IMAGE SEGMENTATION WITH A CLASS-ADAPTIVE SPATIALLY CONSTRAINED MIXTURE MODEL

Christophoros Nikou, Nikolaos Galatsanos, Aristidis Likas and Konstantinos Blekas

University of Ioannina, Department of Computer Science,
P.O. Box 1186, 45110 Ioannina, Greece,
phone: + (30) 26510 98802
{cnikou,galatsanos,arly,kblekas}@cs.uoi.gr

ABSTRACT

We propose a hierarchical and spatially variant mixture model for image segmentation where the pixel labels are random variables. Distinct smoothness priors are imposed on the label probabilities and the model parameters are computed in closed form through maximum a posteriori (MAP) estimation. More specifically, we propose a new prior for the label probabilities that enforces spatial smoothness of different degree for each cluster. By taking into account spatial information, adjacent pixels are more probable to belong to the same cluster (which is intuitively desirable). Also, all of the model parameters are estimated in closed form from the data. The proposed conducted experiments indicate that our approach compares favorably to both standard and previous spatially constrained mixture model-based segmentation techniques.

1. INTRODUCTION

Image segmentation is the process of grouping image pixels based on the coherence of some attributes such as intensity, color or texture. Given a known number of classes, each class can be associated with a label, so that image segmentation (pixel labeling or clustering) consists of assigning a numeric label to each pixel. Many approaches have been proposed to solve the image segmentation problem [8, 15]. Methods based on grey level thresholding [10, 14] and statistical modeling of the image [4] were investigated. While the former assume absence of statistical noise and partial volume effects the latter are computationally expensive.

The relatively recent application of finite mixture models (FMM) [1, 13, 16] to image segmentation is based on the assumption that the intensity of each pixel is a sample from a finite mixture of distributions. Finite mixtures of distributions have provided a mathematical-based approach to the statistical modeling of random phenomena [11]. The parameters of the model can be estimated through likelihood maximization (ML) using the Expectation-Maximization (EM) [5] algorithm. Since pixel observations are considered to be independent samples, a significant drawback of the ML approach is that spatial information is not taken into account.

To overcome this difficulty, the spatially variant finite mixture model (SVFMM) considers a maximun *a posteriori* (MAP) approach by introducing a prior distribution for the parameters following the Gibbs function [6, 7, 17]. However, the SVFMM considers a global statistical model for the whole image. It does not take into account intra-cluster statistics which, in general may differ significantly. The smoothness constraint in image segments implied by the SVFMM may be violated not only in cases of noise and missing data but also by the nature of the data (e.g. textured images). In this study, we improve the SVFMM by introducing a new class-adaptive regularization to control the strength of the prior which is now different for each image cluster. Furthermore, all of the model parameters are computed in closed form through the MAP estimation and the EM algorithm.

In the following, the standard spatially variant finite mixture model is described in section 2 and our adaptive spatially variant mixture model (A-SVFMM) is presented in section 3. Experimental results are presented in section 4 and conclusions are drawn in section 5.

2. THE STANDARD SPATIALLY VARIANT FINITE MIXTURE MODEL

Let x^i denote the intensity of the i^{th} pixel of an image ($i = 1\dots,N$) modeled as independently distributed random variables. The SVFMM [6] provides a modification of the classical FMM approach [1, 11, 16] for pixel labeling. It assumes a mixture model with K components each one having its own vector of density parameters θ^j.

Pixel i is characterized by its probability vector $\overrightarrow{\pi}^i = \left[\pi_1^i \pi_2^i \dots \pi_K^i\right]^T$ where K is the number of clusters that a pixel may belong. $\mathbf{\Pi} = \{(\overrightarrow{\pi}^1)^T, (\overrightarrow{\pi}^2)^T, \dots (\overrightarrow{\pi}^N)^T\}$ is the set of probability vectors and $\mathbf{\Theta} = \{\theta^1, \theta^2, \dots, \theta^K\}$ the set of component parameters. The set of probabilities $\pi_j^i = P(j|x_i)$ of the i^{th} pixel to belong to the j^{th} cluster (or class label) must satisfy the constraints $0 \leq \pi_j^i \leq 1$ and $\sum_{j=1}^{K} \pi_j^i = 1$.

