A Bayesian Framework for Image Segmentation With Spatially Varying Mixtures

Christophoros Nikou, *Member, IEEE*, Aristidis C. Likas, *Senior Member, IEEE*, and
Nikolaos P. Galatsanos, *Senior Member, IEEE*

Abstract—A new Bayesian model is proposed for image segmentation based upon Gaussian mixture models (GMM) with spatial smoothness constraints. This model exploits the Dirichlet compound multinomial (DCM) probability density to model the mixing proportions (i.e., the probabilities of class labels) and a Gauss–Markov random field (MRF) on the Dirichlet parameters to impose smoothness. The main advantages of this model are two. First, it explicitly models the mixing proportions as probability vectors and simultaneously imposes spatial smoothness. Second, it results in closed form parameter updates using a maximum *a posteriori* (MAP) expectation-maximization (EM) algorithm. Previous efforts on this problem used models that did not model the mixing proportions explicitly as probability vectors or could not be solved exactly requiring either time consuming Markov Chain Monte Carlo (MCMC) or inexact variational approximation methods. Numerical experiments are presented that demonstrate the superiority of the proposed model for image segmentation compared to other GMM-based approaches. The model is also successfully compared to state of the art image segmentation methods in clustering both natural images and images degraded by noise.

Index Terms—Bayesian model, Dirichlet compound multinomial distribution, Gauss–Markov random field prior, Gaussian mixture, image segmentation, spatially varying finite mixture model.

I. INTRODUCTION

MANY approaches have been proposed to solve the image segmentation problem [1], [2]. Among them, clustering based methods rely on arranging data into groups having common characteristics [3], [4]. During the last decade, the main research directions in the relevant literature are focused on graph theoretic approaches [5]–[8], methods based upon the mean shift algorithm [9], [10] and rate distortion theory techniques [11], [12].

Modeling the probability density function (pdf) of pixel attributes (e.g., intensity, texture) with finite mixture models (FMM) [13]–[15] is a natural way to cluster data because it automatically provides a grouping based upon the components of the mixture that generated them. Furthermore, the likelihood of a FMM is a rigorous metric for clustering performance [14]. FMM based pdf modeling has been used successfully in a number of applications ranging from bioinformatics [16] to image retrieval [17]. The parameters of the FMM model with Gaussian components can be estimated through maximum likelihood (ML) estimation using the Expectation-Maximization (EM) algorithm [13], [14], [18]. However, it is well-known that the EM algorithm finds, in general, a local maximum of the likelihood. Furthermore, it can be shown that Gaussian components allow efficient representation of a large variety of pdf. Thus, Gaussian mixture models (GMM), are commonly employed in image segmentation tasks [14].

A drawback of the standard ML approach for image segmentation is that commonality of location is not taken into account when grouping the data. In other words, the prior knowledge that adjacent pixels most likely belong to the same cluster is not used. To overcome this shortcoming, spatial smoothness constraints have been imposed.

Imposing spatial smoothness is key to certain image processing applications since it is an important *a priori* known property of images [19]. Examples of such applications include denoising, restoration, inpainting and segmentation problems. In a probabilistic framework, smoothness is expressed through a prior imposed on image features. A common approach is the use of an MRF. Many MRF variants have been proposed, see for example [20]. However, determination of the amount of the imposed smoothness automatically requires knowledge of the normalization constant of the MRF. Since this is not known analytically, learning strategies were proposed [21]–[23].

Research efforts in imposing spatial smoothness for image segmentation can be grouped into two categories. In the methods of the first category, spatial smoothness is imposed on the discrete *hidden* variables of the FMM that represent class labels, see for example [7], [24]–[26]. These approaches may be categorized in a more general area involving simultaneous image recovery and segmentation which is better known as *image modeling* [27]–[30]. More specifically, spatial regularization is achieved by imposing a discrete Markov random field (DMRF) on the classification labels of neighboring pixels that penalizes solutions where neighboring pixels belong to different classes. Another method in this category is proposed in [7] which is based upon the optimization of an energy function having a term for the quality of the clustering and a term for the spatial tightness. Minimization of the energy function is accomplished using graph cuts [31].

The Gaussian scale mixtures (GSM) and their extension of mixtures of projected GSM (MPGSM) and the fields of GSM (FoGSM) were also used in image denoising in the wavelet domain in [32]–[34]. In GSM denoising [32], clusters of wavelet coefficients are modeled as the product of a Gaussian random vector and a positive scaling variable. In MPGS denoising [34], the model is extended to handle different local image characteristics and incorporates dimensionality reduction through linear projections. By these means, the number of model parameters is reduced and fast model training is obtained. In the case of FoGSM [33], multiscale subbands are modeled by a product of an exponentiated homogeneous Gaussian Markov random field (hGMRF) and a second independent hGMRF. In [33], it is demonstrated that samples drawn from a FoGSM model have marginal and joint statistics similar to subband coefficients of photographic images.

To estimate the smoothness parameters, Woolrich *et al.* proposed in [35] and [36] a model based upon a logistic transform that approximates the previously mentioned DMRF with a *continuous* Gaussian Markov random field. However, for this model inference of the *contextual mixing proportions* (posterior class label probabilities) of each pixel cannot be obtained in closed form. Thus, in [35], inference based upon Markov Chain Monte Carlo (MCMC) is proposed, while in [36] inference based upon Variational Bayes (VB) is employed. Although MCMC methods have been studied in statistics for a long time and several general criteria have been proposed to determine their convergence [37], [38], inference based upon them may be notoriously time consuming. On the other hand, VB-based inference is approximate and there is no easy way to assert the tightness of the variational bound. Moreover, similar in spirit approaches to avoid local maxima of the likelihood, which is a drawback of the ML solution, rely on the stochastic EM and its variants [39], [40].

In the second category of methods, the MRF-based smoothness constraint is not imposed on the labels but on the contextual mixing proportions. This model is called spatially variant finite mixture model (SVFMM) [41] and avoids the inference problems of DMRFs. In this model maximum *a posteriori* (MAP) estimation of the contextual mixing proportions via the MAP-EM algorithm is possible. However, the main disadvantage of this model is that the M-step of the proposed algorithm cannot be obtained in closed form and is formulated as a constrained optimization problem that requires a projection of the solution onto the unit simplex (positive and summing up to one components) [41], [42]. Consequently, the parameters that control the spatial smoothness cannot be estimated automatically from the data.

In [43], a new family of smoothness priors was proposed for the contextual mixing proportions based upon the Gauss–Markov random fields that takes into account cluster statistics, thus, enforcing different smoothness strength for each cluster. The model was also refined to capture information in different spatial directions. Moreover, all the parameters controlling the degree of smoothness for each cluster, as well as the label probabilities for the pixels, are estimated in closed form via the maximum *a posteriori* (MAP) methodology. The advantage of this family of models is that inference is obtained using an EM algorithm with closed form update equations. However, the implied model still does not take into account

explicitly that the mixing proportions are probabilities, thus, the constraint that they are positive and must sum to one is not guaranteed by the update equations. As a result, the M-step of this EM algorithm also requires a *reparatory* projection step which is ad-hoc and not an implicit part of the assumed Bayesian model. A synergy between this category of priors and line processes, to account for edge preservation, was presented in [44].

In this paper, we present a new hierarchical Bayesian model for mixture model-based image segmentation with spatial constraints. This model assumes the *contextual mixing proportions* to follow a Dirichlet compound multinomial (DCM) distribution. More precisely, the class to which a pixel belongs is modeled by a discrete multinomial distribution whose parameters follow a Dirichlet law [45]. Furthermore, spatial smoothness is imposed by assuming a Gauss–Markov random field (GMRF) prior for the parameters of the Dirichlet. The parameters of the multinomial distribution are integrated out in a fully Bayesian framework and the updates of the parameters of the Dirichlet are computed in closed form through the EM algorithm.

The Dirichlet distribution has been previously proposed as a prior for text categorization [46], [47], object recognition and detection [48] and scene classification [49]. The difference of the proposed model with respect to existing methods is twofold. At first, text, scene or object categorization are supervised learning problems while the proposed segmentation method is unsupervised. Also, in the existing studies, estimation of the parameters of the Dirichlet distribution is generally accomplished by variational inference or by simplified logistic models. The advantage of the herein proposed model is that, not only the E-step can be expressed in closed form, but also our model explicitly assumes that the contextual mixing proportions are probability vectors. Inference through the EM algorithm leads to a third degree polynomial equation for the parameters of the Dirichlet distribution. Therefore, the closed form M-step yields parameter values automatically satisfying the necessary probability constraints.

