# Image and Video Segmentation by Combining Unsupervised Generalized Gaussian Mixture Modeling and Feature Selection

Mohand Saïd Allili, *Member, IEEE,* Djemel Ziou, Nizar Bouguila, and Sabri Boutemedjet

***Abstract*—In this letter, we propose a clustering model that efficiently mitigates image and video under/over-segmentation by combining generalized Gaussian mixture modeling and feature selection. The model has flexibility to accurately represent heavy-tailed image/video histograms, while automatically discarding uninformative features, leading to better discrimination and localization of regions in high-dimensional spaces. Experimental results on a database of real-world images and videos showed us the effectiveness of the proposed approach.**

***Index Terms*—Feature selection, image/video segmentation, minimum message length (MML), mixture of generalized Gaussian distributions (MoGG).**

## I. Introduction

Segmentation is an important topic in computer vision and image/video processing. For example, newly established multimedia standards, such as MPEG-4 and MPEG-7, are based on the object content of videos [13]. Therefore, successful representation and processing in these standards require efficient segmentation algorithms. This efficiency lies—among other things—on the capacity of the segmentation to yield meaningful regions (i.e., reduce over/under-segmentation) [8]. We note that over-segmentation occurs when the number of regions is over-estimated, leading to *insignificant* small regions; whereas under-segmentation occurs when real regions are erroneously fused. Both problems may compromise the application at hand using segmentation (e.g., video coding and indexing based on object shape).

In the last few years, finite mixture models have been established as a powerful technique to achieve segmentation in unsupervised fashion [1], [3], [6]. In this technique, clusters are represented by the mixture components whose number can be automatically determined using criteria, such as the minimum description length (MDL), the Akaike (AIC), etc. [7]. Segmentation, then, amounts to dividing the feature space into different compact clusters which correspond to regions in the image/video space. Recently, a new mixture model based on the generalized Gaussian distribution (GGD) (called MoGG) was proposed in [3], for which the number of components is determined using the minimum message length (MML) criterion [7], [15]. The MoGG mitigates efficiently image over-segmentation induced by features with non-Gaussian distribution. We recall that *non-Gaussianity* is an intrinsic property of many image/video features distribution [e.g., discrete cosine transform (DCT) and wavelet coefficients, pixels difference, etc.] [14]. Also, the presence of noise and/or outliers in images/videos can produce non-Gaussian data. Both cases arise in heavy-tailed histograms, with sharped or flat-shaped modes. Since the number of mixture components is determined automatically, the number of regions can be easily over-estimated using the Gaussian mixture model [3].

In general, the use of multiple features (e.g., color, texture, etc.) ensures better performance for segmentation. However, the performance of mixture modeling may be substantially deteriorated in the presence of many irrelevant features [11]. In segmentation, for instance, texture features on one particular orientation may discriminate a texture pattern in an image/video, while features on other orientations can be uniform (i.e., irrelevant) and useless for this segmentation. Therefore, introducing a *feature selection* (FS) mechanism, to remove irrelevant features, is of prominent importance to estimate the real number of regions. Finally, note that due to the interrelation between the chosen feature subset and the number of mixture components, FS should be established in unsupervised way and simultaneously with model learning (i.e., the number of regions with their parameters) [5], [9]. Supervised FS has been used in the past to improve the performance of segmentation [2], [12]. However, unsupervised FS together with mixture model learning have not been addressed so far for image/video segmentation.

In this letter, we propose an efficient segmentation framework which incorporates unsupervised FS in the MoGG model. This combination aims at enhancing the representation of non-Gaussian data and eliminating irrelevant features, leading to significant reduction of over/under-segmentation. Our model selects automatically the optimal number of regions

Manuscript received May 23, 2008; revised July 9, 2009, September 28, 2009, and March 1, 2010; accepted May 6, 2010. Date of publication September 20, 2010; date of current version October 8, 2010. This research was funded by the Natural Sciences and Engineering Research Council of Canada, Ottawa, ON, Canada. This paper was recommended by Associate Editor S. Pankanti.

