

A Class-Adaptive Spatially Variant Mixture Model for Image Segmentation

Christophoros Nikou, Nikolaos Galatsanos and Aristidis Likas

Department of Computer Science, University of Ioannina,

PO Box186, 45110 Ioannina, Greece,

Phone: + (30) 26510 98802

{cnikou, galatsanos, arly}@cs.uoi.gr

Abstract

We propose a new approach for image segmentation based on a hierarchical and spatially variant mixture model. According to this model the pixel labels are random variables and a smoothness prior is imposed on them. The main novelty of this work is a new family of smoothness priors for the label probabilities in spatially variant mixture models. These Gauss-Markov random field-based priors allow all their parameters to be estimated in closed form via the maximum a posteriori (MAP) estimation using the Expectation-Maximization (EM) methodology. Thus, it is possible to introduce priors with multiple parameters that adapt to different aspects of the data. Numerical experiments are presented where the proposed MAP algorithms were tested in various image segmentation scenarios. These experiments demonstrate that the proposed segmentation scheme compares favorably to both standard and previous spatially constrained mixture model-based segmentation.

Index Terms

Image and texture segmentation, clustering-based segmentation, Gaussian mixture model, spatial smoothness constraints, smoothness prior, Gauss-Markov random field, simultaneously autoregressive prior, maximum *a posteriori* (MAP) estimation, Expectation-Maximization (EM) algorithm.

EDICS: SEG-CLST, SEG-STAT

I. INTRODUCTION

Image segmentation is the process of grouping image pixels based on the coherence of certain attributes such as intensity, color, or texture. Many approaches have been proposed to solve the image segmentation problem. For surveys of this topic the reader may refer to [1], [2]. In this paper, we will focus our attention to image segmentation methods based on clustering. Clustering is the process of arranging data into groups having common characteristics and is a fundamental problem in many fields of science [3], [4]. Thus, image segmentation can be viewed a special type of clustering. Usually, in image segmentation, our data, the image pixels have spatial locations associated with them. Thus, apart from the commonality of attributes such as intensity, color or texture, commonality of location is an important characteristic of the grouping that we are seeking in image segmentation.

More specifically, in this paper we will focus our attention on clustering methods based on the modeling of the probability density function (pdf) of the data via finite mixture models (FMM) [5], [6], [7]. Modeling the pdf of data with FMM is a natural way to cluster data because it automatically provides a grouping of the data based on the components of the mixture that generated them. More specifically, FMM are based on the assumption that each datum originates from one component of the mixture according to some probability. Thus, this probability can be used to be assign each datum to the component that has most likely generated it. Furthermore, the likelihood of an FMM is a rigorous measure for the clustering performance [7]. FMM based pdf modeling has been used successfully in a number of applications ranging from bioinformatics [8] to image retrieval [9].

The parameters of the FMM model with Gaussian components can be estimated very efficiently through maximum likelihood (ML) using the Expectation-Maximization (EM) algorithm [5], [10], [7]. Furthermore, it can be shown that Gaussian components allow efficient representation of any pdf. Thus, mixture models with Gaussian components, in other words Gaussian mixture models (GMM), are the ones used in most applications [7].

In mixture model-based image segmentation the image pixels are considered independent, thus it is straightforward to use a GMM and ML for this problem. In [11], the performance of different algorithms that estimate GMMs for image segmentation was evaluated. A drawback of the ML approach for this application 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, several approaches were proposed. Local Bayesian segmentation methods with a comparison of mixture estimation algorithms are investigated in [12]. A Markov random field (MRF) approach was proposed in [13]. Caillol *et al.* [14] introduced fuzziness in Gaussian

mixtures and modeled spatial information in both the segmentation and parameter estimation levels.

The spatially variant finite mixture model (SVFMM) was proposed [15]. The SVFMM considers the pixel labels as random variables instead of parameters as the "classical" EM-ML formulation. It assumes a MRF prior on the data. A maximum *a posteriori* (MAP) estimation using a Gibbs MRF-based prior for the pixel labels is used [15]. This prior enforces spatial smoothness of the pixel labels, and generates clusters that are spatially continuous. Based on this model, a number of approaches have been proposed for various applications [16], [17].

The MAP algorithm proposed in [15] cannot find the pixel labels in closed form and used a gradient projection algorithm. In [18], the problem of computing the pixel labels for the SVFMM was improved using quadratic programming. This was shown to produce both better segmentations and values of the criterion function. In this paper, we build upon the SVFMM framework rather than the works in [16] and [17].

The main shortcoming of SVFMM in [15] and [18] is the use of a smoothness prior which is rather inflexible and does not adapt to the data. This is a consequence of the fact that it uses only one parameter that cannot be estimated automatically from the image. Our experience with this method indicates that appropriate values for this parameter can be found. However, they require a tedious trial-and error process. This clearly makes incorporation of more than one such parameter in the prior prohibitive. In this paper, this difficulty is bypassed since the new prior that we propose allows estimation of all its parameters in closed form the data. Thus, the introduction of priors that can adapt to different aspects of the data is possible. Herein, we use priors that are cluster and directionally adaptive and all their parameters are automatically estimated from the image. We also demonstrate via numerical experiments for image segmentation problems with both single and multi-dimensional feature vectors that the proposed priors improve the segmentation performance of the SVFMM.

In what follows, background for SVFMM is given in section II. The proposed priors are presented in section III. In section IV, the proposed MAP algorithm is presented. Numerical experiments are presented in section V and finally conclusions and directions for future research are given in section VI.

II. BACKGROUND

Let $x^i = (x_1^i, x_2^i, \dots, x_L^i)^T$ denote the vector of features representing i^{th} spatial location (pixel), ($i = 1, \dots, N$), of an L -dimensional image modeled as independently distributed random variables. Here, L is the number of features per pixel (e.g. intensity, textural features, location etc.). The SVFMM [15] provides a modification of the classical FMM approach [7], [5] 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 $\pi^i = (\pi_1^i \pi_2^i \dots \pi_K^i)^T$ where K is the number of components. We define $\mathbf{\Pi} = \{(\pi^1)^T, (\pi^2)^T, \dots, (\pi^N)^T\}$ as the set of probability vectors and $\Theta = \{\theta^1, \theta^2, \dots, \theta^K\}$ the set of component parameters. The variables π_j^i represent the probabilities of the i^{th} pixel to belong to the j^{th} cluster (or class) and must satisfy the constraints