Another approach to handle non stationary images and relying on MRF is the triplet Markov field (TMF) model [50] which was also applied to image segmentation [51], [52]. The main difference of TMF with respect to our model is that, in TMF, the random field is imposed jointly on the hidden variables, the observation and a set of auxiliary variables which determine the type of the stationarity. In contrast, in our model, the random field is imposed on the *contextual mixing proportions*.

Numerical experiments are presented to assess the performance of the proposed model both with simulated data where the ground truth is known and real natural images where the performance is assessed both visually and quantitatively.

The remainder of the manuscript is organized as follows: background for the spatially variant finite mixture model is given in Section II. The proposed modeling of probabilities of the pixel labels with a DCM distribution is presented in Section III. In Section IV, the MAP-EM algorithm for the estimation of the proposed model parameters is developed. Experimental results of the application of our model to natural and artificial images are presented in Section V and conclusions and directions for future research are given in Section VI.

II. BACKGROUND ON SPATIALLY VARIANT FINITE MIXTURE MODELING

Let $x^i = (x_1^i, x_2^i, \ldots, x_L^i)^T$ denote the vector of features (e.g., intensity, textural features) representing the ith spatial location, $i = 1 \ldots, N$, of a d-dimensional vector valued image modeled as independently distributed random variables. The SVFMM [41]–[43] provides a modification of the classical FMM approach [13], [14] for pixel labeling. It assumes a mixture model with K components each one having a vector of parameters θ_j defining the density function.

Pixel i is characterized by its probability vector $\pi^i = \left(\pi_1^i \pi_2^i \ldots \pi_K^i\right)^T$ where K is the number of components. We define $\mathbf{\Pi} = \{(\pi^1)^T, (\pi^2)^T, \ldots (\pi^N)^T\}$ as the set of probability vectors and $\mathbf{\Theta} = \{\theta_1, \theta_2, \ldots, \theta_K\}$ the set of component parameters. The parameters π_j^i, $j = 1, \ldots, K$ are the *contextual mixing proportions* for each pixel i and represent the probabilities of the ith pixel ($i = 1, \ldots, N$) to belong to the jth class and must satisfy the constraints

$$0 \leq \pi_j^i \leq 1, \quad \sum_{j=1}^{K} \pi_j^i = 1$$
$$i = 1, \ldots, N, \quad j = 1, \ldots, K. \quad (1)$$

The standard finite mixture model [41] assumes that the probability density function of an observation x^i is expressed by

$$f(x^i | \mathbf{\Pi}; \mathbf{\Theta}) = \sum_{j=1}^{K} \pi_j^i \phi(x^i | \theta_j) \quad (2)$$

with $\phi(x^i | \theta_j)$ being a Gaussian distribution with parameters $\theta_j = \{\mu_j, \Sigma_j\}$, where $\mu_j = (\mu_{j,1}, \mu_{j,2}, \ldots \mu_{j,L})$ is the mean vector and Σ_j is the covariance matrix of the L-dimensional Gaussian distribution. This notation implies that $\mathbf{\Pi}$ are considered as random variables and $\mathbf{\Theta}$ as parameters.

The spatially varying finite mixture models use a prior density distribution $p(\mathbf{\Pi})$ for the random variables $\mathbf{\Pi}$. Therefore, denoting $\mathbf{X}$ the set of pixels feature vectors $\{x^i\}$, with $i = 1, \ldots, N$, which we assume to be statistically independent and following Bayes rules, we obtain the posterior probability density function given by

$$p(\mathbf{\Pi} | \mathbf{X}; \mathbf{\Theta}) \propto \prod_{i=1}^{N} p(\mathbf{\Pi}) f(x^i | \mathbf{\Pi}, \mathbf{\Theta}) \quad (3)$$

with the log-density

$$L(\mathbf{\Pi} | \mathbf{X}; \mathbf{\Theta}) = \sum_{i=1}^{N} \left\{ \log f(x^i | \mathbf{\Pi}, \mathbf{\Theta}) + \log (p(\mathbf{\Pi})) \right\}. \quad (4)$$

A typical example of $p(\mathbf{\Pi})$ is the Gauss–Markov random field prior [43], expressed by

$$p(\mathbf{\Pi}) \propto \prod_{j=1}^{K} \beta_j^{-N} \exp \left[-\frac{1}{2} \frac{\sum_{i=1}^{N} \sum_{m \in \mathcal{N}_i} (\pi_j^i - \pi_j^m)^2}{\beta_j^2} \right] \quad (5)$$

where the parameter β_j captures the spatial smoothness of cluster j and enforces different degree of smoothness in each cluster in order to better adapt the model to the data.

The graphical model for the spatially variant version of the finite mixture model (SVFMM) is presented in Fig. 1. In the standard FMM, the feature vector x for a given pixel depends upon the state of the discrete *hidden* variable z denoting the mixture component responsible for generating the observation x. That is, if $z_j = 1$ pixel x belongs to class j. In that case, the mixing proportion for a given class is simply the percentage of pixels belonging to that class. In the case of the SVFMM, each pixel i has its own set of mixing proportions π^i, generally called contextual mixing proportions or probabilities of the pixel labels. These contextual mixing proportions are spatially constrained by a smoothness prior. The strength of this prior could either be unique for the whole set of pixels [42] or could vary based upon the local statistics of each image class, thus, making the model less stationary (5).

The EM algorithm [18] for MAP estimation of the model parameters requires the computation of the conditional expectation values of the hidden variables at the E-step of iteration step t

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

In the M-step, considering that the complete data log-likelihood is linear in the "hidden" variables [18], the maximization of the complete data log-likelihood

$$Q(\mathbf{\Pi}, \mathbf{\Theta} | \mathbf{\Pi}^{(t)}, \mathbf{\Theta}^{(t)}) = \sum_{i=1}^{N} \sum_{j=1}^{K} z_j^{i(t)} \{ \log(\pi_j^i) + \log (\phi(x^i | \theta^j)) \} + \log (p(\mathbf{\Pi})) \quad (7)$$

yields the model parameters. The function $Q(\cdot)$ in (7) can be maximized independently for each parameter providing the following update equations of the mixture model parameters at step $t + 1$

$$\mu_{j,l}^{(t+1)} = \frac{\sum_{i=1}^{N} z_j^{i(t)} x_l^{i(t)}}{\sum_{i=1}^{N} z_j^{i(t)}}$$

$$\sigma_{j,l}^{(t+1)} = \sqrt{\frac{\sum_{i=1}^{N} z_j^{i(t)} \left[x_l^{i(t)} - \mu_{j,l}^{(t+1)} \right]^2}{\sum_{i=1}^{N} z_j^{i(t)}}}. \quad (8)$$

The probabilities π_i^j are computed by setting $\partial Q / \partial \pi_j^i = 0$ which yields a second degree equation with respect to π_j^i

$$|\mathcal{N}_i| \left(\pi_j^i\right)^2 - \left[\sum_{m \in \mathcal{N}_i} \pi_j^m \right] \left(\pi_j^i\right) - z_j^i \beta_j^2 = 0 \quad (9)$$

where $|\mathcal{N}_i|$ is the number of pixels in the neighborhood of the ith pixel.
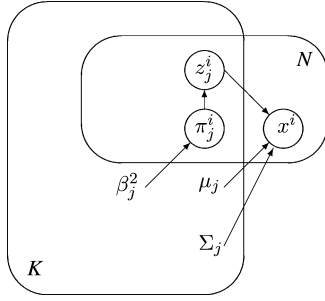
Fig. 1. Graphical model for the spatially variant finite mixture model (SVFMM).

It is easily verified that (9) has always a real nonnegative solution for π_i^j. However, the main drawback of the SVFMM is that it imposes spatial smoothness on π^i without explicitly taking into account that it is a probability vector ($\pi_j^i \geq 0$, $\sum_{j=1}^{K} \pi_j^i = 1$, $i = 1, 2, \ldots, N$). For this purpose, *reparatory* computations were introduced in the M-step to enforce the variables π_i^j to satisfy these constraints. A gradient projection algorithm was used in [41] and quadratic programming was proposed in [42]. This approach was shown to improve both the criterion function (7) and the performance of the model. However, *reparatory* projections compromise the assumed Bayesian model.