The FMM assumes that the density function at an observation x^i is expressed by:

$$f(x^i|\mathbf{\Pi},\mathbf{\Theta}) = \sum_{j=1}^{K} \pi_j^i \phi(x^i|\theta^j), \tag{1}$$

where $\phi(x^i|\theta^j)$ is a Gaussian distribution with parameters its mean and standard deviation $\theta^j = \{\mu_j, \sigma_j\}$. The SVFMM proposes a prior density based on the Gibbs distribution for the parameter set $\mathbf{\Pi}$:

$$p(\mathbf{\Pi}) = \frac{1}{Z} e^{-U(\mathbf{\Pi})} \tag{2}$$

with

$$U(\mathbf{\Pi}) = \beta \sum_{i=1}^{N} V_{\mathcal{N}_i}(\mathbf{\Pi}), \tag{3}$$

where Z is a normalizing constant, β is the Gibbs regularization parameter and the function $V_{\mathcal{N}_i}(\mathbf{\Pi})$ denotes the clique potential function of the pixel label vectors $\{\overrightarrow{\pi}^i\}$ within the neighborhood $\mathcal{N}_i$. In the general case, this function has the form:

$$V_{\mathcal{N}_i}(\mathbf{\Pi}) = \sum_{m \in \mathcal{N}_i} g(d_{i,m})$$

where $u_{i,m}$ specifies the distance between two label vectors $\{\overrightarrow{\pi}^i\}$ and $\{\overrightarrow{\pi}^m\}$:

$$d_{i,m} = ||\{\overrightarrow{\pi}^i\} - \{\overrightarrow{\pi}^m\}||^2 = \sum_{j=1}^{K} (\pi_j^i - \pi_j^m)^2$$

and the neighborhood $\mathcal{N}_i$ is generally the set of horizontally and vertically adjacent pixels to pixel i. A choice for the monotonically increasing and non negative penalty function is

$$g(d_{i,m}) = \frac{1}{1 + \frac{1}{d_{i,m}}}$$

which is a robust to outliers function [9].

Therefore, denoting $\mathbf{X}$ the set of pixels $\{x^i\}$, $i = 1, ..., N$, considering them to be statistically independent and following Bayes rules, we obtain the following probability density function:

$$q(\mathbf{\Pi}, \mathbf{\Theta}|\mathbf{X}) = \prod_{i=1}^{N} p(\mathbf{\Pi}) f(x^i|\mathbf{\Pi}, \mathbf{\Theta}) \tag{4}$$

with the log-density:

$$L(\mathbf{\Pi}, \mathbf{\Theta}|\mathbf{X}) = \sum_{i=1}^{N} \log f(x^i|\mathbf{\Pi}, \mathbf{\Theta}) + \log p(\mathbf{\Pi}) \tag{5}$$

The EM algorithm [5], for MAP estimation, requires the computation of the conditional expectation values, or the posterior probabilities, of the missing variables at the E-step, at iteration step t:

$$z_j^{i(t)} = \frac{\pi_j^{i(t)} \phi(x^i|\theta_j^{(t)})}{\sum_{p=1}^{K} \pi_j^{i(t)} \phi(x^i|\theta_p^{(t)})} \tag{6}$$

In the M-step, considering that the complete data log-likelihood is linear in the missing variables [5], the maximization of

$$Q_{MAP}(\mathbf{\Pi}, \mathbf{\Theta}|\mathbf{\Pi}^{(t)}, \mathbf{\Theta}^{(t)}) = \sum_{i=1}^{N} \sum_{j=1}^{K} z_j^{i(t)} \{\log(\pi_j^i) + \log\left(\phi(x^i|\theta^j)\right)\} - \beta \sum_{i=1}^{N} \sum_{m \in \mathcal{N}_i} g(u_{i,m}) \tag{7}$$

corresponding to the complete data log-likelihood, yields the model parameters. The function $Q_{MAP}(\cdot)$ in (7) can be maximized independently for each parameter with the following update equations of the mixture model parameters at step $t+1$:

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_j^{i(t)} x^{i(t)}}{\sum_{i=1}^{N} z_j^{i(t)}}, \qquad \sigma_j^{(t+1)} = \sqrt{\frac{\sum_{i=1}^{N} z_j^{i(t)} \left[x^{i(t)} - \mu_j^{(t+1)}\right]^2}{\sum_{i=1}^{N} z_j^{i(t)}}} \tag{8}$$

Several methods have been proposed to to compute the mixing proportions π_i^j of each normal density. A generalized EM scheme, based on the gradient projection method was used in [6] and a linear constrained convex quadratic programming approach was proposed in [2].