M. S. Allili is with the Department of Computer Science and Engineering, Université du Québec en Outaouais, Gatineau, QC J8X 3X7, Canada.

D. Ziou and S. Boutemedjet are with the Department of Computer Science, University of Sherbrooke, Sherbrooke, QC J1K 2R1, Canada.

N. Bouguila is with the Concordia Institute for Information Systems Engineering, Concordia University, Montréal, QC H4B 1R6, Canada.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2010.2077483

and the relevant features by optimizing a single objective based on the MML criterion [15]. This unified objective allows for accurate discrimination and identification of the real image/video regions inside high-dimensional feature vectors, while penalizing mixture overfitting.

This letter is organized as follows. Section II outlines the proposed segmentation model and the details of our learning algorithm. Section III shows experiments on real-world image and video segmentation. Section IV presents the computation complexity of our algorithm. We end the letter with a conclusion and some prospective future work.

## II. Proposed Model

In [3], we proposed a multidimensional version of the GGD. Given a $d$-dimensional feature vector $\vec{x} = (x_1, ..., x_d) \in \mathbb{R}^d$, the GGD distribution is defined as follows:

$$p(\vec{x}|\vec{\mu}, \vec{\sigma}, \vec{\lambda}) = \prod_{l=1}^{d} \frac{\lambda_l K(\lambda_l)}{2\sigma_l} \exp\left\{ -A(\lambda_l) \left| \frac{x_l - \mu_l}{\sigma_l} \right|^{\lambda_l} \right\} \quad (1)$$

where $\vec{\mu} = (\mu_1, ..., \mu_d)$, $\vec{\sigma} = (\sigma_1, ..., \sigma_d)$ and $\vec{\lambda} = (\lambda_1, ..., \lambda_d)$ are vectors of location, scale, and shape parameters, respectively. Also, we have $K(\lambda) = \frac{\left[\frac{\Gamma(3/\lambda)}{\Gamma(1/\lambda)}\right]^{\frac{1}{2}}}{\Gamma(1/\lambda)}$ and $A(\lambda) = \left[\frac{\Gamma(3/\lambda)}{\Gamma(1/\lambda)}\right]^{\frac{\lambda}{2}}$, with $\Gamma(\cdot)$ denoting the gamma function. Each parameter $\lambda_l \geq 0$, $l = 1, \ldots, d$, controls the shape of the GGD and determines whether it is peaked or flat in the $l$th dimension.

Given that $M$ regions are present in the image/video, we can use an MoGG to model the distribution of the image/video data. Each region $j$ is modeled, then, as a $d$-dimensional GGD distribution parameterized by $\{\vec{\theta}_{jl}\}, l = 1, \ldots, d$, such that $\vec{\theta}_{jl} = (\lambda_{jl}, \mu_{jl}, \sigma_{jl})$ are the parameters of a univariate GGD component $p(x_{il}|\theta_{jl})$. If we consider $p_j$ as the prior probability of assigning each pixel to the $j$th region, then the core step of segmentation is to first estimate $\vec{\Theta}^* = (p_j^*, \theta_{jl}^*), (j = 1, \ldots, M$ and $l = 1, \ldots, d)$. Note that the MoGG assumes that all the features have equal importance. However, with high-dimensional descriptors such as color, texture, etc., features do not contribute equally in discriminating among the exiting regions. Indeed, some features may be uniform or unimodal making the distinction between the real regions hard to achieve. Therefore, we extend the MoGG model by expressing the discrimination power of each feature separately. Let $\phi_l$ be a binary variable, set to 0 when the $l$th feature is irrelevant (i.e., uniform) and to 1, otherwise. Then, the distribution of each $x_{il}$ can be approximated as follows [5], [9]:

$$p(x_{il}|\theta_{jl}^*, \vec{\varphi}_l^*, \phi_l) \simeq \left(p(x_{il}|\theta_{jl})\right)^{\phi_l} \left(p(x_{il}|\vec{\varphi}_l)\right)^{1-\phi_l} \quad (2)$$

where now $\vec{\Theta}^* = (p_j^*, \theta_{jl}^*, \vec{\varphi}*_l)$. The star superscript denotes the unknown true distribution of the $l$th feature of the region, and both $p(x_{il}|\theta_{jl})$ and $p(x_{il}|\varphi_l)$ are univariate GGDs. In (2), $\phi_l$ is a hidden variable set to 1 from the data in every case where the $l$th feature is multimodal. We consider each $\phi_l$ is a Bernoulli variable with $p(\phi_l = 1) = \rho_{l1}$ and $p(\phi_l = 0) = \rho_{l2}$, such that $\rho_{l1} + \rho_{l2} = 1$.