III. DIRICHLET COMPOUND MULTINOMIAL MODELING OF CONTEXTUAL MIXING PROPORTIONS

To overcome the limitations of SVFMM, we propose in this section, a new Bayesian model for mixture-based the image segmentation problem based upon a hierarchical prior for the the contextual mixing proportions $\mathbf{\Pi}$, which are assumed to follow a DCM distribution. The DCM distribution is a multinomial whose parameters are generated by a Dirichlet distribution [45], thus π^i are probability vectors. Similar in spirit priors have been proposed, in the totally different context of text modeling [47] where the DCM parameters are estimated through an iterative gradient descent optimization method. Also in a recent work [53], a new family of exponential distributions is proposed capable of approximating the DCM probability law in order to make its parameter estimation faster than [47]. In what follows, we describe how to compute them in closed form. Furthermore, spatial smoothness is imposed on the parameters of the Dirichlet distributions which are computed in closed form through a cubic equation having always one real non negative solution that satisfies the constraints of the Dirichlet parameters.

A. Dirichlet Compound Multinomial Distribution

More precisely, for the ith pixel, $i = 1, 2, \ldots, N$, the class label z^i is considered to be a random variable following a multinomial distribution with probability vector $\xi^i = (\xi_1^i, \xi_2^i, \ldots, \xi_K^i)^T$ with K being the number of classes.

Let also $\Xi = \{(\xi^1)^T, (\xi^2)^T, \ldots (\xi^N)^T\}$ to be the set of parameters for the whole image. By the multinomial definition it holds that

$$p(z^i | \xi^i) = \frac{M!}{\prod\limits_{j=1}^{K} (z_j^i)!} \prod_{j=1}^{K} \left(\xi_j^i\right)^{z_j^i}, \quad i = 1, \ldots, N \quad (10)$$

with

$$\xi_j^i \geq 0, \quad \sum_{j=1}^{K} \xi_j^i = 1, \quad i = 1, 2, \ldots, N. \quad (11)$$

The model described by (10) represents the probability ξ_j^i that pixel i belongs to class j, as one of the K possible outcomes of a multinomial process with M realizations. Each of the K outcomes of the process appears with probability ξ_j^i, $j = 1, 2 \ldots, K$. Generally speaking, this is a generative model for the image. When the multinomial distribution is used to generate a clustered image, the distribution of the number of emissions (i.e., counts) of an individual class follows a binomial law.

The DCM distribution assumes that parameters Ξ of the multinomial follow a Dirichlet distribution parameterized by $\mathbf{A} = \{(\alpha^1)^T, (\alpha^2)^T, \ldots (\alpha^N)^T\}$ where $\alpha^i = (\alpha_1^i, \alpha_2^i, \ldots, \alpha_K^i)^T$, $i = 1, 2 \ldots, N$, is the vector of Dirichlet parameters for ξ^i

$$p(\xi^i | \alpha^i) = \frac{\Gamma\left(\sum\limits_{j=1}^{K} \alpha_j^i\right)}{\prod\limits_{j=1}^{K} \Gamma(\alpha_j^i)} \prod_{j=1}^{K} \left(\xi_j^i\right)^{(\alpha_j^i - 1)} \quad (12)$$

where $\alpha_j^i \geq 0$, $i = 1, 2, \ldots, N$, $j = 1, 2 \ldots, K$ and $\Gamma(\cdot)$ is the Gamma function.

Under the Bayesian framework, the probability label for the ith image pixel is obtained by marginalizing the parameters Ξ

$$p(z^i | \alpha^i) = \int_0^1 p(z^i | \xi^i) p(\xi^i | \alpha^i) d\xi^i, \quad i = 1, 2, \ldots, N. \quad (13)$$

Substituting (10) and (12) into (13) we obtain , with some easy manipulation, the following expression for the label probabilities:

$$p(z^i | \alpha^i) = \frac{M!}{\prod\limits_{j=1}^{K} (z_j^i)!} \frac{\Gamma\left(\sum\limits_{j=1}^{K} \alpha_j^i\right)}{\Gamma\left(\sum\limits_{j=1}^{K} \alpha_j^i + z_j^i\right)} \prod_{j=1}^{K} \frac{\Gamma(\alpha_j^i + z_j^i)}{\Gamma\left(\alpha_j^i\right)} \quad (14)$$

with $i = 1, 2, \ldots, N$.

B. Hierarchical Image Model

We now assume a generative model for the image where the determination of component j generating the ith pixel is an outcome of a DCM process with only one realization. Consequently, the vector of the hidden variables z^i (6) has the jth

component equal to one and all the others set to zero. This is also illustrated in the contextual mixing proportions which in that case are the posterior probabilities

$$\pi_j^i = p(z_j^i = 1|x^i) = 1$$
$$\pi_m^i = p(z_m^i = 1|x^i) = 0 \quad m = 1, \ldots, K, \, m \neq j. \quad (15)$$

Thus, taking into account that we have one realization of the DCM process $(M = 1)$ and that $\Gamma(x + 1) = x\Gamma(x)$, the label probabilities for the ith pixel in (14) become

$$\pi_j^i = p(z_j^i = 1|x^i) = \frac{\alpha_j^i}{\sum_{l=1}^{K} \alpha_l^i}, \quad j = 1, 2 \ldots, K. \quad (16)$$

The new model may become spatially variant by introducing a spatial prior on the parameters $\mathbf{A}$ of the Dirichlet distribution. More specifically, we assume a Gauss–Markov type random field prior for our model since its parameters may be estimated in closed form [43]

$$p(\mathbf{A}) \propto \prod_{j=1}^{K} \beta_j^{-N} \exp\left[-\frac{1}{2}\frac{\sum_{i=1}^{N}\sum_{m\in\mathcal{N}_i}(\alpha_j^i - \alpha_j^m)^2}{\beta_j^2}\right]. \quad (17)$$

The main characteristic of this prior is that it enforces smoothness of different degree in each cluster, thus providing better adaptation to the data [43].

The graphical model for this hierarchical approach is presented in Fig. 2. We will refer to this new model as the Dirichlet Compound Multinomial-based Spatially Variant Finite Mixture Model (DCM-SVFMM). The generative image model works as follows: a sample ξ (probability vector) is first drawn from a Dirichlet distribution with parameter α, thus obtaining a multinomial distribution with parameter ξ. The "hidden" variable z, denoting the class of observation x, is the outcome of a multinomial process parameterized by ξ. Moreover, the parameters α of the Dirichlet distribution are spatially constrained by a smoothness prior as it is also the case for the standard SVFMM.

IV. MAP-EM Estimation

Employing the proposed model one can derive the corresponding MAP algorithm using the EM methodology. Applying the Gauss–Markov prior in (17) to parameters $\mathbf{A}$ yields the following MAP function to be maximized in the M-step of the
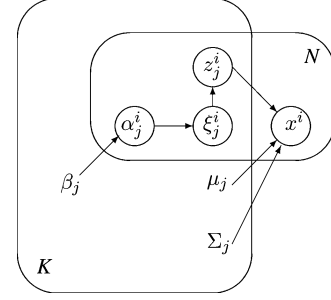

Fig. 2. Graphical model for feature vector x following a spatially variant finite mixture model (SVFMM) with a Dirichlet compound multinomial (DCM) prior.

EM algorithm, see (18) at the bottom of the page, where we define

$$A^i = \sum_{j=1}^{K} \alpha_j^i.$$

To compute the model parameters $\alpha_j^{i(t+1)}$ at the M-step we have to maximize (18) with respect to α_j^i, that is, to compute its partial derivative and set the result to zero. Considering a neighborhood with $|\mathcal{N}_i| = 8$ pixels and setting $\partial Q/\partial \alpha_j^i = 0$ gives a third degree polynomial equation with respect to $\alpha_j^{i(t+1)}$, for $i = 1, \ldots, N$ and $j = 1, \ldots, K$

$$(\alpha_j^i)^3 + \left[A_{-j}^i - \frac{\sum_{m\in\mathcal{N}_i}\alpha_j^m}{|\mathcal{N}_i|}\right](\alpha_j^i)^2$$
$$+ \frac{A_{-j}^i \sum_{m\in\mathcal{N}_i}\alpha_j^m}{|\mathcal{N}_i|}(\alpha_j^i) - \frac{z_j^i A_{-j}^i \beta_j^2}{2|\mathcal{N}_i|} = 0 \quad (19)$$

where

$$A_{-j}^i = \sum_{\substack{m=1 \\ m\neq j}}^{K} \alpha_m^i, \quad i = 1, 2, \ldots, N.$$

Based upon polynomial theory, it can be proved that (19) has only one real nonnegative solution satisfying the constraint in (12), that is $\alpha_j^i \geq 0$. Specifically, the constant term of the third degree (19) is negative, thus the product of the roots should be positive. This implies that, for three real roots, either three roots should be positive or two roots should be negative and one positive. The coefficient of the quadratic term is positive which implies that the sum of the roots should be negative. Therefore, two roots should be negative and one positive. For a pair of complex conjugate roots and a real root, the real root is always positive

$$Q(\mathbf{A}, \mathbf{\Theta}|\mathbf{A}^{(t)}, \mathbf{\Theta}^{(t)}) = \sum_{i=1}^{N}\sum_{j=1}^{K}\left\{z_j^i\left[\log\left(\frac{\alpha_j^i}{A^i}\right) + \log\left(\phi(x^i|\theta^j)\right)\right] - \log(\beta_j^N) - \frac{1}{2}\frac{\sum_{m\in\mathcal{N}_i}(\alpha_j^i - \alpha_j^m)^2}{\beta_j^2}\right\} \quad (18)$$

since the product of the three roots should be positive. Moreover, the root can be obtained in closed form, see [54], [55].