3. A FINITE MIXTURE MODEL WITH ADAPTIVE PRIOR DISTRIBUTION

The main drawback of the SVFMM is that the scalar parameter β must be fixed in Eq. (7). This parameter controls the strength of the prior $p(\mathbf{\Pi})$ with respect to the standard mixture model. Large values of β produce strongly smooth image segments while small values of the parameter result in a soft smoothness constraint.

To overcome this drawback, we propose a prior distribution incorporating information on the strength of the smoothness constraint. Each cluster j has a distinct parameter β_j which is computed in close form through the EM algorithm. More precisely, we consider a prior probability for each of the parameters $\mathbf{\Pi}$ in (5):

$$p\left(\overrightarrow{\pi}^{\mathbf{i}}\right) \propto \prod_{j=1}^{K} (\beta_j^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \frac{\sum_{m \in \mathcal{N}_i} (\pi_j^i - \pi_j^m)^2}{\beta_j^2}\right]. \tag{9}$$

This prior probability is based on the assumption that the local differences of the label probabilities are Gaussian distributed with zero mean and different variance for each cluster. Similar in spirit priors have been used to enforce smoothness in other image processing problems (e.g. image restoration [3, 12]). Hence, the parameter β_j^2 enforces spatial smoothness of different degree for each cluster j and can be seen as the variance of cluster j. Notice that different configurations of the neighborhood $\mathcal{N}_i$ for the i^{th} pixel lead to different forms of probability distribution for vector $\overrightarrow{\pi}^{\mathbf{i}}$. The prior density in eq. (9) yields the following MAP function to be maximized:

$$Q_{MAP}(\mathbf{\Pi}, \mathbf{\Theta}|\mathbf{\Pi}^{(t)}, \mathbf{\Theta}^{(t)}) = \sum_{i=1}^{N} \sum_{j=1}^{K} z_j^{i(t)} \{\log(\pi_j^i) + \log\left(\phi(x^i|\theta^j)\right)\} - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{K} \log(\beta_j^2) - \frac{1}{2} \frac{\sum_{m \in \mathcal{N}_i} (\pi_j^i - \pi_j^m)^2}{\beta_j^2} \tag{10}$$

To compute the model parameters $\pi_j^{i(t+1)}$ and $\beta_j^{2(t+1)}$ at time step $(t+1)$ we have to maximize (10) with respect to π_j^i we have to compute its partial derivative and set the result to zero. Notice that we have to take into consideration that every π_j^i in the summation term $\sum_{m \in \mathcal{N}_i} (\pi_j^i - \pi_j^m)^2$ occurs once as the probability of the central pixel and 8 times as a neighbor π_j^m of different pixels. Thus, $\frac{\partial Q_{MAP}}{\partial \pi_j^i} = 0$ gives a second degree equation with respect to $\pi_j^{i(t+1)}$:

$$16\left(\pi_j^{i(t+1)}\right)^2 - 2\pi_j^{i(t+1)} \sum_{m \in \mathcal{N}_i} \pi_j^m - z_j^{i(t)} \beta_j^{2(t)} = 0 \qquad (11)$$

or

$$\pi_j^{i(t+1)} = \frac{\sum_{m \in \mathcal{N}_i} \pi_j^m \pm \sqrt{\left(\sum_{m \in \mathcal{N}_i} \pi_j^m\right)^2 + 16 z_j^{i(t)} \beta_j^{2(t)}}}{16}, \qquad (12)$$

for $i = 1,...,N$ and $j = 1,...,K$, expressing the probability of the i^{th} pixel to belong to the j^{th} class at time $(t+1)$.

Also, the solution for the class variances are obtained by setting $\frac{\partial Q_{MAP}}{\partial \beta_j^2} = 0$ and solving for β_j^2 at time step $(t+1)$:

$$\beta_j^{2(t+1)} = \frac{1}{N} \sum_{i=1}^{N} \sum_{m \in \mathcal{N}_i} \left(\pi_j^{i(t+1)} - \pi_j^m\right)^2, \quad j = 1,...,K \qquad (13)$$

It must be noticed that the neighborhood $\mathcal{N}_i$ in expressions (12) and (13) may include pixels with updated label parameter vectors (step at time $t+1$), as well as pixels whose label vectors have not yet been updated and their value comes from step at time t of the EM algorithm.