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

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

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

where $\phi(x^i | \theta^j)$ is a Gaussian distribution with parameters $\theta_j = \{\mu_j, \Sigma_j\}$, where $\mu_j = (\mu_{j,1}, \mu_{j,2}, \dots, \mu_{j,L})^T$ 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 Θ as parameters. In this study, we have considered a diagonal covariance matrix $\Sigma_j = \text{diag}\{\sigma_{j,l}\}$, with $l = 1 \dots L$ for each class j . The SVFMM in [18] uses a prior density based on the Gibbs distribution for the random variables $\mathbf{\Pi}$ given by:

$$p(\mathbf{\Pi}) = \frac{1}{Z} e^{-U(\mathbf{\Pi})} \quad (3)$$

with

$$U(\mathbf{\Pi}) = \beta \sum_{i=1}^N V_{\mathcal{N}_i}(\mathbf{\Pi}), \quad (4)$$

where Z is a normalizing constant called "Partition Function", β is the prior parameter that controls the degree to which smoothness is imposed. This parameter is analogous to the *regularization parameter* used to impose smoothness in ill-posed inverse problems, see for example [19]. The function $V_{\mathcal{N}_i}(\mathbf{\Pi})$ denotes the clique potential function of the pixel label vectors $\{\pi^i\}$ 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 $d_{i,m}$ specifies the distance between two label vectors $\{\vec{\pi}^i\}$ and $\{\vec{\pi}^m\}$:

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

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

$$g(d_{i,m}) = \frac{1}{1 + \frac{1}{d_{i,m}}} \quad (5)$$

which is robust to outliers [20]. The value of the partition function Z in general depends on β . However, this relationship is unknown unless the penalty function has a very simple form [21], [22]. For many penalty functions including the one in equation (5) this relation is unknown.

Therefore, denoting \mathbf{X} the set of pixels $\{x^i\}$, with $i = 1, \dots, N$, which we assume to be statistically independent and following Bayes rules, we obtain the posterior probability density function given by:

$$q(\mathbf{\Pi}|\mathbf{X}; \Theta) \propto \prod_{i=1}^N p(\mathbf{\Pi}) f(x^i|\mathbf{\Pi}, \Theta) \quad (6)$$

with the log-density:

$$L(\mathbf{\Pi}|\mathbf{X}; \Theta) = \sum_{i=1}^N \log f(x^i|\mathbf{\Pi}, \Theta) + \log p(\mathbf{\Pi}) \quad (7)$$

The EM algorithm [10] for MAP estimation, based on the SVFMM [15], requires the computation of the conditional expectation values of the "hidden" variables at the E-step of iteration step t :

$$z_j^{i(t)} = \frac{\pi_j^{i(t)} \phi(x^i|\theta_j^{(t)})}{\sum_{p=1}^K \pi_j^{i(t)} \phi(x^i|\theta_p^{(t)})} \quad (8)$$

In the M-step, the expected log-likelihood of the complete data is used [10]. This log-likelihood for mixture models is well-known [5], [7] and is linear in the "hidden" variables. The details of the derivation of this log-likelihood for the SVFMM can be found in [15]. The maximization of

$$Q(\mathbf{\Pi}, \Theta|\mathbf{\Pi}^{(t)}, \Theta^{(t)}) = \sum_{i=1}^N \sum_{j=1}^K z_j^{i(t)} \{ \log(\pi_j^i) + \log(\phi(x^i|\theta_j^i)) \} - \beta \sum_{i=1}^N \sum_{m \in \mathcal{N}_i} g(d_{i,m}) \quad (9)$$

corresponding to the complete data log-likelihood, yields the model parameters. The function $Q(\cdot)$ in (9) can be maximized independently for each parameter with the following update equations of the mixture model parameters at step $t + 1$:

$$\mu_{j,l}^{(t+1)} = \frac{\sum_{i=1}^N z_j^{i(t)} x_l^{i(t)}}{\sum_{i=1}^N z_j^{i(t)}}, \quad \left(\sigma_{j,l}^{(t+1)} \right)^2 = \frac{\sum_{i=1}^N z_j^{i(t)} \left[x_l^{i(t)} - \mu_{j,l}^{(t+1)} \right]^2}{\sum_{i=1}^N z_j^{i(t)}}. \quad (10)$$

Several methods have been proposed to compute the contextual mixing proportions π_i^j of each normal density. A generalized EM scheme, based on the gradient projection algorithm was used in [15]. In [18], an approach based on quadratic programming was proposed. This approach was shown to improve both the criterion function (9) and the classification of the SVFMM.

III. THE NEW PRIOR FOR THE PIXEL LABELS

As already mentioned, the main drawbacks of the SVFMM [15], [18] are that one parameter, the scalar β in Eq. (9) is used to capture the smoothness of all the clusters and in all directions. Furthermore, the value of this parameter β cannot be estimated from the data in an easy way. This stems from the fact that for penalties of the form of Eq. (5) the relationship of the partition function and β is difficult to find. If the penalty function is quadratic this relationship is known and closed-form estimates of β from the data are easily obtained.

Motivated by the above we consider a Gauss-Markov random field prior probability for $\mathbf{\Pi}$ in (7) which is given by

$$p(\mathbf{\Pi}) \propto \beta^{-NK} \exp \left[-\frac{1}{2} \frac{\sum_{i=1}^N \sum_{j=1}^K \left(\sum_{m \in \mathcal{N}_i} (\pi_j^i - \pi_j^m) \right)^2}{\beta^2} \right], \quad (11)$$

where \mathcal{N}_i is the neighborhood for the i^{th} pixel.

The statistical assumptions used in this prior can be clearly explained if the predictor of the label π_j^i is defined by the mean of its spatial neighbors as

$$\hat{\pi}_j^i = \frac{1}{|\mathcal{N}_i|} \sum_{m \in \mathcal{N}_i} \pi_j^m \quad (12)$$

where $|\mathcal{N}_i|$ is the number of pixels in the neighborhood, and the prediction error

$$\pi_j^i - \hat{\pi}_j^i = \epsilon_j^i. \quad (13)$$

This prior is based on the assumption that $p(|\mathcal{N}_i| \epsilon_j^i) = N(0, \beta^2)$. In other words, the prediction errors of the labels, for all spatial locations and all clusters are independent identically distributed Gaussian random variables with zero mean and variance β^2 .

Obviously, the prior in Eq. (11) is based on a simplistic assumption and does not capture the fact that the statistics of each cluster might be different. To capture this property one can use a distinct variance β_j^2 for each cluster. Then, the prior is given by

$$p(\mathbf{\Pi}) \propto \prod_{j=1}^K \beta_j^{-N} \exp \left[-\frac{1}{2} \frac{\sum_{i=1}^N \left(\sum_{m \in \mathcal{N}_i} (\pi_j^i - \pi_j^m) \right)^2}{\beta_j^2} \right]. \quad (14)$$

In this prior, the parameter β_j^2 captures the spatial smoothness of cluster j . Thus, this prior can enforce smoothness of different degree in each cluster and adapts better to the data.