Finally, the solution for the class variances are obtained by setting $\partial Q / \partial \beta_j^2 = 0$ and solving for β_j^2 at time step $(t+1)$

$$\beta_j^2 = \frac{1}{N} \sum_{i=1}^{N} \sum_{m \in \mathcal{N}_i} \left(\alpha_j^i - \alpha_j^m \right)^2, \quad j = 1, \ldots, K. \quad (20)$$

The overall MAP-EM algorithm is summarized as follows:
- Initialize the mixture model component parameters $\theta_j = \{\mu_j, \Sigma_j\}, j = 1, \ldots, K$ and the vectors of parameters α^i of the Dirichlet distribution for each pixel $i = 1, \ldots, N$.
- Do until the MAP functional (18) does not change significantly.
  — E-step
    - Calculate the posterior probabilities $z_j^{i(t)}$ that the ith pixel belongs to the jth class (6).
  — M-step
    - Calculate the new mixture model parameters (8).
    - Update the parameters of the Dirichlet distribution by keeping only the real nonnegative solution of (19).
    - Update the probabilities for the pixel labels (16).
    - Update the class variances (20).
- End

The principal advantage of the DCM approach is that the obtained solution satisfies the constraints in (1) for the contextual mixing proportions as shown in (16) which is a natural consequence of the DCM prior. Therefore, no *reparatory* projection step is needed in the EM algorithm in contrast to previous methods [41], [42].

V. EXPERIMENTAL RESULTS

Since the EM algorithm is sensitive to initialization, in our experiments, 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. In the termination criterion of the EM algorithm, considered here, convergence was defined as the percentage of change in the log-likelihood (4) between two consecutive iterations to be less than 0.001%, or $|\Delta L/L| < 10^{-5}$.

At first, to justify the necessity for the proposed DCM prior with respect to the previously proposed SVFMM [43], we present examples of segmentation of piecewise constant images of varying complexity. Moreover, evaluation of our method to the segmentation of natural images and comparison to the state of the art methods was also performed. We applied the DCM-SVFMM algorithm to the 300 images of the Berkeley image data base [56]. In that context, we compared our method with two unsupervised segmentation algorithms: the Gaussian blurring mean-shift (GBMS) [10] and the normalized cut (Ncut) [5]. In both cases (artificial and natural images), the robustness of our method to Gaussian noise was also investigated.

A. Piecewise Constant Image Segmentation

We begin with a simple example of two-class image segmentation in order to clarify why a DCM-SVFMM model naturally
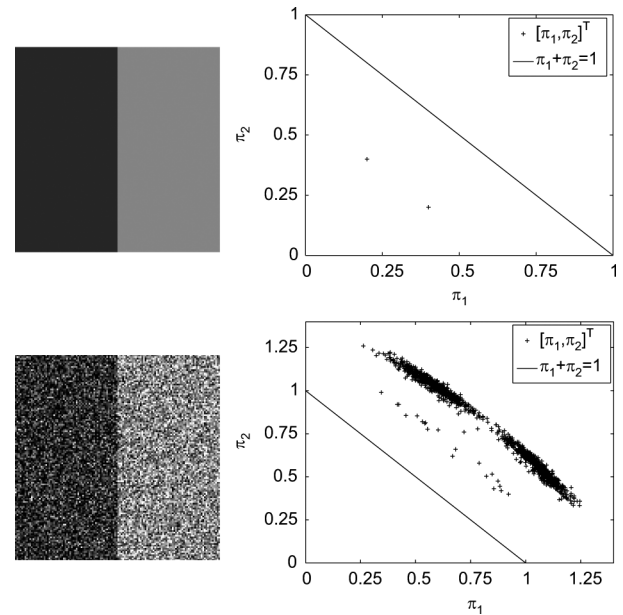

Fig. 3. Two-class image segmentation and the contextual mixing proportions (π_1^i, π_2^i), for each pixel $i = 1, \ldots, N$ for the first iteration of the EM algorithm of the SVFMM. Top row: a piecewise constant image of two distinct regions of intensities 50 and 150 with equal mixing proportions. Notice that all of the values of the mixing proportions coincide at a single point. Bottom row: The image is degraded by zero mean white Gaussian noise in order to achieve $\text{SNR} = 7$ dB. A representative percentage of the points are plotted for visualization purposes. In both cases, the contextual mixing proportions need to be projected onto the line $\pi_1 + \pi_2 = 1$.

completes the Bayesian framework of the standard SVFMM. The piecewise constant image at the top left of Fig. 3 has two distinct regions of intensities 50 and 150. Segmenting the image into two classes by the SVFMM [43] leads to the estimation of the posterior probabilities $\pi_j^i, j = 1, 2$ for each pixel $i = 1, \ldots, N$. As the classes are clearly distinct, the values of the probabilities of the pixel labels coincide with the two crosses at the bottom left corner of the figure (each cross represents half of the image pixels). The values of the contextual mixing proportions must satisfy (1), that is, they must lie on the straight line $\pi_1 + \pi_2 = 1$. However, as the clustering of the image is easy to obtain the projection step does not alter the segmentation.

When the image is degraded by white Gaussian noise at 7 dB the values of π_j^i have wider diversity. Projecting these values onto the line $\pi_1 + \pi_2 = 1$ through a *reparatory* step leads to a percentage of correctly classified pixels which is consistently 0.4–0.8% inferior to the result obtained by the DCM-SVFMM where no projection needs to be performed as the contextual mixing proportions lie on the line $\pi_1 + \pi_2 = 1$. This difference may not be visible by the human eye but it is more pronounced when the number of segments increases.

For the 3-class segmentation of the piecewise constant image in Fig. 4 with the SVFMM, the contextual mixing proportions are projected onto the surface $\pi_1 + \pi_2 + \pi_3 = 1$ which leads to a more intricate configuration. This is not the case for the DCM-SVFMM where the contextual mixing proportions are naturally computed as probability vectors. In that case, the difference between the segmentations provided by the two models favors the DCM-SVFMM at about 1.5–2.5%. Notice that the
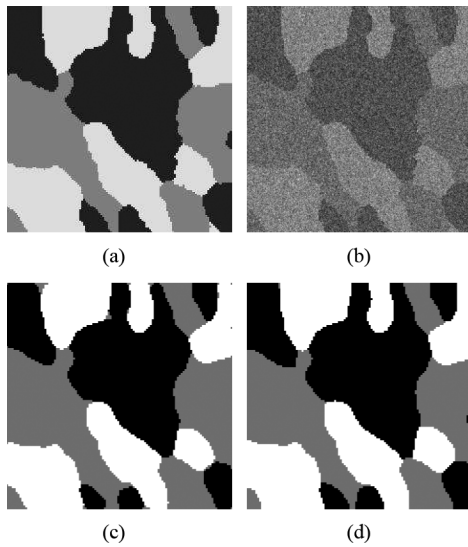
Fig. 4. (a) Three-class piecewise constant image. Intensity means are 30, 125, and 220. The contextual mixing proportions are 0.37, 0.30, and 0.33 respectively. (b) Image degraded by zero mean white Gaussian noise in order to achieve $\mathrm{SNR} = 7$ dB. Segmentation into three classes using the (c) SVFMM (95.1% correct classification) and the (d) DCM-SVFMM (97.3% correct classification).
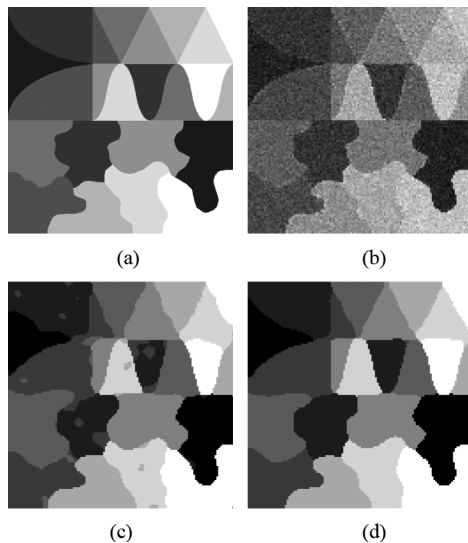

Fig. 5. (a) 8-class piecewise constant image. Intensity means are 32, 64, 96, 128, 160, 192, 224 and 255. The mixing proportions are respectively 0.12, 0.16, 0.15, 0.13, 0.11, 0.10, 0.12 and 0.11. (b) Image is degraded by zero mean white Gaussian noise in order to achieve $\mathrm{SNR} = 16$ dB. Segmentation into 8 classes using the (c) SVFMM (88.2% correct classification) and the (d) DCM-SVFMM (93.4% correct classification).

