004*. LNCS, vol. 3216, pp. 10–17. Springer, Heidelberg (2004)
5. Tasdizen, T., Awate, S., Whitaker, R., Foster, N.: MRI tissue classification with neighborhood statistics: a nonparametric entropy-minimizing approach. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005*. LNCS, vol. 3750, pp. 517–525. Springer, Heidelberg (2005)
6. Pohl, K., Bouix, S., Nakamura, M., Rohfling, T., McCarley, R., Kikinis, R., Grimson, L., Shenton, M., Wells, W.: A hierarchical algorithm for MR brain image parcellation. *IEEE Transactions on Medical Imaging* 26, 1201–1212 (2007)
7. Niessen, W.J., Vincken, K.L., Weickert, J., Viergever, M.A.: Three dimensional MR brain segmentation. In: *Proceedings of the 6th International Conference on Computer Vision (ICCV 1998)*, pp. 53–58 (1998)
8. Laidlaw, D.H., Fleischer, K.W., Barr, A.H.: Partial-volume Bayesian classification of material mixtures in MR volume data using voxel histograms. *IEEE Transactions on Medical Imaging* 17(1), 74–86 (1998)
9. Kapur, T., Grimson, W.E.L., Kikinis, R., Wells III, W.M.: Enhanced spatial priors for segmentation of magnetic resonance imagery. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) *MICCAI 1998*. LNCS, vol. 1496, pp. 457–468. Springer, Heidelberg (1998)

10. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging* 18, 897–908 (1999)
11. Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging* 20, 677–688 (2001)
12. Greenspan, H., Ruf, A., Goldberger, J.: Constrained Gaussian mixture model framework for automatic segmentation of MR brain images. *IEEE Transactions on Medical Imaging* 25, 1233–1245 (2006)
13. Peng, Z., Wee, W., Lee, J.H.: Automatic segmentation of MR brain images using spatial-varying Gaussian mixture and Markov random field approach. In: *Proceedings of the Computer Vision and Pattern Recognition Workshop (CVPRW 2006)* (2006)
14. Awate, S.P., Tasdizen, T., Foster, N., Whitaker, R.: Adaptive Markov modeling for mutual information based, unsupervised MRI brain tissue classification. *Medical Image Analysis* 10, 726–739 (2006)
15. Awate, S.P., Zhang, H., Gee, J.C.: A fuzzy, nonparametric segmentation framework for DTI and MRI analysis: With applications to DTI tract extraction. *IEEE Transactions on Medical Imaging* 26, 1525–1536 (2007)
16. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer, Heidelberg (2006)
17. Nikou, C., Galatsanos, N., Likas, A.: A class-adaptive spatially variant mixture model for image segmentation. *IEEE Transactions on Image Processing* 16(4), 1121–1130 (2007)
18. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(6), 721–741 (1984)
19. Molina, R., Mateos, J., Katsaggelos, A., Vega, M.: Bayesian multichannel image restoration using compound Gauss-Markov random fields. *IEEE Transactions on Image Processing* 12, 1642–1654 (2003)
20. Kanemura, A., Maeda, S., Ishii, S.: Hyperparameter estimation in Bayesian image superresolution with a compound Markov random field prior. In: *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2007)*, Thessaloniki, Greece (2007)
21. Blekas, K., Likas, A., Galatsanos, N., Lagaris, I.: A spatially constrained mixture model for image segmentation. *IEEE Transactions on Neural Networks* 16(2), 494–498 (2005)
22. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
23. Collins, D., Zijdenbos, A., Kollokian, V., Sled, J., Kabani, N., Holmes, C., Evans, A.: Design and construction of a realistic digital brain phantom. *IEEE Transactions on Medical Imaging* 17(3), 463–468 (1998)
24. Kwan, R., Evans, A., Pike, G.: MRI simulation-based evaluation of image-processing and classification methods. *IEEE Transactions on Medical Imaging* 18, 1085–1097 (1999)