Moreover, this prior can be further refined by allowing smoothness that varies both within cluster and along different spatial directions. In other words, parameters $\beta_{j,d}^2$ can be used to express not only the class variance for cluster j but also the variance within cluster j at a certain spatial direction d (e.g. horizontal, vertical and diagonal pixel variances). In that case, the prior is given by

$$p(\mathbf{\Pi}) \propto \prod_{d=1}^D \prod_{j=1}^K \beta_{j,d}^{-N} \exp \left[-\frac{1}{2} \frac{\sum_{i=1}^N \left(\sum_{m \in \mathcal{N}_i^d} (\pi_j^i - \pi_j^m) \right)^2}{\beta_{j,d}^2} \right], \quad (15)$$

where D is the total number of the considered directions (generally 4), $\beta_{j,d}^2$ is the variance of class j only considered for pixels having adjacency type d and \mathcal{N}_i^d , $d = 1, \dots, D$ is the neighborhood of the i^{th} pixel in direction d . In the general case, $D = 4$ directions (horizontal, vertical and 2 diagonal directions).

This prior can be also explained in a similar fashion as the one in eq. (11). One can define in the same manner as in eq. (12) and (13) D directional predictors and prediction error for each one of the K data clusters. Then, this prior assumes that the $D \times K$ prediction errors for each direction and cluster are independent and identically distributed for each pixel Gaussian random variables. In other words, we have $p(|\mathcal{N}_i^d| \epsilon_{j,d}^i) = N(0, \beta_{j,d}^2)$ where $|\mathcal{N}_i^d|$ the number of pixels in neighborhood \mathcal{N}_i^d . Priors of this nature have been applied successfully for images (pixel values) to regularize various ill-posed inverse problems in order to generate smooth image estimates from noisy observations. Priors of this nature are called *simultaneously autoregressive* (SAR) (see [19], [23], [24] and [25]). In this work we adopt such priors, by applying them to the pixel labels, for SVFMM based image segmentation.

IV. MAP ESTIMATION

Using the previously proposed priors one can derive the corresponding MAP algorithms using the EM methodology in a similar manner as in [15], [18]. Since the only difference from the previous SVFMM are the priors all parameters of the SVFMM model except the pixel labels and the parameters of priors are identically computed. Since the priors in equations (11) and (14) can be considered as special cases of the prior in eq. (15) we will derive only the MAP algorithm for latter. The prior in eq. (15) yields the following MAP function to be maximized for the M-step of the EM algorithm

$$Q(\mathbf{\Pi} | \mathbf{\Pi}^{(t)}; \Theta, \Theta^{(t)}) =$$

$$\sum_{i=1}^N \sum_{j=1}^K \left\{ z_j^{i(t)} [\log(\pi_j^i) + \log(\phi(x^i|\theta^j))] - \sum_{d=1}^D \left[\frac{1}{2} \log(\beta_{j,d}^2) - \frac{1}{2} \frac{\left(\sum_{m \in \mathcal{N}_i^d} (\pi_j^i - \pi_j^m) \right)^2}{\beta_{j,d}^2} \right] \right\} \quad (16)$$

To compute the model parameters $\pi_j^{i(t+1)}$ and $\beta_{j,d}^{2(t+1)}$ at time step $(t+1)$ of the EM algorithm we have to maximize (16) with respect to π_j^i or 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 $\left(\sum_{m \in \mathcal{N}_i^d} (\pi_j^i - \pi_j^m) \right)^2$ occurs once as the probability of the central pixel and $|\mathcal{N}_i^d|$ times as a neighbor π_j^m of different pixels if a first order neighborhood is used. Thus, $\frac{\partial Q}{\partial \pi_j^i} = 0$ gives a second degree equation with respect to $\pi_j^{i(t+1)}$:

$$2|\mathcal{N}_i^d| \sum_{p=1}^D \prod_{\substack{d=1 \\ d \neq p}}^D \beta_{j,d}^{2(t)} \left(\pi_j^{i(t+1)} \right)^2 - 2 \sum_{p=1}^D \prod_{\substack{d=1 \\ d \neq p}}^D \beta_{j,d}^{2(t)} \sum_{m \in \mathcal{N}_i^p} \pi_j^m \pi_j^{i(t+1)} - z^{i(t)} \prod_{d=1}^D \beta_{j,d}^{2(t)} = 0 \quad (17)$$

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

$$B_j^{(t)} = \sum_{p=1}^D \prod_{\substack{d=1 \\ d \neq p}}^D \beta_{j,d}^{2(t)}, \quad (18)$$

which is the sum of all combinations of products between $\beta_{j,p}^2$, for $j = 1, \dots, D$, when a specific $\beta_{j,p}^2$, for $p = 1, \dots, D$ is excluded from the product, equation (17) becomes:

$$2|\mathcal{N}_i^d| B_j^{(t)} \left(\pi_j^{i(t+1)} \right)^2 - 2B_j^{(t)} \sum_{m \in \mathcal{N}_i^p} \pi_j^m \pi_j^{i(t+1)} - z^{i(t)} \prod_{d=1}^D \beta_{j,d}^{2(t)} = 0 \quad (19)$$

The solution of Eq. (19) for π_j^i at time step $t+1$ of the EM algorithm is:

$$\pi_j^{i(t+1)} = \frac{B_j^{(t)} \sum_{m \in \mathcal{N}_i^p} \pi_j^m \pm \sqrt{\left[B_j^{(t)} \sum_{m \in \mathcal{N}_i^p} \pi_j^m \right]^2 + 2|\mathcal{N}_i^d| z^{i(t)} B_j^{(t)} \prod_{d=1}^D \beta_{j,d}^{2(t)}}}{2|\mathcal{N}_i^d| B_j^{(t)}} \quad (20)$$

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

$$\beta_{j,d}^{2(t+1)} = \frac{1}{N} \sum_{i=1}^N \left(\sum_{m \in \mathcal{N}_i^d} (\pi_j^{i(t+1)} - \pi_j^m) \right)^2, \quad (21)$$

for $j = 1, \dots, K$ and $d = 1, \dots, D$. Analogous are the log-likelihoods to be maximized for the models involving the priors in (11) and (14).

In the standard EM algorithm, the neighborhood \mathcal{N}_i^d in expressions (20) and (21), for the computation of the parameters at time step $t + 1$, includes pixels with label parameter vectors computed at time t . In our experiments, we have noticed that if pixels with updated label parameter vectors (computed for time $t + 1$), as well as pixels whose label vectors have not yet been updated (computed at time t) are combined then the algorithm converges faster. However, there is no theoretical proof neither for the speed-up of the algorithm nor for the convergence itself.