SVFMM classifies erroneously the pixels at the region boundaries, while the DCM-SVFMM is able to preserve the edges [Fig. 4(c) and (d)]. The SVFMM completely misses the little region at the middle right part of the image and creates two new small regions at the top left part of the image (outside the white region).

As the number of clusters increases, the performance deterioration due to the reparatory projection step is more apparent. In the 8-class segmentation example in Fig. 5, the segmentation provided by the SVFMM, at 16 dB signal to noise ratio, is
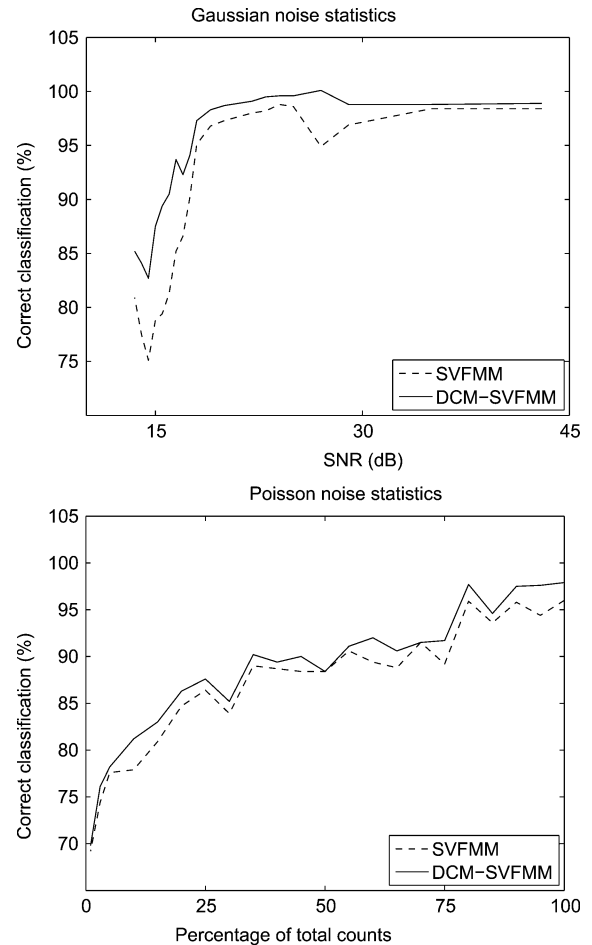

Fig. 6. Variation of the percentage of correctly classified pixels when the 8-class image in Fig. 5(a) is degraded by noise. Top: correct classification versus signal to noise ratio when the image is degraded by white Gaussian noise. Bottom: correct classification versus percentage of total counts when the image is degraded by Poisson noise.

erroneous not only at the region boundaries but also the piecewise constant image areas. At the same noise level, the DCM-SVFMM only slightly deforms the region boundaries (notice the triangular regions at the top right part of the segmented images).

The variation of the percentage of correctly classified pixels in Fig. 5(a) to different levels of Gaussian and Poisson noise is presented in Fig. 6. In the case of Gaussian noise, its variance was adjusted in order to obtain different SNR between 13 and 43 dB. In the case of photon limited images, every pixel intensity was replaced by a random number generated by a Poisson process having as mean value a fraction of the pixel intensity. Thus, the total photon counts of the image were reduced by different fractions. Poisson noise is signal dependent and difficult to handle. However, the DCM-SVFMM segments the image with high accuracy and in all cases provides better results than the SVFMM.

B. Natural Image Segmentation

We have also evaluated the DCM-SVFMM on the segmentation of the 300 images of the Berkeley image data base [56]. To this end, we have compared our method with the Ncut [5] and
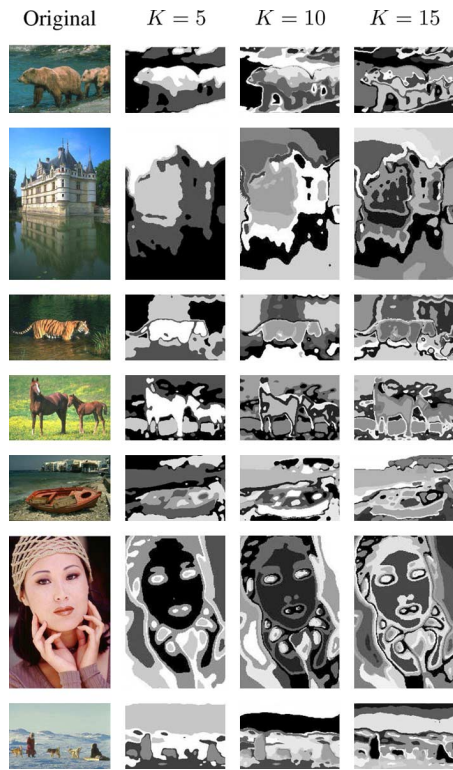
Fig. 7. Natural image segmentation examples using the DCM-SVFMM with the the MRF texture features described in [58]. The features were computed on 7×7 pixel neighborhoods and underwent dimensionality reduction by projection on a 8-dimensional space using PCA.

Fig. 8. Natural image segmentation examples using the DCM-SVFMM with the the Blobworld texture features described in [57].

the GBMS algorithm [10]. We have applied the DCM-SVFMM and the Ncut with number of segments $K = \{5, 10, 15, 20\}$ and the GBMS with the entropy based termination criterion [10].

An important issue in natural image segmentation is what features to use as input to a segmentation algorithm, especially when texture segmentation is considered. Apart from standard filter banks, in the recent literature, the two major approaches to texture description are the Blobworld features [57] and the MRF features [58]. Blobworld features generate 6-D vectors containing color (in *Lab* space) and texture (polarity, anisotropy, contrast) information and one of the key points of the process is the correct estimation of the scale of the texture. On the other hand, an MRF feature is simply a vectorized version of a cutoff window around a pixel followed by dimensionality reduction using PCA.

We have experimented with both approaches for texture description. Representative results are presented in Fig. 7 for the MRF texture model and in Fig. 8 for the Blobworld representation. As it can be observed, the description based upon the Blobworld features preserves better the boundaries between textured regions. MRF texture features provide a distinct response at textured boundaries as it is also noticed in [12]. However, the PCA smoothing step makes MRF texture features more insensitive to noise than Blobworld features. Further investigation revealed that Blobworld features from degraded natural images provided very favorable results for our DCM-SVFMM method with respect to ncut and GBMS. As one of our purposes is to investigate
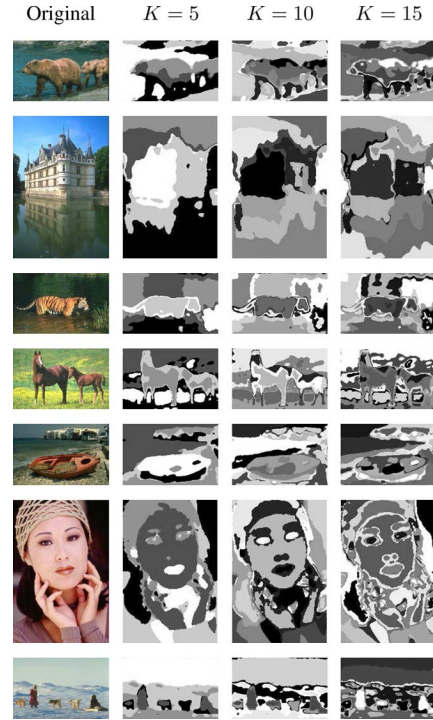
TABLE I
STATISTICS ON THE PROBABILISTIC RAND (PR) INDEX FOR THE COMPARED SEGMENTATION METHODS OVER THE 300 IMAGES OF THE BERKELEY IMAGE DATA BASE. ALL OF THE COMPARED METHODS EMPLOYED THE MRF TEXTURE FEATURES DESCRIBED IN [58]. THE FEATURES WERE COMPUTED ON 7×7 PIXEL NEIGHBORHOODS AND UNDERWENT DIMENSIONALITY REDUCTION BY PROJECTION ON A 8-D SPACE USING PCA

PR index - MRF texture features			
Method	mean	median	mean-10%
DCM ($K = 5$)	0.6923	0.7135	0.6991
DCM ($K = 10$)	0.7077	0.7436	0.7205
DCM ($K = 15$)	0.7062	0.7432	0.7204
DCM ($K = 20$)	0.7042	0.7417	0.7187
Ncut ($K = 5$)	0.6715	0.6993	0.6800
Ncut ($K = 10$)	0.7044	0.7420	0,7173
Ncut ($K = 15$)	0.7157	0.7594	0.7304
Ncut ($K = 20$)	0.7219	**0.7625**	0.7372
GBMS	**0.7362**	0.7516	**0.7492**

the behavior of the compared methods to noise without any bias, we have decided to perform the comparative experiments using the MRF texture features for all the methods.