Note that (2) leads to false positives (i.e., uninformative features that are identified as relevant) when a feature is defined by only overlapped components. Therefore, we generalize the definition of feature relevance by considering the irrelevant component $p(.|\vec{\varphi}_l)$ as a common mixture of GGDs independent of the region labels [5]. This choice is also motivated by the ability of the mixture to approximate almost any arbitrary distribution of the irrelevant features. We consider $K$ as the number of components in this common mixture, with the parameters $\varphi_{1l}, \ldots, \varphi_{Kl}$ for each feature. Let $\pi_{kl}$ (with $\sum_{k=1}^{K} \pi_{kl} = 1$) be the prior probability that $x_{il}$ is generated by the $k$th component of the common mixture, given that the $l$th feature is irrelevant (i.e., $\phi_l = 0$). From this, we derive the final model for segmentation with FS as

$$p(\vec{x}_i|\vec{\Theta}) = \sum_{j=1}^{M} p_j \prod_{l=1}^{d} \left( \rho_{l1}\, p(x_{il}|\theta_{jl}) + \rho_{l2} \sum_{k=1}^{K} \pi_{kl}\, p(x_{il}|\varphi_{kl}) \right). \quad (3)$$

We consider the notation $\mathbf{p} = (p_1, ..., p_M)$, $\vec{\rho}_l = (\rho_{l1}, \rho_{l2})$ and $\vec{\pi}_l = (\pi_{l1}, ..., \pi_{lK})$. The set of all model parameters is $\vec{\Theta} = (\mathbf{p}, \theta_{jl}, \varphi_{kl}, \vec{\rho}_l, \vec{\pi}_l)$, $j = 1, ..., M$ and $l = 1, ..., d$.

Note that the parameters $\vec{\Theta}$ and the numbers $M$ and $K$ are unknown and need to be identified from the data. The maximum likelihood method is the most commonly used approach for the estimation of $\vec{\Theta}$. However, ML favors both higher values for $K$ and $M$, which leads to over-segmentation. To avoid this problem, we use and MML approach which identifies models with less complexity [5], [9], [15]. The message length of the image/video data $\mathcal{X}$ is, then, given by

$$\begin{aligned} MessL \simeq & -\log p(\vec{\Theta}) + \frac{1}{2}\log|I(\vec{\Theta})| + \frac{c}{2}\left(1 + \log\frac{1}{12}\right) \\ & - \log p(\mathcal{X}|\vec{\Theta}) \end{aligned} \quad (4)$$

where $p(\vec{\Theta})$, $I(\vec{\Theta})$, and $p(\mathcal{X}|\vec{\Theta})$ denote the prior distribution, the Fisher information matrix, and the likelihood, respectively. The constant $c = M + d + 3dM + 4dK$ in (4) is the total number of parameters. To facilitate the calculation of the MML, it is common sense to assume the independence of the different groups of parameters, which allows the factorization of both $|I(\vec{\Theta})|$ and $p(\vec{\Theta})$. We approximate the Fisher information $|I(\Theta)|$ from the complete likelihood which assumes labeled observations [4], [7]. In absence of any other knowledge, we adopt the uninformative Jeffrey's prior for each group of parameters as prior distribution. From this, we obtain the following MML objective:

$$\begin{aligned} MessL(M, K) = & -\log p(\mathcal{X}|\vec{\Theta}) + \frac{c}{2}\log N + \frac{3d}{2}\sum_{j=1}^{M}\log p_j \\ & + \frac{3}{2}\sum_{l=1}^{d}\sum_{k=1}^{K}\log \pi_{kl} + \frac{c}{2}\left(1 + \log\frac{1}{12}\right) \\ & + \frac{3M}{2}\sum_{l=1}^{d}\log\rho_{l1} + \frac{3K}{2}\sum_{l=1}^{d}\log\rho_{l2} \end{aligned} \quad (5)$$

which we minimize under the constraints $0 < p_j \leq 1$, $0 < \rho_{l1} \leq 1$, $0 < \pi_{kl} \leq 1$, and $\sum_{j=1}^{M} p_j = 1$, $\sum_{k=1}^{K} \pi_{kl} = 1$, in a manner similar to [7]. To accelerate convergence, we initialize the expectation-maximization using the Fuzzy $C$-means algorithm. The update formulas used by EM are provided in Appendix I.

## III. Experiments

We conducted experiments on several examples of real-world image and video segmentation. We compared the segmentation results obtained using the proposed approach (denoted by MoGG+FS) with those obtained using: 1) an MoGG model without FS [3]; 2) an MoGG model with FS, but having a fixed parameter $K = 1$; 3) a Gaussian mixture model (denoted by MoG) [6]; and 4) an MoG with FS (denoted by MoG+FS). We develop the applications of image and video segmentation in two separate subsections.

### A. Application to Color/Texture Image Segmentation

Here, we use for each pixel $(u, v)$ a feature vector $\vec{x}(u, v)$ which combines color and texture information. For color, we use the RGB color space. For texture, we use 24 features calculated from the color correlogram of the pixel neighborhood as defined in [1]. To test the performance of the proposed method, we compared the segmentation accuracy of all the tested models against the ground truth (GT) (i.e., segmentation performed manually). For this purpose, we use a database of 2000 images which include the Berkeley benchmark [10] and images downloaded from the Corel and Freefoto datasets. To quantitatively measure the segmentation accuracy, we use the following objective criteria:

*Boundary localization error* ($\epsilon_1$): measures the misalignment of regions between a tested segmentation and the GT. This error is defined by [10] as

$$\epsilon_1 = \frac{1}{N} \sum_{(u,v)} \min \left\{ \epsilon_{(u,v)}(TM, GT), \epsilon_{(u,v)}(GT, TM) \right\} \quad (6)$$

where $\epsilon_{(u,v)}(TM, GT) = \frac{|S_i - S'_j|}{|S_i|}$ and $\epsilon_{(u,v)}(GT, TM) = \frac{|S'_j - S_i|}{|S_j|}$; $S_i(u, v)$ and $S'_j(u, v)$ are the segments[1] containing the pixel $(u, v)$ in a TM and the GT results, respectively. The symbol "$-$" stands for the set difference operator and $N$ is the number of pixels in the image.

*Amount of over/under-segmentation* ($\epsilon_2$): measures the amount of over/under-segmentation produced by each TM, by comparison to the GT. A set of segments $S_{m1}, ..., S_{m\ell}$ in the TM over-segment a segment $S'_m$ in the GT *iff*: $\forall i \in \{1, ..., \ell\}$ : $|S_{mi} \cap S'_m| \geq k|S_{mi}|$ and $\sum_{i=1}^{\ell} |S_{mi} \cap S'_m| \geq k|S'_m|$, where $k$ is a threshold that we set here to 0.75 as suggested in [1]. We define the error $\epsilon_2$ as the sum of the number of segments in the GT that are over-segmented in the TM, and the number of segments in the TM that are over-segmented in the GT.