The overall EM algorithm may be summarized as follows:

- Initialize the mixture model component parameters $\theta_j = \{\mu_j, \sigma_j\}$, $j = 1, \dots, K$ and the contextual mixing proportion vectors π^i for each pixel $i = 1, \dots, N$.
- Do until the MAP functional (16) does not change significantly.
 - E-step
 - * Calculate the posterior probabilities $z_j^{i(t)}$ for the i^{th} pixel to belong to the j^{th} class (8).
 - M-step
 - * Calculate the new mixture model parameters (10).
 - * Calculate the new contextual mixing proportions (20).
 - * Project the contextual mixing proportions onto the constraints (1) using the quadratic programming algorithm in [18].
 - * Calculate the class variances (21).
- End

V. NUMERICAL EXPERIMENTS

The performance of the proposed approach is illustrated with a number of numerical examples. Since 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. The termination criterion of the EM algorithm, considered here, was convergence defined as the percentage of change in the log-likelihood (16) between two consecutive iterations to be less than 0.001%, or $\left| \frac{\Delta Q}{Q} \right| < 10^{-5}$.

We present comparisons for image segmentation between the standard finite mixture model (FMM) [6], the spatially variant finite mixture model (SVFMM) [15] improved by the quadratic optimization method we proposed in [18] and the herein proposed directional class-adaptive spatially variant finite mixture model (DCA-SVFMM). We also present the comparisons of the intermediate versions of our final prior model: the adaptive spatially variant finite mixture model (A-SVFMM) having only one global variance for the prior distribution (11) and the class-adaptive spatially variant finite mixture

model (CA-SVFMM) (14) having one variance per cluster (14). In all our experiments we used a 8 closest pixels neighborhood system for \mathcal{N}_i and a one neighbor system for the directional system \mathcal{N}_i^d . Results on piece-wise constant, color and textured images are presented. Furthermore, the proposed algorithm was tested on the Berkeley image segmentation data base [26].

A. Piece-wise constant image segmentation

Figure 1(a) shows a simulated three-class image with intensities for the three classes 30, 125 and 220. The contextual mixing proportions are 0.37, 0.30 and 0.33 respectively. Figure 1(b) shows the same image corrupted by zero mean white Gaussian noise. Because image contrast is what we are most interested in for quantifying the segmentation, we define the signal to noise ratio (SNR) as [16]:

$$SNR = \frac{\text{mean inter-class contrast}}{\text{standard deviation of the noise}}$$

Thus, for the image in figure (1)(b) we have $SNR = 1.0$.

We also define the measure of correct classification ratio (CCR):

$$CCR = \sum_{j=1}^K \frac{|GT_j \cap Seg_j|}{|GT|} \quad (22)$$

where GT_j is the ground truth for the j^{th} cluster, Seg_j describes the pixels the algorithm classified to cluster j and $GT = \bigcup_{j=1}^K GT_j$.

Figure 2 shows the segmentation of the corrupted image in fig. 1(b) obtained by the SVFMM (fig. 2(a)) with a fixed regularization parameter and the variants of our approach: the A-SVFMM (fig. 2(b)), the CA-SVFMM (fig. 2(c)) and the the DCA-SVFMM (fig. 2(d)). The best results are obtained with the final refinement of our model, with the consideration of neighborhood directions jointly with the class adaptive prior. Table I presents the percentage of correctly classified pixels (CCR metric) for the different methods. Finally, the robustness of our approach is illustrated in table II where the statistics of the CCR measure for 30 realizations of the noise configuration are shown.

In order to demonstrate the limits of location based features in image segmentation and support the use of a spatially variant mixture model we present the segmentation of the image in fig. 1(a) using location as a feature. Each pixel is represented by a three dimensional vector whose components are the pixel intensity and the two image coordinates. The image was segmented by the standard FMM algorithm without any prior assumption on spatial variations. As can be seen in fig. 1(c), location based features have the tendency to produce clusters with erroneous spatial arrangements because they assign the same weight to image intensity and coordinates.

In a second experiment we evaluate our method in a more difficult configuration. We have modified the values of the original image (fig. 1) in order to decrease the inter-class distances. We have fixed

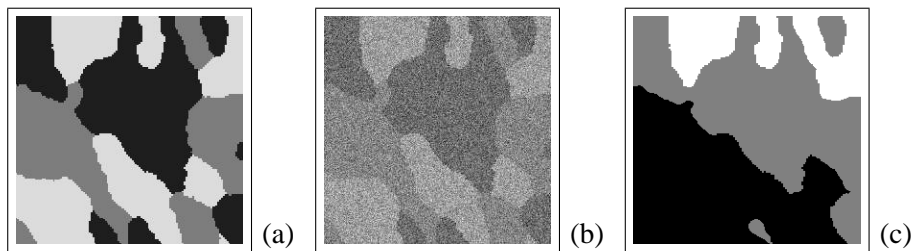


Fig. 1. (a) The 3-class test image used in the experiments described in the text. Intensity means are 30, 125 and 220. The contextual mixing proportions are 0.37, 0.30 and 0.33 respectively. (b) The image in (a) degraded by zero mean white Gaussian noise with SNR=1.0. (c) Segmentation of the image in (a) combining intensity and location features. Each pixel is represented by a three dimensional vector whose components are the pixel intensity and the two image coordinates. The segmentation algorithm is the standard FMM without spatial prior.

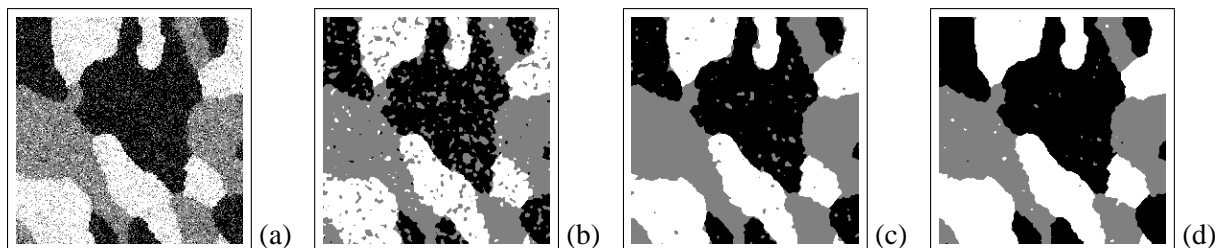


Fig. 2. Three class segmentation of the image presented in figure 1(b) using (a) SVFMM, (b) A-SVFMM, (c) CA-SVFMM and (d) DCA-SVFMM.

TABLE I

Percentage of correctly classified pixels (CCR) for the segmentation of the degraded image in figure 1.

| 3-class CCR (%) - Problem 1 | | |
|-----------------------------|---------|---------|
| | SNR=2.0 | SNR=1.0 |
| FMM | 77.3 | 64.5 |
| SVFMM | 86.5 | 80.7 |
| A-SVFMM | 93.3 | 89.8 |
| CA-SVFMM | 96.3 | 96.1 |
| DCA-SVFMM | 98.4 | 97.5 |

the three classes to gray levels of 110, 130 and 150 respectively and added white Gaussian noise to the image. Thus, we have obtained images with signal to noise ratios of 2.0 and 1.0. We then applied our algorithm to those images and compared the segmentation results with respect to the performance of the SVFMM. The results are summarized in table III where the superior performance of our method is underpinned.