Following the comparative road map in [12], we have quantitatively compared the segmentation methods using four performance measures: the probabilistic Rand index (PR) [59], the variation of information (VI) [60], the global consistency error (GCE) [56] and the Boundary displacement error (BDE) [61]. The PR index measures the consistency between human segmentations and the computed segmentation map. VI measures the amount of information one segmentation conveys about the

TABLE II
STATISTICS ON THE VARIATION OF INFORMATION (VI) FOR THE COMPARED
SEGMENTATION METHODS OVER THE 300 IMAGES OF THE BERKELEY IMAGE
DATA BASE. ALL OF THE COMPARED METHODS EMPLOYED THE MRF TEXTURE
FEATURES DESCRIBED IN [58]. THE FEATURES WERE COMPUTED ON 7×7
PIXEL NEIGHBORHOODS AND UNDERWENT DIMENSIONALITY REDUCTION BY
PROJECTION ON A 8-D SPACE USING PCA

VI - MRF texture features			
Method	mean	median	mean-10%
DCM ($K = 5$)	**2.9502**	**2.9533**	**2.9437**
DCM ($K = 10$)	3.5376	3.4617	3.5242
DCM ($K = 15$)	3.8678	3.8016	3.8534
DCM ($K = 20$)	4.1062	4.0605	4.0914
Ncut ($K = 5$)	3.4753	3.5104	3.4867
Ncut ($K = 10$)	3.8619	3.8891	3.8831
Ncut ($K = 15$)	4.0481	4.1570	4.0692
Ncut ($K = 20$)	4.1439	4.2889	4.1608
GBMS	3.5701	3.5521	3.5555

TABLE III
STATISTICS ON THE GLOBAL CONSISTENCY ERROR (GCE) FOR THE COMPARED
SEGMENTATION METHODS OVER THE 300 IMAGES OF THE BERKELEY IMAGE
DATA BASE. ALL OF THE COMPARED METHODS EMPLOYED THE MRF TEXTURE
FEATURES DESCRIBED IN [58]. THE FEATURES WERE COMPUTED ON 7×7
PIXEL NEIGHBORHOODS AND UNDERWENT DIMENSIONALITY REDUCTION BY
PROJECTION ON A 8-D SPACE USING PCA

GCE - MRF texture features			
Method	mean	median	mean-10%
DCM ($K = 5$)	**0.4077**	0.4263	0.4116
DCM ($K = 10$)	0.4339	0.4601	0.4381
DCM ($K = 15$)	0.4195	0.4434	0.4221
DCM ($K = 20$)	0.4084	**0.4216**	**0.4095**
Ncut ($K = 5$)	0.4945	0.5207	0.5027
Ncut ($K = 10$)	0.4842	0.5102	0.4910
Ncut ($K = 15$)	0.4487	0.4741	0.4534
Ncut ($K = 20$)	0.4127	0.4286	0.4150
GBMS	0.4405	0.4620	0.4445

TABLE IV
STATISTICS ON THE BOUNDARY DETECTION ERROR (BDE) FOR THE COMPARED
SEGMENTATION METHODS OVER THE 300 IMAGES OF THE BERKELEY IMAGE
DATA BASE. ALL OF THE COMPARED METHODS EMPLOYED THE MRF TEXTURE
FEATURES DESCRIBED IN [58]. THE FEATURES WERE COMPUTED ON 7×7
PIXEL NEIGHBORHOODS AND UNDERWENT DIMENSIONALITY REDUCTION BY
PROJECTION ON A 8-D SPACE USING PCA

BDE - MRF texture features			
Method	mean	median	mean-10%
DCM ($K = 5$)	17.7190	15.5480	16.9566
DCM ($K = 10$)	15.4710	13.4410	14.7653
DCM ($K = 15$)	15.0200	13.1390	14.3170
DCM ($K = 20$)	**14.8480**	**12.8180**	**14.1545**
Ncut ($K = 5$)	18.0130	15.6360	17.0902
Ncut ($K = 10$)	16.9400	14.7130	16.1475
Ncut ($K = 15$)	16.7920	14.5740	16.0207
Ncut ($K = 20$)	16.7820	14.4350	15.9999
GBMS	15.8600	14.0340	15.2440

other (mutual information). GCE measures the degree of refinement between two segmentations. Finally, BDE measures the average chamfer distance between the boundaries of two segmentation maps. For each measure the mean and the median were computed. As a supplemental robustness indicator, we also present the 10% trimmed mean values where the 5% best and the 5% worst results were excluded.

The algorithms were provided with the MRF features without any preprocessing step or an initialization that could significantly give important advantage to any method. However, better numerical results could be obtained by an initialization based upon superpixels [62], [63] as it is the case in the extensive comparative study presented in [12].

The results for the PR index are summarized in Table I where it is clear that the GBMS provides the better performance in terms of the mean and the 10%-trimmed mean values and the Ncut with $K = 20$ clusters reveals more robust (see its median value). In Table II, where the statistics for the VI are shown, the DCM-SVFMM with $K = 5$ performs better in all cases. As we have observed, the relatively small number of segmnets $(K = 5)$ maximizes the mutual information with respect to the cases where the value of K is larger. This occurs because, generally, the joint histograms present less disorder due to the smaller number of bins. In any case, the Ncut algorithm has inferior results with regard to DCM-SVFMM while the GBMS asymptotically approaches, but never reaches, the values of our method for $K = 10$. Similar remarks could be made for Table III, presenting the GCE, where our method clearly outperforms the Ncut and GBMS segmentation algorithms. Finally, the BDE statistics (Table IV) are also favorable to the DCM-SVFMM segmentation algorithm.

It must also be noted that the DCM-SVFMM method is parameter-free while both the Ncut and the GBMS methods are parameter dependent. This means that in order to apply these methods one must take into account that they should undergo a trial-and-error procedure in order to determine these parameters. Also, the involved parameters depend both upon the image

size and the expected spatial coherence. From this point of view, our method is more flexible.

Furthermore, we have examined the performance of the compared methods to the segmentation of noisy images. To this end, 20 images of the Berkeley image data base [56] were selected. These images contain both color and texture and represent a variety of natural scenes. The images were then degraded by white Gaussian noise of varying strength with the resulting SNR being between 12 dB and -4 dB. A SNR below 0 dB results in a severe degradation of the image. However, in MRF feature space, the corruption is less intense as PCA smoothing eliminates part of the noise. This is why we have preferred to use the MRF texture features instead of the Blobworld description which would be more favorable to the DCM-SVFMM algorithm.