The formulation described above can be integrated through the EM algorithm with equation (6) being the E-step and equations (8), (12) and (13) being the M-step of the algorithm. All of the unknown parameters are computed in closed form from the data.

Moreover, the parameter β_j^2 can be extended to express not only the class variance for cluster j but also the variance within cluster j at a certain spatial direction (e.g. horizontal, vertical and diagonal pixel variances). In that case, the prior probability becomes

$$p(\mathbf{\Pi}) \propto \prod_{d=1}^{D} \prod_{j=1}^{K} (\beta_{j,d}^2)^{-\frac{N}{2}} \exp\left[-\frac{1}{2} \frac{\mathbf{\Pi}^T \mathbf{Q}_d^T \mathbf{Q}_d \mathbf{\Pi}}{\beta_{j,d}^2}\right] \qquad (14)$$

where D is the total number of the considered pixel adjacencies (generally 4), $\beta_{j,d}^2$ is the variance of class j only considered for pixels having adjacency type d and $\mathbf{Q}_d$ is a first order difference operator in direction d. In that case, derivation of the expression of the log-likelihood with respect to π_j^i leads also to a second degree equation whose coefficients depend on the parameters $\beta_{j,d}^2$ and z_j^i.

4. EXPERIMENTAL RESULTS

The performance of our approach is illustrated with a number of examples. To ensure that the computed values for π_j^i in (12) satisfy the constraints

$$0 \le \pi_j^i \le 1, \quad \sum_{j=1}^{K} \pi_j^i = 1$$

we have applied the vector projection algorithm proposed in [2] in all of the techniques. Also, as the EM algorithm is sensitive to initialization, we have executed a number of iterations of the EM algorithm with a set of randomly generated initial conditions and kept the one giving the maximum value for the log-likelihood. Finally, we have not imposed any maximum number of iterations to the EM algorithm. This strategy was adopted for all of the techniques compared here.

We present a comparison between the standard finite mixture model (FMM) [1], the spatially variant finite mixture model (SVFMM) [17] improved by the optimization method we proposed in [2] and our adaptive spatially variant finite mixture model (A-SVFMM) for segmenting piecewise-constant images with a small number of classes. Figure 1 shows a simulated three-class image with intensities for the three classes 70, 90 and 110. The mixing proportions are 0.37, 0.30 and 0.33 respectively. Figure 2 shows the same image corrupted by zero mean Gaussian noise. The noise standard deviation was computed in order to achieve a signal to noise ratio of 4.0 dB , 2.0 dB and 1.0 dB repsectively.

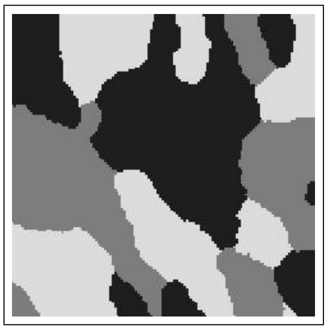


Figure 1: The 3-class test image used in the experiments described in the text. Intensity means are 70, 90 and 110. The mixing proportions are 0.37, 0.30 and 0.33 respectively.

The top row of figure 3 shows the segmentation obtained by FMM, the middle row shows the segmentation achieved by the SVFMM and the bottom row shows the segmentation by our method (A-SVFMM). The results for the A-SVFMM are more accurate than the other methods.

This is also confirmed by the percentage of correctly classified pixels which is illustrated in table 1.

	4 dB	2 dB	1 dB
FMM	90.4	81.2	76.8
SVFMM	92.6	86.7	79.4
A-SVFMM	99.2	96.4	91.3

Table 1: *Percentage of correctly classified pixels for the degraded images of figure 2. See text for technique abbreviations.*

We have to notice, that in the case of the SVFMM algorithm we have fixed the normalization parameter of the Gibbs distribution (7,3) to the same value for the three noise cases since there is no trivial method to estimate the parameter β from the data. In contrast, in the proposed approach the parameters of the prior are easily estimated and this is one of the main strengths of the proposed method. Moreover, the FMM and SVFMM techniques converge more slowly than the A-SVFMM method, especially in presence of significant

Figure 2: Noisy versions of the 3-class image of figure 1 with (a) SNR=4.0 dB, (b) 2.0 dB and (c) 1.0 dB .