A last experiment was conducted in order to illustrate how the variance of the noise affects the

TABLE II

Statistical behavior of the CCR. The experiment presented in table I was repeated 30 times with different realizations of the noise. The median, mean and standard deviation of the CCR for the DCA-SVFMM is presented.

| 3-class CCR statistics - Problem 1 | | | |
|------------------------------------|--------|-------|---------|
| | Median | Mean | S. dev. |
| SNR=2.0 | 97.0% | 96.9% | 1.3% |
| SNR=1.0 | 95.8% | 96.1% | 1.8% |

TABLE III

Percentage of correctly classified pixels for the segmentation of degraded versions of the image in figure 1 with decreased inter-cluster contrast. The original values of the cluster means are $\mu_1 = 110$, $\mu_2 = 130$ and $\mu_3 = 150$.

| 3-class CCR (%) - Problem 2 | | |
|-----------------------------|---------|---------|
| | SNR=2.0 | SNR=1.0 |
| FMM | 67.8 | 61.3 |
| SVFMM | 82.2 | 70.2 |
| A-SVFMM | 95.3 | 88.7 |
| CA-SVFMM | 95.3 | 90.3 |
| DCA-SVFMM | 96.5 | 95.6 |

intra-class variances of the priors for the pixel probability labels β_j , $j = 1, 2, 3$. We have corrupted the three regions of the image in fig. 1 with different amounts of noise and applied the class-adaptive segmentation algorithm which implies one variance per cluster (CA-SVFMM). The results are presented in table IV. Looking at such results one has to keep in mind that parameters β_j do not refer directly to the noise in the observations but to the probability of the pixel belonging to the respective cluster taking into account its neighbors. However, we observe that parameters β_j may vary across different clusters by orders of magnitude. This demonstrates the ability of the model to adapt to the data.

B. Application to texture segmentation

We have also evaluated and compared the proposed algorithms for the segmentation of multidimensional images. This type of images may be obtained directly from sensors, such as multispectral satellite images or RGB color television images. Also, they may be created through preprocessing in order to extract significant features characterizing the image content (e.g. textural features). Here, we present an example from both cases.

A first experiment concerns the segmentation of a textured image. Fig. 3(a) shows a composite

TABLE IV

The mean and standard deviation for each cluster estimated by the CA-SVFMM for the 3-class segmentation of image of figure 1 when different amount of noise is added to each class. The values of the estimated class variances β_j , $j = 1, 2, 3$ of the contextual mixing proportions are also presented. The CCR was 97%. The original values of the cluster means are $\mu_1 = 30$, $\mu_2 = 125$ and $\mu_3 = 220$ and the image was degraded with noise variances of 30, 60 and 90 respectively.

| 3-class segmentation - Problem 3 | | | | | | | | | |
|----------------------------------|---------|------------|---------------------|---------|------------|----------------------|---------|------------|---------------------|
| | μ_1 | σ_1 | β_1 | μ_2 | σ_2 | β_2 | μ_3 | σ_3 | β_3 |
| Simulated | 30.9 | 29.8 | — | 124.2 | 59.4 | — | 218.7 | 88.9 | — |
| CA-SVFMM | 28.5 | 33.8 | 30×10^{-3} | 126.0 | 62.6 | 117×10^{-3} | 222.5 | 90.8 | 44×10^{-3} |

image of 4 natural textures from the Brodatz collection [27]. We have degraded the textured image with Gaussian noise in order to make segmentation more challenging and obtained two different images of SNR of 1.0 and 0.5 respectively (fig.3(b)-(c)). We have extracted textural features using a filter bank of 40 bandpass Gabor filters for 8 equally spaced angles and 5 different radial lengths [28]. Eight features were used for segmentation. These features were created by selecting 8 from the available 40 Gabor filter responses and then processed them as in [29]. This approach has been also successfully evaluated for texture segmentation in [30]. For segmentation with multidimensional features we used a diagonal covariance matrix for the mixture of the Gaussian distributions.

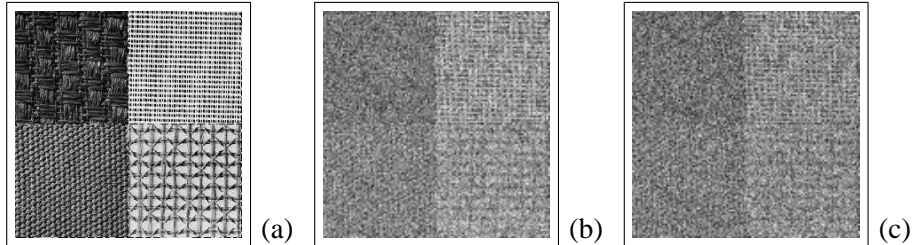


Fig. 3. (a) A composite 4-texture image and its noisy versions with (b) SNR=1 and (c) SNR=0.5.

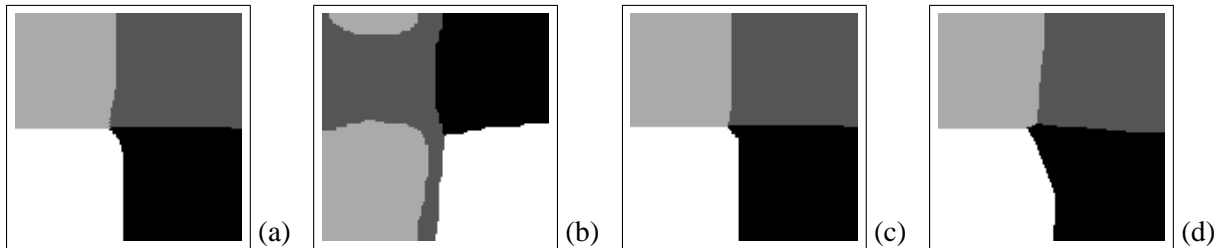


Fig. 4. Four class segmentation of the images presented in fig. 3(b)-(c) by applying the (a)-(b) SVFMM and (c)-(d) DCA-SVFMM methods to the dyadic Gabor filter bank responses of the original images.

The segmentation results for the degraded textured images are summarized in table V where the CCR percentage is presented for each method. Also, fig. 4 shows that the DCA-SVFMM approach performs very well even in the presence of significant amount of noise. Notice how the cross-like separation is preserved in fig. 4(d) which is not the case in fig. 4(b).