The results for the mean value of the PR index as a function of the SNR are presented in Fig. 9. The curves for the Ncut and DCM-SVFMM for different values of the number of segments
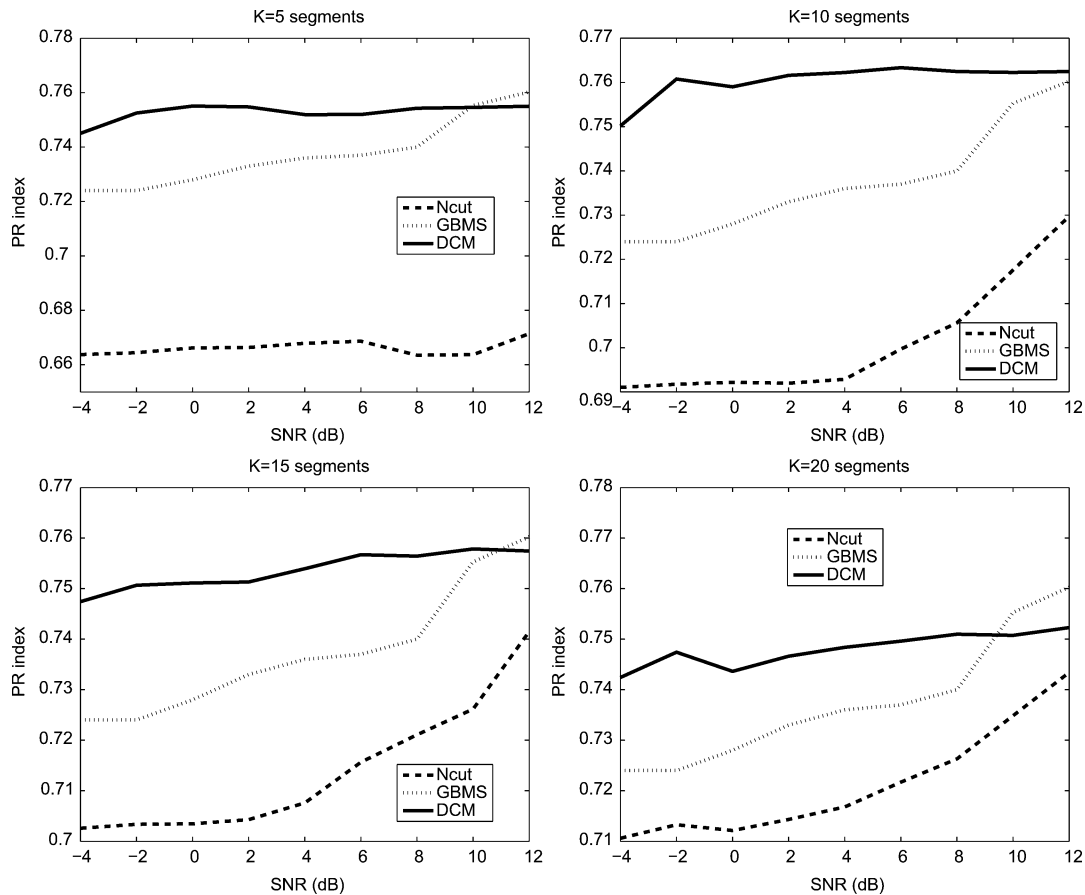
Fig. 9. Evaluation of the segmentation of 20 images of the Berkeley image data base degraded by white Gaussian noise of varying strength. The mean value of the PR index as a function of the SNR is presented for the compared methods. All of the compared methods employed the MRF texture features described in [58]. The features were computed on 7×7 pixel neighborhoods and underwent dimensionality reduction by projection on a 8-D space using PCA.

are shown. The curve for the GBMS is the same in all of the plots. As it can be observed, for higher SNR the GBMS and the DCM-SVFMM perform in a similar way. When the SNR values decrease, the DCM-SVFMM has better PR indices. The Ncut is clearly more influenced by the noise. Let us notice that the results presented in Fig. 9 are the more favorable to Ncut and especially to GBMS with respect to the qualitative evaluation measures examined in this study. The statistics for the other measures (VI, GCE, BDE) are widely in favor of the DCM-SVFMM even for higher SNR.

Finally, an important remark of our experimental investigation is that when only (*Lab*) color features are used, the comparison revealed a very high performance for the DCM-SVFMM with respect to the other segmentation methods. The explanation is that when the added noise is not smoothed out by PCA as it is the case in MRF texture features, the Ncut and GBMS methods are not robust and provide erroneous segmentations. This may be significant for applications involving nontextured images.

The execution time of the algorithm depends upon many parameters such as the dimensions of the image, the number of segments and most importantly the dimension of the feature vectors. The experiments presented here were performed using MATLAB on a standard PC having a 2 GHz dual core processor with 2 GB RAM. For the 3-class image of 256×256 pixels,

using as feature only the intensity, the algorithm performs one EM iteration per second and convergence may be achieved in 10–50 iterations, depending upon the amount of noise. On the other hand, the Berkeley images are much larger (480×320) and using the 8-D MRF features yields an execution time of 30 s per iteration in average.

VI. CONCLUSION AND FUTURE WORK

We have presented a fully Bayesian methodology for image segmentation and modeling using a hierarchical spatially constrained mixture model. The model takes into account spatial information by imposing distinct smoothness priors on the probabilities of each cluster. The contextual mixing proportions for each pixel are constrained to follow a Dirichlet compound multinomial distribution, thus, avoiding the projection step in the standard EM algorithm [42]. A closed form solution for updates of the coefficients of the Dirichlet distributions was derived in a MAP framework and the contextual mixing proportions are explicitly estimated. The experimental results confirm the theoretical significance of the proposed method by showing improvements with respect to the performance of spatially varying mixture models, employing a projection step to constrain the values of the mixing proportions. Comparison with state of the art techniques for image segmentation reveals that the proposed

method performs at the same level in natural image segmentation. Moreover, in presence of Gaussian noise, the herein presented algorithm provides more coherent segmentations in all cases where the state-of-the-art methods fail. A notable advantage of the DCM-SVFMM method is that it does not need any parameter to be fixed before training which is not the case neither in graph based methods nor in the mean-shift algorithm where the result strongly depends upon the selected parameters.

Important open questions in a segmentation algorithm concern the estimation of the number of image segments as well as the automatic determination of salient features in the case of multidimensional feature vectors [64], [65]. Recently proposed image models have shown to be very efficient for image denoising [33]. A comparison of the method proposed here to such models for image segmentation is also envisaged.

Additional shortcomings may be related to a specific segmentation algorithm. In our case, a first issue is the initialization of the mixtures which is a well known problem. Another limitation is the lack of an implicit or explicit boundary that could be obtained by a presegmentation of the image using for instance a superpixel representation. These issues constitute the subject of on going research.

REFERENCES

[1] R. Haralick and L. G. Shapiro, "Survey: Image segmentation techniques," *Comput. Vis. Graph. Image Process.*, vol. 29, pp. 100–132, 1985.

[2] N. Pal and S. Pal, "A review of image segmentation techniques," *Pattern Recognit.*, vol. 26, pp. 1277–1294, 1993.

[3] R. Xu and D. Wunsch, II, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, May 2005.

[4] M. M. A. Jain and P. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, 1999.

[5] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[6] P. F. Felzenswalb and D. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, 2004.

[7] R. Zabih and V. Kolmogorov, "Spatially coherent clustering using graph cuts," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2004, vol. 2, pp. 437–444.

[8] T. Cour, F. Benezit, and J. Shi, "Spectral segmentation with multiscale graph decomposition," in *Proc. 2005 IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 2005, vol. 2, pp. 1124–1131.

[9] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[10] M. A. Carreira-Perpinan, "Fast nonparametric clustering with Gaussian blurring mean-shift," in *Proc. Int. Conf. Machine Learning*, 2006, pp. 153–160.

[11] Y. Ma, H. Derksen, W. Hong, and J. Wright, "Segmentation of multivariate mixed data via lossy data coding and compression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1546–1562, Sep. 2007.

[12] A. Yang, J. Wright, S. Sastry, and Y. Ma, "Unsupervised segmentation of natural images via lossy data compression," *Comput. Vis. Image Understand.*, vol. 110, no. 2, pp. 212–225, 2008.

[13] D. Titterington, A. Smith, and U. Makov, *Statistical Analysis of Finite Mixture Distributions*.   Hoboken, NJ: Wiley, 1985.

[14] G. McLachlan, *Finite Mixture Models*.   Hoboken, NJ: Wiley, 2000.

[15] C. M. Bishop, *Pattern Recognition and Machine Learning*.   New York: Springer-Verlag, 2006.

[16] K. Blekas, D. I. Fotiadis, and A. Likas, "Greedy mixture learning for multiple motif discovery in biological sequences," *Bioinformatics*, vol. 19, no. 5, pp. 607–617, 2003.

[17] H. Greenspan, G. Dvir, and Y. Rubner, "Context-dependent segmentation and matching in image databases," *Comput. Vis. Image Understand.*, vol. 93, no. 1, pp. 86–109, 2004.

[18] P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *J. Roy. Statist. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[19] A. Blake and A. Zisserman, *Visual Reconstruction*.   Cambridge, MA: MIT Press, 1987.

[20] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[21] G. Potamianos and J. Goutsias, "Stochastic approximation algorithms for partition function estimation of Gibbs random fields," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1948–1965, Nov. 1997.

[22] C. Zhu and D. Mumford, "Prior learning and Gibbs reaction-diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 11, pp. 1236–1250, Nov. 1997.

[23] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 25–47, 2000.

[24] J. Marroquin, E. Arce, and S. Botello, "Hidden Markov measure field models for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1380–1387, 2003.

[25] 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 Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.

[26] Z. Kato and T. C. Pong, "A Markov random field image segmentation model for color textured images," *Image Vis. Comput.*, vol. 24, no. 10, pp. 1103–1114, 2006.