Fig. 1 shows the segmentation for a sample of images chosen randomly from the dataset. For each image, we run the tested models 10 times and we report the average ($\bar{\epsilon}_{i=1,2}$) of the obtained errors values. We show the result of the first execution of each model, where a pixel in each segmentation map takes the color mean value in the mixture component (region) to which the pixel is assigned. Table I shows the values of $\bar{\epsilon}_1$ and $\bar{\epsilon}_2$ obtained for each model using the whole dataset. Two conclusions can be drawn from the results. First, the MoGG (resp. MoGG+FS) outperforms the MoG (respectively, MoG+FS) in both performance criteria. We noticed in some images that over-segmentation of regions can be caused by self-shadowing or/and nonuniform illumination, where color homogeneity inside such regions is affected. The flexibility of the MoGG allows to include those areas inside the main regions, while the MoG tends to create small regions to fit them. Second, using FS in both MoGG and MoG models yields better performance than without using FS. Finally, using a mixture of arbitrary number of components for irrelevant features enhances the segmentation accuracy compared to using only a fixed number of components (i.e., $K = 1$). The added performance using the proposed model, with respect to $\epsilon_1$, is approximately 51% against MoG, 34% against MoG+FS, 48% against MoGG, and 24% against MoGG+FS with fixed $K = 1$. With respect to $\epsilon_2$, the added performance is 52% against MoG, 34% against MoG+FS, 45% against MoGG, and 20% against MoGG+FS with fixed $K = 1$. These results clearly demonstrate the advantage of combining FS and GGD mixture modeling for reducing over/under-segmentation.

[1]A segment is a connected set of 05 pixels or more, equally labeled.

TABLE I
Values of the Errors $\epsilon_1$ and $\epsilon_2$ for the Compared Image Segmentation Models

| $[\bar{\epsilon}_1, \bar{\epsilon}_2]$ | | | | |
|---|---|---|---|---|
| MoG | MoG+FS | MoGG | MoGG+FS (K=1) | MoGG+FS (K=*) |
| [0.21, 23] | [0.15, 16] | [0.19, 20] | [0.13, 13] | [**0.10**, **11**] |

### B. Application to Video Segmentation

In real-world videos, irrelevant data (i.e., outliers) can be induced by noise, shadowing, or even insignificant objects entering momentarily in the scene, which decrease the segmentation accuracy. To mitigate the effect of these outliers, we propose to use the MoGG+FS model for video segmentation. Let $M$ be the real number of spatiotemporal regions that compose the moving objects and the background, and $T$ be the number of frames in the video sequence. An efficient way to remove the outlying data is by selecting the best (relevant) frames that exhibit a good separation among the moving objects in the video. In other words, the frames where the objects are not well separated (because of object occlusions, noise, shadows, etc.) will see their contribution decreased in the segmentation. By including the frame relevance, the formulation of the video segmentation becomes as follows:

$$p(\vec{x}_{it}|\vec{\Theta}) = \sum_{j=1}^{M} p_j \prod_{t=1}^{T} \left( \rho_{t1} p(\vec{x}_{it}|\theta_j) + \rho_{t2} \sum_{k=1}^{K} \pi_k p(\vec{x}_{it}|\varphi_k) \right) \quad (7)$$

where $t$ designates the frame number, $\rho_{t1}$ and $\rho_{t2}$ define the relevancy of the frame $t$, and $\vec{x}_{it}$ is the $i$th pixel of the same frame. From (7), the value of a pixel $\vec{x}_{it}$ on the frame $t$ can be generated either by a component of the mixture of the real objects or by the common mixture (i.e., modeling the features distribution in the irrelevant frames). The estimation of the parameters of the model is performed in the same way as for image segmentation, by using the EM algorithm.

Fig. 2 shows four examples of video segmentation using the tested models. From each video, we show a frame drawn randomly from the sequence. Table II shows the errors $\epsilon_1$ and