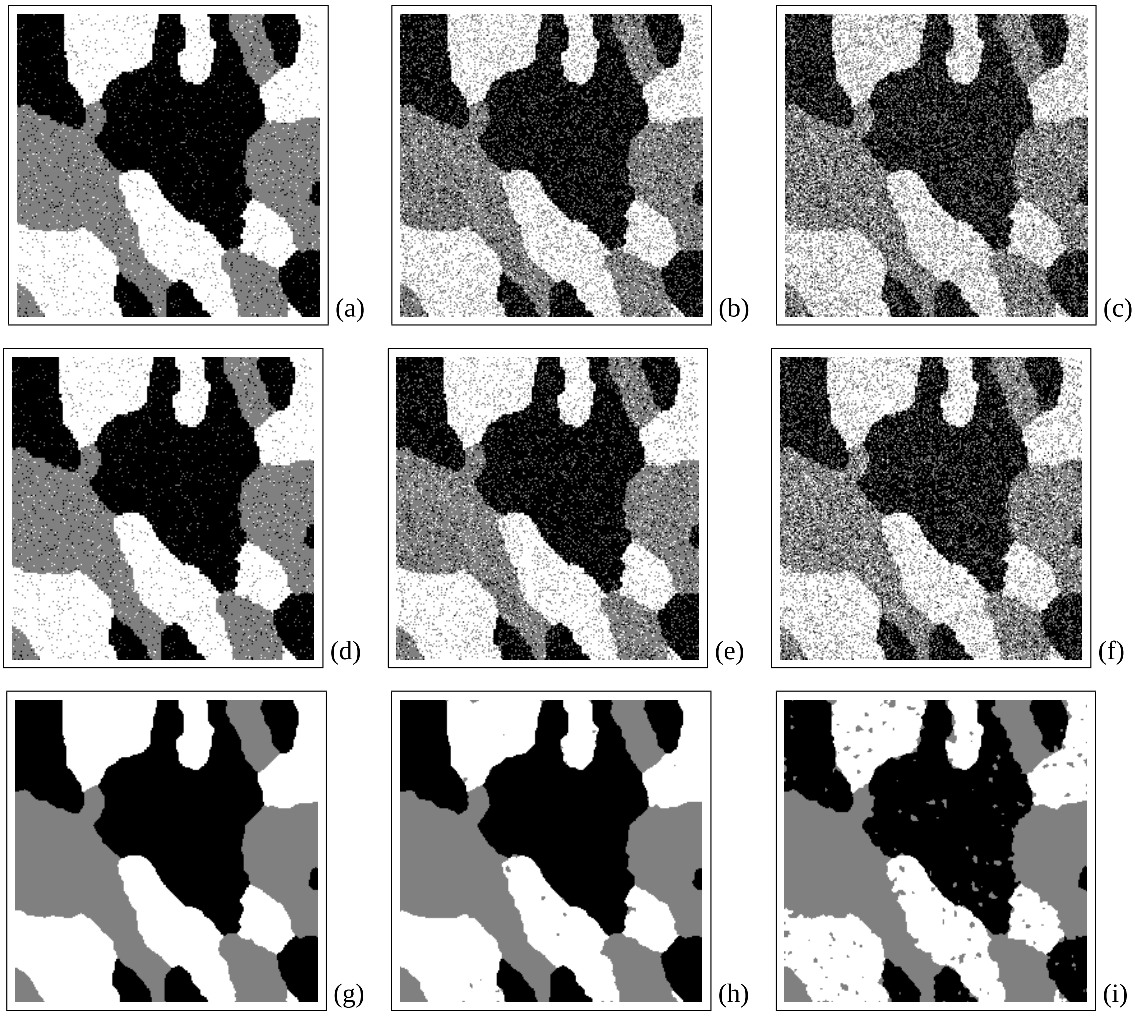


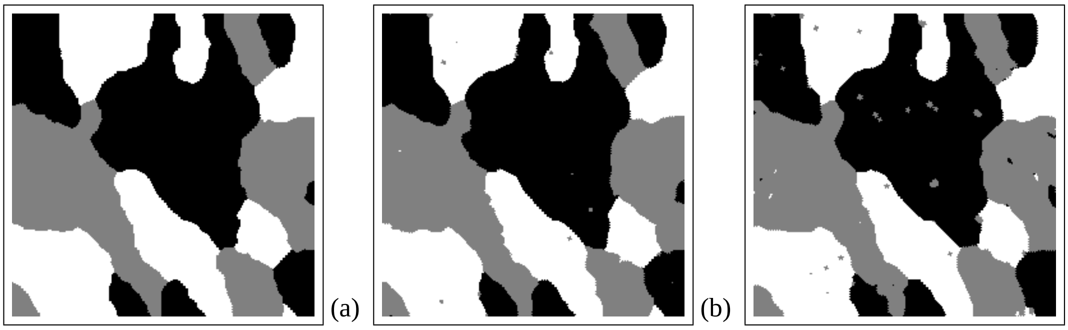
Figure 3: Three class segmentation of the images presented in figure 2 using (a)-(c) FMM, (d)-(f) SVFMM and (g)-(i) A-SVFMM. See text for techniques abbreviations.

amount of noise. The former methods need 10-30 iterations for the 1 dB noisy image while the latter needs only 4-8 iterations of the EM algorithm for the same amount of noise.

Finally, we have applied our directional, class-adaptive finite mixture model segmentation algorithm with the prior probability density (14) to the same data (fig. 4). The results clearly illustrate that the introduced directional adaptivity further improves the segmentation results both visually and in terms of correct pixel classification.

5. CONCLUSION

We have presented a hierarchical and spatially constrained mixture model for image segmentation. The model takes into account spatial information by imposing distinct smoothness priors on the probabilities of each cluster. Experimental re-



(a) (b) (c)

Figure 4: Three class segmentation of the images presented in figure 2 using the directional version of the A-SVFMM described by the prior density in eq. (14). The percentages of correctly classified pixels are 99.6%, 98.%5 and 95.8% respectively.

sults have shown that our approach improves significantly not only the standard mixture model-based segmentation but also its spatially variant version. Future work consists in applying the methods to real world segmentation problems in medical imaging and bioinformatics. Also, we plan to consider color and texture image segmentation.

REFERENCES

[1] C. M. Bishop. Neural networks for pattern recognition. *Oxford University Press Inc.*, New York, 1995.

[2] K. Blekas, A. Likas, N. Galatsanos and I. Lagaris. A spatially constrained mixture model for image segmentation. *IEEE Transactions on Neural Networks*, 16(2):494-498, 2005.