TABLE V

Percentage of correctly classified pixels for the segmentation of the degraded images of figure 3.

| 4-class texture CCR (%) | | |
|-------------------------|---------|---------|
| | SNR=1.0 | SNR=0.5 |
| FMM | 62.1 | 50.8 |
| SVFMM | 93.8 | 85.4 |
| A-SVFMM | 93.8 | 86.9 |
| CA-SVFMM | 95.2 | 87.1 |
| DCA-SVFMM | 95.5 | 92.2 |

C. Application to RGB image segmentation

Moreover, we have experimented on the segmentation of RGB natural images. We present two examples from the Berkeley data base [26]. Fig. 5(a)-(d) shows a color image and its RGB components. The image consists mainly of three color components: the white church wall, the red cupola and the dark blue sky. Other red and dark blue regions are also present (mainly on the windows). We have added to the image zero mean white Gaussian noise with different standard deviation for each color component leading to SNRs of 2dB, 4dB and 3dB for the R, G and B components respectively. The degraded images are shown in fig. 5(e)-(h). We have applied the SVFMM and the DCA-SVFMM algorithms to obtain a 3-class segmentation of the degraded RGB image. At first, we applied the segmentation algorithms to the noise-free image in order to obtain a baseline of their behavior. As depicted in fig. 6 (a)-(b), the DCA-SVFMM method segments better the red component by preserving the sharp edge between the cupola and the sky. Furthermore, the bottom right red window is more accurately extracted by our method. Also, we can observe that the narrow shadows on the wall right at the bottom of the cupola are slightly better preserved by the SVFMM. However, this was obtained after manual tuning of the parameter β of the Gibbs distribution (4) as mentioned at the end of the section.

The results for the degraded images are presented in fig. 6 (c)-(d). Due to absence of ground truth, the evaluation of the segmentation is only qualitative. As in the single-dimensional case, our new method provides a better segmentation, especially in the red component which underwent the most

significant degradation. In the case of the SVFMM, many pixels of the cupola were classified as sky or wall which is not the case for the DCA-SVFMM approach.

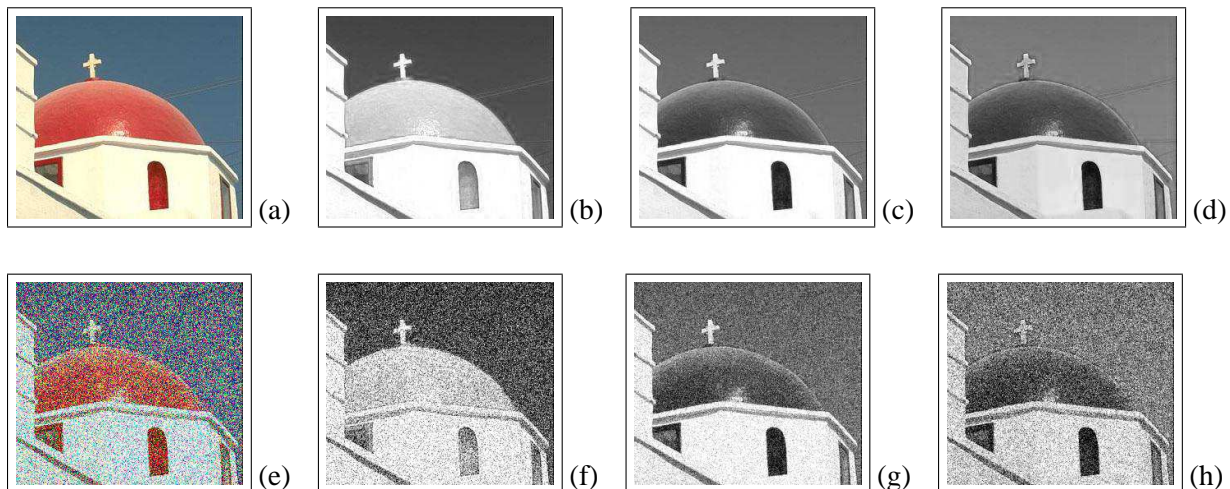


Fig. 5. (a) A color image and (b)-(d) its RGB components. (e) The color image reconstructed after having corrupted its color components by zero-mean white Gaussian noise with different standard deviation for each component. The corresponding SNR are 2dB, 4dB and 3dB for the R, G and B components respectively. (f)-(h) The degraded RGB components.

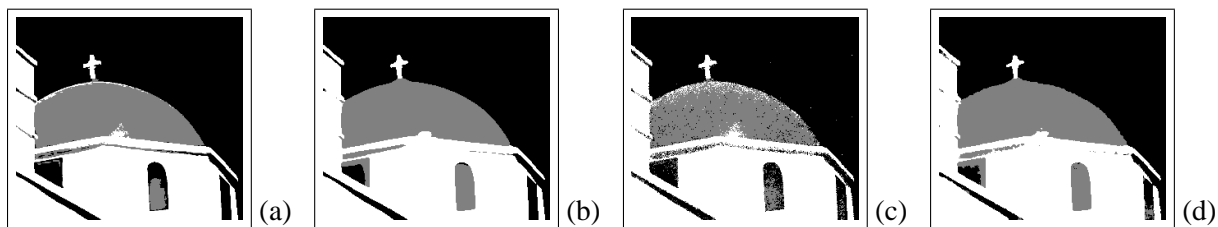


Fig. 6. Three class segmentation of the noise-free image presented in fig. 5(a) by applying the (a) SVFMM and (b) DCA-SVFMM methods to the RGB components of the image. Three class segmentation of the degraded image presented in fig. 5(e) by applying the (c) SVFMM and (d) DCA-SVFMM methods to the RGB components of the image.

An interesting property of the proposed prior is that it can take into account intra-cluster statistics. For the segmentation problem, this property seems interesting and eventually well suited to take into account, to a certain extent, the so-called *shading effect* sometimes found in real world images (e.g. regions with globally smooth shading variations, gradually changing color, such as sky, lake, wall etc.). This is the case at the top of the red cupola where we can observe reflections as well as at the wall where the color is not constantly white due to shadows 5(a). This effect generally induces inappropriate and undesirable over-segmentation which is not the case here. The directional and class adaptive parameters $\beta_{j,d}$ are summarized in table VI which illustrates that these effects were taken

into account by assigning relatively larger variances to these regions.