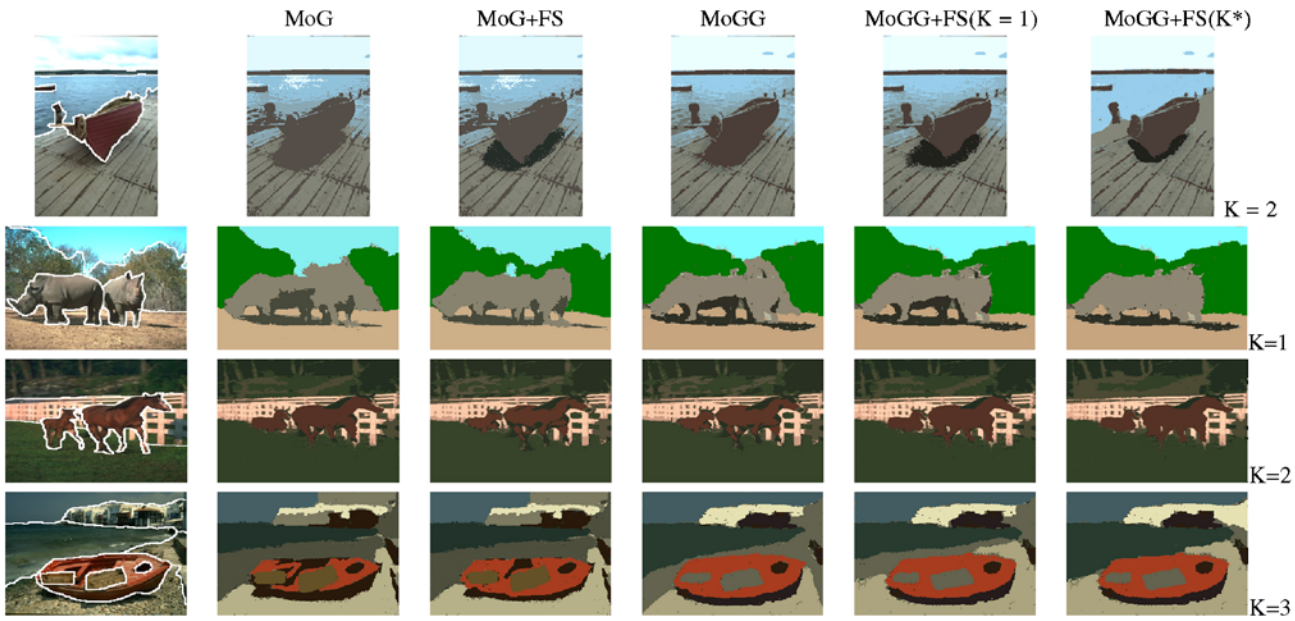


Fig. 1. Examples of real-world image segmentation. In each row, the first column represents the GT. In the right most of each row, we show the optimal value of $K$ found for the corresponding example.


Fig. 2. Examples of video segmentation: from top to bottom, we show segmentation of frames chosen randomly from each video. In the right most of each row, we show the optimal value of $K$ found for the corresponding example.

TABLE II

VALUES OF THE ERRORS $\epsilon_1$ AND $\epsilon_2$ FOR THE COMPARED VIDEO SEGMENTATION MODELS

| Video | Size | $[\bar{\epsilon}_1,\bar{\epsilon}_2]$ | | | | |
|---|---|---|---|---|---|---|
| | | MoG | MoG+FS | MoGG | MoGG+FS(K=1) | MoGG+FS(K=*) |
| *Akiyo* | 300 frames | [0.23, 22.5] | [0.16, 17.8] | [0.22, 20.5] | [0.12, 13.2] | [0.12, 13.2] |
| *Suzie* | 150 frames | [0.25, 27.4] | [0.18, 19.1] | [0.24, 23.7] | [0.15, 17.5] | [0.12, 11.3] |
| *Grandma* | 870 frames | [0.19, 20.5] | [0.15, 14.8] | [0.17, 19.3] | [0.13, 14.5] | [0.12, 13.9] |
| *Irene* | 300 frames | [0.27, 28.2] | [0.19, 20.8] | [0.23, 26.6] | [0.16, 17.3] | [0.13, 14.4] |

$\epsilon_2$ for the video sequences. We can see clearly the improvement brought by the proposed model against the compared ones. These results are confirmed visually in the segmentations shown in Fig. 2, where the quality of object segmentation is clearly improved using the proposed approach.