[27] X. Yu, T. D. Bui, and A. Krzyzak, "Robust estimation for range image segmentation and reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 530–538, May 1994.

[28] S. Z. Li, "MAP image restoration and segmentation by constrained optimization," *IEEE Trans. Image Process.*, vol. 7, no. 12, pp. 1730–1735, Dec. 1998.

[29] A. N. Rajagopalan and S. Chaudhuri, "An MRF model-based approach to simultaneous recovery of depth and restoration from defocused images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 7, pp. 577–589, Jul. 1999.

[30] L. Quan, J. Wang, P. Tan, and L. Yuan, "Image-based modeling by joint segmentation," *Int. J. Comput. Vis.*, vol. 75, no. 1, pp. 135–150, 2007.

[31] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[32] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[33] S. Lyu and E. P. Simoncelli, "Modeling multiscale subbands of photographic images with fields of Gaussian scale mixtures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 4, pp. 693–706, Apr. 2009.

[34] B. Goossens, A. Pizurica, and W. Philips, "Image denoising using mixtures of projected Gaussian scale mixtures," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1689–1702, Aug. 2009.

[35] M. Woolrich, T. Behrens, C. Beckmann, and S. Smith, "Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data," *IEEE Trans. Med. Imag.*, vol. 24, no. 1, pp. 1–11, Jan. 2005.

[36] M. Woolrich and T. Behrens, "Variational Bayes inference of spatial mixture models for segmentation," *IEEE Trans. Med. Imag.*, vol. 25, no. 10, pp. 1380–1391, Dec. 2006.

[37] J. G. Propp and D. B. Wilson, "Exact sampling with coupled Markov chains and applications to statistical mechanics," *Random Struct. Algorithms*, vol. 9, no. 1 and 2, pp. 223–252, 1996.

[38] P. Bremaud, *Markov Chains: Gibbs fields, Monte Carlo Simulation and Queues*.   New York: Springer–Verlag, 2001.

[39] G. Celeux and J. Diebolt, "Asymptotic properties of a stochastic EM algorithm for estimating mixing proportions," *Commun. Statist. Stochastic Models*, vol. 9, pp. 599–613, 1993.

[40] J. Diebolt and E. H. S. Ip, "Stochastic EM: Method and application," in *Markov Chain Monte Carlo in practice*, W. R. Gilks, S. Richardson, and D. J. Speigelhalter, Eds.   London, U.K.: Chapman & Hall, 1996.

[41] S. Sanjay-Gopal and T. Hebert, "Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm," *IEEE Trans. Image Process.*, vol. 7, no. 7, pp. 1014–1028, Jul. 1998.

[42] K. Blekas, A. Likas, N. Galatsanos, and I. Lagaris, "A spatially constrained mixture model for image segmentation," *IEEE Trans. Neural Netw.*, vol. 16, no. 2, pp. 494–498, Mar. 2005.

[43] C. Nikou, N. Galatsanos, and A. Likas, "A class-adaptive spatially variant mixture model for image segmentation," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1121–1130, Apr. 2007.

[44] G. Sfikas, C. Nikou, N. Galatsanos, and C. Heinrich, "Spatially varying mixtures incorporating line processes for image segmentation," *J. Math. Imag. Vis.*, vol. 36, no. 2, pp. 91–110, 2010.

[45] N. Johnson, S. Kotz, and N. Balakrishnan*, Discrete Multivariate Distributions*. Hoboken, NJ: Wiley, 1997.

[46] D. M. Blei, A. Y. Ng, and M. Jordan, "Latent Dirichlet allocations," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[47] R. Madsen, D. Kauchak, and C. Elkan, "Modeling word burstiness using the Dirichlet distribution," in *Proc. 22nd Int. Conf. Machine Learning*, Bonn, Germany, 2005, pp. 545–552.

[48] E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Wilsky, "Learning hierarchical models of scenes, objects and parts," in *Proc. IEEE Int. Conf. Computer Vision*, 2005, vol. 2, pp. 1331–1336.

[49] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Int. Conf. Computer Vision*, 2005, vol. 2, pp. 1816–1823.

[50] W. Pieczynski, D. Benboudjema, and P. Lanchantin, "Statistical image segmentation using triplet Markov fields," in *Proc. SPIE's Int. Symp. Remote Sensing*, 2002, pp. 22–27.

[51] W. Pieczynski and D. Benboudjema, "Multisensor triplet Markov random fields and theory of evidence," *Image Vis. Comput.*, vol. 24, no. 1, pp. 61–69, 2006.

[52] D. Benboudjema and W. Pieczynski, "Unsupervised statistical segmentation of nonstationary images using triplet Markov fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1367–1378, 2007.

[53] C. Elkan, "Clustering documents with an exponential-family approximation of the Dirichlet compound multinomial distribution," in *Proc. 23rd Int. Conf. Machine Learning*, Pittsburgh, PA, 2006, pp. 289–296.

[54] E. T. Whittaker and G. Robinson, "The solution of the cubic," in *The Calculus of Observations: A Treatise on Numerical Mathematics*. New York: Dover, 1967, ch. 62, pp. 124–126.

[55] G. Birkhoff and S. McLane*, A Survey of Modern Algebra*. Natick, MA: AK Peters, 1997.

[56] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int. Conf. Computer Vision*, Jul. 2001, vol. 2, pp. 416–423.

[57] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1026–1038, Aug. 2002.

[58] M. Varma and A. Zisserman, "Texture classification: Are filter banks necessary?," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 2003, vol. 2, pp. 691–698.

[59] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 929–944, Jun. 2007.

[60] M. Meila, "Comparing clusterings: An axiomatic view," in *Proc. Int. Conf. Machine Learning*, 2005, pp. 577–584.

[61] J. Freixenet, X. Munoz, D. Raba, J. Marti, and X. Cuff, "Yet another survey on image segmentation: Region and boundary information integration," in *Proc. Eur. Conf. Computer Vision*, 2002, pp. 408–422.

[62] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. IEEE Int. Conf. Computer Vision*, 2003, vol. 1, pp. 10–17.

[63] G. Mori, "Guiding model search using segmentation," in *Proc. 10th IEEE Int. Conf. Computer Vision*, 2005, vol. 2, pp. 1417–1423.

[64] C. Constantinopoulos, M. K. Titsias, and A. Likas, "Bayesian feature and model selection for Gaussian mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 6, pp. 1013–1018, Jun. 2006.

[65] C. Constantinopoulos and A. Likas, "Active Bayesian mixture learning for image modeling and segementation using low level features," in *Proc. IEEE Workshop on Machine Learning for Signal Processing*, Maynooth, Ireland, 2006, pp. 319–323.

Christophoros Nikou (S'97–M'05) received the Diploma degree in electrical engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1994, and the DEA and Ph.D. degrees in image processing and computer vision from Louis Pasteur University, Strasbourg, France, in 1995 and 1999, respectively.

During 2001, he was a Senior Researcher with the Department of Informatics, Aristotle University of Thessaloniki. From 2002 to 2004, he was with Compucon S.A., Thessaloniki. Since 2004, he has been with the Department of Computer Science, University of Ioannina, Greece, where he was a Lecturer (2004–2009) and he is now an Assistant Professor. His research interests mainly include image processing and computer vision and their application to medical imaging.

Aristid C. Likas (S'91–M'96–SM'03) received the Diploma degree in electrical engineering and the Ph.D degree in electrical and computer engineering, both from the National Technical University of Athens, Greece.

Since 1996, he has been with the Department of Computer Science, University of Ioannina, Greece, where he is currently an Associate Professor. His research interests include machine learning, neural networks, statistical signal processing, and bioinformatics.

Nikolas P. Galatsanos (SM'95) received the Diploma of Electrical Engineering from the National Technical University of Athens, Greece, in 1982, and the M.S.E.E. and Ph.D. degrees from the Electrical and Computer Engineering Department, University of Wisconsin, Madison, in 1984 and 1989, respectively.

He was on the faculty of the Electrical and Computer Engineering Department, Illinois Institute of Technology, Chicago, from 1989 to 2002. He was with the Department of Computer Science, University of Ioannina, Ioannina, Greece, from 2002–2008. Presently, he is with the Department of Electrical and Computer Engineering, University of Patras, Patras, Greece. He coedited a book titled *Image Recovery Techniques for Image and Video Compression and Transmission* (Kluwer, 1998). His research interests center around image processing and statistical learning problems for medical imaging and visual communications applications.

Dr. Galatsanos has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and the *IEEE Signal Processing Magazine*, and he served as an Associate Editor for the *Journal of Electronic Imaging*.