TABLE VI

The parameters $\beta_{j,d}$, $j = 1, 2, 3$, $d = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ for the segmentation of the noise free image in fig 5(a) by the DCA-SVFMM.

| $\beta_{j,d} \times 10^{-3}$ | | | | |
|------------------------------|---------------|----------------|----------------|-----------------|
| Direction | $d = 0^\circ$ | $d = 45^\circ$ | $d = 90^\circ$ | $d = 135^\circ$ |
| $\beta_{1,d}$ (cupola) | 139 | 293 | 155 | 293 |
| $\beta_{2,d}$ (sky) | 30 | 39 | 10 | 38 |
| $\beta_{3,d}$ (wall) | 119 | 279 | 161 | 280 |

We have to notice that in all of the experiments, in the case of the SVFMM algorithm, we have presented the results for the *best* regularization β parameter of the Gibbs distribution (4). This parameter was obtained after a tedious search performed heuristically since there is no trivial method to estimate this parameter from the data for this model. More precisely, the best parameter β varies between 2.1 and 2.5 for the 3-class images and between 1.8 and 2.2 for the 4-class (textured images) and 5-class cases. In contrast, for the herein proposed approach all the parameters of the prior are estimated automatically from the data and this is one of the main strengths of our approach.

Finally, it is important to bring to notice the faster convergence of the proposed approach as compared to SVFMM in [15], [18]. The different variants of our model required approximately 10-30 iterations as compared to the previous SVFMM, that required over 100 iterations for convergence.

D. Evaluation with the Berkeley image segmentation data base

Finally, we have compared the results of our algorithm to the manual segmentations provided by the Berkeley image segmentation data base [26]. This benchmark consists of a set of images along with their ground truth segmentation maps which were provided by different individuals. Evaluation of a segmentation algorithm when multiple ground truths are available is a non trivial task. However, a probabilistic evaluation can be achieved by the probabilistic Rand (PR) index [31]. This index was conceived for the case of hard segmentation maps and was used in [32] for the comparisons of segmentations of the Berkeley data base.

The PR index, between a segmentation map S_{test} to be evaluated and a set of M ground truth images $S_g = \{S_1, S_2, \dots, S_M\}$ is given by:

$$PR(S_{test}, S_g) = \frac{2}{N(N-1)} \sum_{\substack{i,j \\ i < j}} [c_{ij}p_{ij} + (1 - c_{ij})(1 - p_{ij})] \quad (23)$$

where $c_{ij} = 1$ if pixels i and j belong to the same cluster and $c_{ij} = 0$ if pixels i and j belong to different clusters. The normalization term is the inverse of the number of all possible unordered pairs of N pixels and p_{ij} is the ground truth probability that pixels i and j belong to the same cluster, computed as the mean pixel pair *relationship* among all the ground truth images. The term *relationship* implies a binary value *true* or *false* meaning that the pixels of the pair belong or not to the same cluster.

The PR index takes values between 0 and 1. A score of 0 indicates that every pixel pair in the machine segmentation has the opposite *relationship* as every pair in the ground truth segmentations while a score of 1 indicates that every pixel pair in the machine segmentation has the same *relationship* as every pair in the ground truth images. If two pixels are in the same region in most of the ground truth images they are penalized accordingly for not being in the same region in the machine segmentation and vice versa. As a consequence, the PR index is robust to segmentation maps that result by splitting or merging segments of the ground truth which is a desirable property [33].

The CA-SVFMM algorithm was applied to a set of 30 color images of the Berkeley segmentation data base with several configurations for the number of clusters (parameter K) for each image. We have chosen both textured and non textured images in order to evaluate the algorithm. The PR indices for a subset of the data with the value of parameter K that provided the maximum index is presented in table VII. For comparison purposes, we present, in the same table, for the same parameter K , the PR index for the SVFMM algorithm [18], with $\beta = 2.0$ which is a good compromise according to the performed experiments. In all cases, the proposed class adaptive algorithmic approach outperforms the standard SVFMM algorithm. Some segmentation examples from the application of CA-SVFMM to images of the Berkeley data base, including the segmentation maps with the best K , are shown in fig. 7.

VI. CONCLUSION AND FUTURE WORK

We have presented a hierarchical and spatially constrained mixture model for image segmentation. This model takes into account spatial information by imposing distinct smoothness priors on the probabilities of each cluster and pixel neighborhoods. Experimental results have shown that our approach improves significantly not only standard FMM segmentation but also its spatially variant version. Moreover, the number of iterations of the EM algorithm is reduced significantly compared to the SVFMM [15], [18].

Important open questions for FMM-based clustering are how the number of model components can be selected automatically and which features (in the multidimensional case) should be used.





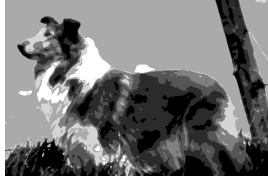


| | | | |
|---|---|--|---|
|  |  |  |  |
| 232038 | $K = 5, PR = 0.57$ | $K = 10, PR = 0.62$ | $K = 15, PR = 0.59$ |
|  |  |  |  |
| 102061 | $K = 7, PR = 0.68$ | $K = 12, PR = 0.70$ | $K = 15, PR = 0.73$ |
|  |  |  |  |
| 108073 | $K = 3, PR = 0.44$ | $K = 4, PR = 0.49$ | $K = 5, PR = 0.45$ |
|  |  |  |  |
| 247085 | $K = 5, PR = 0.54$ | $K = 6, PR = 0.61$ | $K = 7, PR = 0.64$ |
|  |  |  |  |
| 118020 | $K = 9, PR = 0.67$ | $K = 12, PR = 0.70$ | $K = 15, PR = 0.66$ |
|  |  |  |  |
| 310007 | $K = 7, PR = 0.70$ | $K = 8, PR = 0.68$ | $K = 9, PR = 0.67$ |

Fig. 7. Image segmentation results from the Berkeley segmentation data base [26]. The first column presents the original color images and the rest of the columns present the segmentation maps of our CA-SVFMM method along with the predetermined number of clusters K and the PR index for each case.

TABLE VII

The maximum PR index and the corresponding number of clusters for 30 images of the Berkeley segmentation data base [26] when the CA-SVFMM algorithm is applied. For comparison purposes, the PR index for the SVFMM algorithm, with $\beta = 2.0$, for the same parameter K is presented.