## IV. Computational Complexity

In what follows, $N$ represents the number of pixels in an image or a video. The mixture parameters are initialized using fuzzy $c$-means, which is of complexity $O(NMd)$. Before each EM iteration, we compute $NMd$ density probabilities for the main mixture and $NKd$ for the common mixture models, respectively. Since the cost of calculating all the $p(j|\vec{x}_i)$'s is $O(Nd)$, and generally $K < M$, the computational complexity of the E-step is $O(NMd)$. For the M-step, (9)–(11) can be updated using $O(NMKd)$ arithmetic operations, while (12)–(14) require $O(MNd)$ operations. The computation of the Hessian matrix of the parameters $\theta_{jl}$ and $\varphi_{kl}$, involved in Fisher scoring, requires $O(NMKd)$ operations. It follows that the M-step time complexity is $O(NMKd)$. We note that for most of the tests, the EM algorithm takes a maximum of 40 iterations to converge. Finally, the assignment of pixel labels takes roughly $O(N)$ operations.

## V. Conclusion

We proposed a new model which combines the GGD formulation and feature selection in robust mixture modeling for segmentation. Our results demonstrate the usefulness and the effectiveness of the proposed model in reducing over/under-segmentation due to heavy-tailed and high-dimensional data. Future work will investigate adding spatial information in order to yield more semantically meaningful segmentation. Moreover, segmentation can be enhanced by including hierarchical schemes, where multiple outputs can be generated for object recognition, for example.

## Acknowledgments

We thank NSERC for their financial support and the reviewers for their helpful comments.

## Appendix

The EM steps are given by

**E-step**

$$t_{ij} = p(j|\vec{x}_i) = \frac{p_j \prod_{l=1}^{d} \left[\beta_j(x_{il})\right]}{\sum_{j=1}^{M} p_j \prod_{l=1}^{d} \left[\beta_j(x_{il})\right]} \tag{8}$$

where $\beta_j(x_{il}) = \rho_{l1}\, p(x_{il}|\theta_{jl}) + \rho_{l2}\, p(x_{il}|\vec{\varphi}_l)$, and $p(x_{il}|\vec{\varphi}_l) = \sum_{k=1}^{K} \pi_{kl}\, p(x_{il}|\varphi_{kl})$. We estimate the parameters $\Theta$ in the M-step as follows:

**M-step**

$$\hat{p}_j = \frac{\max\left(\sum_{i=1}^{N} t_{ij} - \frac{3d}{2}, 0\right)}{\sum_{j=1}^{M} \max\left(\sum_{i=1}^{N} \hat{z}_{ij} - \frac{3d}{2}, 0\right)} \tag{9}$$

$$\frac{1}{\hat{\rho}_{l1}} = 1 + \frac{\max\left(\sum_{i=1}^{N}\sum_{j=1}^{M} t_{ij} \frac{\rho_{l2}\, p(x_{il}|\vec{\varphi}_l)}{\beta_j(x_{il})} - \frac{3K}{2}, 0\right)}{\max\left(\sum_{i=1}^{N}\sum_{j=1}^{M} t_{ij} \frac{\rho_{l1}\, p(x_{il}|\theta_{jl})}{\beta_j(x_{il})} - \frac{3M}{2}, 0\right)} \tag{10}$$

$$\hat{\pi}_{kl} = \frac{\max\left(\sum_{i=1}^{N}\sum_{j=1}^{M} t_{ij} \frac{\rho_{l2}\, \pi_{kl}\, p(x_{il}|\varphi_{kl})}{\beta_j(x_{il})} - \frac{3}{2}, 0\right)}{\sum_{k=1}^{K} \max\left(\sum_{i=1}^{N}\sum_{j=1}^{M} t_{ij} \frac{\rho_{l2}\, \pi_{kl}\, p(x_{il}|\varphi_{kl})}{\beta_j(x_{il})} - \frac{3}{2}, 0\right)} \tag{11}$$

$$\hat{\mu}_{jl}^{\theta} = \frac{\sum_{i=1}^{N} t_{ij} \frac{\rho_{l1}\, p(x_{il}|\theta_{jl}) |x_{il} - \mu_{jl}^{\theta}|^{\lambda_{jl}^{\theta}-2}}{\beta_j(x_{il})} x_{il}}{\sum_{i=1}^{N} t_{ij} \frac{\rho_{l1}\, p(x_{il}|\theta_{jl}) |x_{il} - \mu_{jl}^{\theta}|^{\lambda_{jl}^{\theta}-2}}{\beta_j(x_{il})}} \tag{12}$$