[3] J. Chantas, N. Galatsanos and A. Likas. Maximum a posteriori image restoration based on a new directional continuous edge image prior. *IEEE Proceedings of the International Conference on Image Processing (ICIP) 2005*, 11-14 Sept. 2005, Genova, Italy, (1):941 - 944.

[4] H. Derin and H. Elliot. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:39-55, 1987.

[5] A. P. Dempster, N. M. Laird and D. B. Bubin. Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society*, ser. B, 39(1):1-38, 1977.

[6] S. Sanjay-Gopal and T. Hebert. Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm. *IEEE Transactions on Image Processing*, 7(7):1014-1028, 1998.

[7] P. J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging*, 9(1):84-93, 1990.

[8] R. Haralick and L. G. Shapiro. Survey: image segmentation techniques. *Computer Vision, Graphics and Image Processing*, 29:100-132, 1985.

[9] J. Huber. Robust statistics John Wiley and sons, New York, 1981.

[10] J. Kittler and J. Illingworth. Minimum error threholding. *Pattern Recognition*,19(1):41-47, 1986.

[11] G. McLachlan. Finite mixture models. Wiley-Interscience, 2000.

[12] R. Molina, and B. D. Ripley. Using spatial models as priors in astronomical images analysis. *Journal of Applied Statistics*, (16):193-206, 1989.

[13] D. Ormoneit and V. Tresp. Averaging, maximum penalized likelihood and bayesian estimation for improving gaussian mixture probability density estimates. *IEEE Transactions on Image Processing*, 9(4):639-650, 1998.

[14] N. Otsu. A threshold selection method from grey-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62-66, 1979.

[15] N. Pal and S. Pal. A review of image segmentation techniques. *Pattern Recognition*, 26:1277-1294, 1993.

[16] D. Titterington, A. Smith and U. Makov. Statistical analysis of finite mixture distributions. John Wiley and sons, 1985.

[17] Y. Zhang, M. Brady and S. Smith. 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.