| PR index | | | |
|----------|-----|----------|-------|
| Image # | K | CA-SVFMM | SVFMM |
| 102061 | 15 | 0.73 | 0.70 |
| 80099 | 2 | 0.81 | 0.78 |
| 108073 | 4 | 0.49 | 0.43 |
| 118035 | 3 | 0.65 | 0.63 |
| 134008 | 2 | 0.63 | 0.62 |
| 135069 | 2 | 0.87 | 0.81 |
| 188091 | 4 | 0.65 | 0.59 |
| 207056 | 5 | 0.64 | 0.60 |
| 232038 | 10 | 0.62 | 0.54 |
| 310007 | 7 | 0.70 | 0.62 |
| 323016 | 10 | 0.67 | 0.63 |
| 238011 | 3 | 0.84 | 0.82 |
| 299091 | 5 | 0.58 | 0.54 |
| 314016 | 5 | 0.52 | 0.45 |
| 368016 | 6 | 0.77 | 0.75 |
| 271031 | 5 | 0.63 | 0.57 |
| 253036 | 4 | 0.72 | 0.66 |
| 247085 | 7 | 0.64 | 0.60 |
| 28075 | 6 | 0.63 | 0.56 |
| 113044 | 5 | 0.62 | 0.61 |
| 12003 | 3 | 0.46 | 0.44 |
| 106024 | 4 | 0.68 | 0.64 |
| 183055 | 3 | 0.51 | 0.48 |
| 118020 | 12 | 0.70 | 0.63 |
| 90076 | 5 | 0.52 | 0.49 |
| 41044 | 5 | 0.74 | 0.66 |
| 163014 | 5 | 0.60 | 0.58 |
| 42049 | 7 | 0.62 | 0.58 |
| 112082 | 4 | 0.62 | 0.58 |
| 124084 | 3 | 0.56 | 0.50 |
| Mean | | 0.64 | 0.60 |
| S. dev. | | 0.09 | 0.10 |
| Median | | 0.65 | 0.60 |

These questions are still the subject of on going research [34], [35], [36]. For FMM-based image segmentation these questions are also very important for certain segmentation scenarios. For example, in texture segmentation selecting the appropriate subset of responses from the Gabor filter bank is an important question. However, the use of smoothness priors makes the FMM even more complex. Thus, addressing such questions is out of the scope of this paper and subject of future research.

REFERENCES

- [1] R. Haralick and L. G. Shapiro, "Survey: image segmentation techniques," *Computer Vision, Graphics and Image Processing*, vol. 29, pp. 100–132, 1985.
- [2] N. Pal and S. Pal, "A review of image segmentation techniques," *Pattern Recognition*, vol. 26, pp. 1277–1294, 1993.
- [3] R. Xu and D. Wunsch II, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.
- [4] M. M. A. Jain and P. Flynn, "Data clustering: a review," *ACM Computing surveys*, vol. 31, no. 3, pp. 264–323, 1999.
- [5] D. Titterton, A. Smith, and U. Makov, *Statistical analysis of finite mixture distributions*. John Wiley and Sons, 1985.
- [6] C. M. Bishop, *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press Inc., 1995.
- [7] G. McLachlan, *Finite mixture models*. Wiley-Interscience, 2000.
- [8] 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.
- [9] H. Greenspan, G. Dvir, and Y. Rubner, "Context-dependent segmentation and matching in image databases," *Computer Vision and Image Understanding*, vol. 93, no. 1, pp. 86–109, 2004.
- [10] P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.
- [11] T. Lei and J. Udupa, "Performance evaluation of finite normal mixture model-based image segmentation techniques," *IEEE Transactions on Image Processing*, vol. 12, no. 10, pp. 1153–1169, 2003.
- [12] A. Peng and W. Pieczynski, "Adaptive mixture estimation and unsupervised local Bayesian image segmentation," *Graphical Models and Image Processing*, vol. 57, no. 5, pp. 389–399, 1995.
- [13] B. Chalmond, "An iterative Gibbsian technique for reconstruction of m-ary images," *Pattern Recognition*, vol. 22, no. 6, pp. 747–761, 1989.
- [14] H. Caillol, W. Pieczynski, and A. Hillon, "Estimation of fuzzy Gaussian mixture and unsupervised statistical image segmentation," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 425–440, 1997.
- [15] S. Sanjay-Gopal and T. Hebert, "Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm," *IEEE Transactions on Image Processing*, vol. 7, no. 7, pp. 1014–1028, 1998.
- [16] 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*, vol. 20, no. 1, pp. 45–57, 2001.
- [17] M. Woolrich, T. Behrens, C. Beckmann, and S. Smith, "Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data," *IEEE Transactions on Medical Imaging*, vol. 24, no. 1, pp. 1–11, 2005.
- [18] K. Blekas, A. Likas, N. Galatsanos, and I. Lagaris, "A spatially constrained mixture model for image segmentation," *IEEE Transactions on Neural Networks*, vol. 16, no. 2, pp. 494–498, 2005.

- [19] N. Galatsanos and A. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 322–336, 1992.
- [20] J. Huber, *Robust statistics*. New York: John Wiley and Sons, 1985.
- [21] G. Potamianos and J. Goutsias, "Stochastic approximation algorithms for partition function estimation of Gibbs random fields," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1948 – 1965, 1997.
- [22] D. M. Higdon, J. E. Bowsher, V. E. Johnson, T. G. Turkington, D. R. Gilland, and R. J. Jaszczak, "Fully bayesian estimation of Gibbs hyperparameters for emission computed tomography data," *IEEE Transactions on Medical Imaging*, vol. 16, no. 5, pp. 516–526, 1997.
- [23] D. B. Ripley, *Spatial Statistics*. New York: John Wiley and Sons, 1981.
- [24] R. Molina and B. D. Ripley, "Using spatial models as priors in astronomical images analysis," *Journal of Applied Statistics*, vol. 16, pp. 193–206, 1989.
- [25] J. Chantas, N. Galatsanos, and A. Likas, "Bayesian restoration using a new nonstationary edge preserving image prior," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 2987–2997, 2006.
- [26] 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 *Proceedings of the 8th International Conference on Computer Vision*, vol. 2, July 2001, pp. 416–423.
- [27] P. Brodatz, *Textures: a photographic album for artists and designers*. New York: Dover, 1966.
- [28] A. C. Bovik, "Multichannel texture analysis using localized spatial filters," *IEEE Transactions on Pattern Recognition and Machine Intelligence*, vol. 12, no. 1, pp. 55–73, 1990.
- [29] A. Jain and F. Farrokhinia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition*, vol. 24, no. 12, pp. 1167–1186, 1991.
- [30] T. Randen and J. H. Husøy, "Filtering for texture classification: a comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 291–310, 1999.
- [31] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "A measure for objective evaluation of image segmentation algorithms," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 3, June 2005, pp. 34–41.
- [32] C. Pantofaru and M. Hebert, "A comparison of image segmentation algorithms," The Robotics Institute, Carnegie Mellon University, Tech. Rep. CMU-R1-TR-05-40, 2005.
- [33] R. Unnikrishnan and M. Hebert, "Measures of similarity," in *Proceedings of the IEEE Workshop on Computer Vision Applications*, 2005, pp. 394–400.
- [34] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1188, 2003.
- [35] C. Constantinopoulos, M. K. Titsias, and A. Likas, "Bayesian feature and model selection for Gaussian mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, pp. 1013–1018, 2006.
- [36] L. Wolf and A. Shashua, "Feature selection for unsupervised and supervised inference: the emergence of sparsity in a weight-based approach," *Journal of Machine Learning Research*, vol. 3, pp. 1855–1887, 2005.