$$\hat{\sigma}_{jl}^{\theta} = \sqrt[\lambda_{jl}^{\theta}]{\frac{\sum_{i=1}^{N} t_{ij} \frac{\rho_{l1}\, p(x_{il}|\theta_{jl}) \lambda_{jl}^{\theta} A(\lambda_{jl}^{\theta}) |x_{il} - \mu_{jl}^{\theta}|^{\lambda_{jl}^{\theta}}}{\beta_j(x_{il})}}{\sum_{i=1}^{N} t_{ij} \frac{\rho_{l1}\, p(x_{il}|\theta_{jl})}{\beta_j(x_{il})}}}. \tag{13}$$

Finally, we estimate the parameters $\vec{\lambda}_j^{\theta}$ and $\vec{\lambda}_k^{\varphi}$, with $j = 1, ..., M$ and $k = 1, ..., K$ using the Newton–Raphson method

$$\hat{\lambda}_{\circ l}^{\star} \simeq \hat{\lambda}_{\circ l}^{\star} - \left[\frac{\partial^2 MessL(M, K)}{\partial {\lambda_{\circ l}^{\star}}^2}\right]^{-1} \left[\frac{\partial MessL(M, K)}{\partial \lambda_{\circ l}^{\star}}\right] \tag{14}$$

where the symbol $\circ$ refers to either the index $j$ or $k$, and the symbol $\star$ for either $\theta$ or $\varphi$.

## References

[1] M. S. Allili and D. Ziou, "Globally adaptive region information for color-texture image segmentation," *Pattern Recognit. Lett.*, vol. 28, no. 15, pp. 1946–1956, Nov. 2007.

[2] M. S. Allili and D. Ziou, "Object of interest segmentation and tracking using feature selection and active contours," in *Proc. IEEE Conf. CVPR*, Jun. 2007, pp. 1–8.

[3] M. S. Allili, N. Bouguila, and D. Ziou, "Finite general Gaussian mixture modeling and application to image and video foreground segmentation," *J. Electron. Imaging*, vol. 17, pp. 013005.1–013005.13, Jan.–Mar. 2008.

[4] N. Bouguila and D. Ziou, "High-dimensional unsupervised selection and estimation of finite generalized Dirichlet mixture model based on the minimum message length," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1716–1731, Oct. 2007.

[5] S. Boutemedjet, N. Bouguila, and D. Ziou, "Feature selection for non-Gaussian mixture models," in *Proc. IEEE Workshop MLSP*, Aug. 2007, pp. 69–74.

[6] 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.

[7] M. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.

[8] H. Gao, W.-C. Siu, and C.-H. Hou, "Improved techniques for automatic image segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 12, pp. 1273–1280, Dec. 2001.

[9] M. H. C. Law, M. A. T. Figueiredo, and A. K. Jain, "Simultaneous feature selection and clustering using mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1154–1166, Sep. 2004.

[10] 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. IEEE ICCV*, Jul. 2001, pp. 416–423.

[11] A. Y. Ng, "On feature selection: Learning with exponentially many irrelevant features as training examples," in *Proc. 15th ICML*, Jul. 1998, pp. 404–412.

[12] V. Roth and T. Lange, "Adaptive feature selection in image segmentation," in *Proc. 26th DAGM Symp. Pattern Recognit.*, Aug. 2004, pp. 9–17.

[13] P. Salembier, "Overview of the MPEG-7 standard and of future challenges for visual information analysis," *EURASIP J. Appl. Signal Process.*, vol. 4, pp. 343–353, Apr. 2002.

[14] E. P. Simoncelli and H. Hughes, "Natural image statistics and neural representation," *Ann. Rev. Neurosci.*, vol. 24, pp. 1193–1216, May 2001.

[15] C. Wallace, "Statistical and inductive inference by minimum message length," in *Information Science and Statistics*. Berlin, Germamy: Springer, 2005.

[16] M. Watanabe and K. Yamaguchi, *The EM Algorithm and Related Statistical Models*. Evanston, IL: Routledge, 2004, ch. 7.