



# Spatially Constrained Mixture Model with Feature Selection for Image and Video Segmentation

Ines Channoufi<sup>1,2(✉)</sup>, Sami Bourouis<sup>1,3</sup>, Nizar Bouguila<sup>4</sup>,  
and Kamel Hamrouni<sup>1</sup>

<sup>1</sup> Ecole Nationale d'Ingénieurs de Tunis, LR-SITI Laboratoire Signal Image et Technologies de l'Information, Université de Tunis El Manar, 1002 Tunis, Tunisia  
[kamel.hamrouni@enit.rnu.tn](mailto:kamel.hamrouni@enit.rnu.tn)

<sup>2</sup> ESPRIT School of Engineering, Tunis, Tunisia  
[ines.channoufi@esprit.tn](mailto:ines.channoufi@esprit.tn)

<sup>3</sup> Taif University, Taif, Saudi Arabia  
[s.bourouis@tu.edu.sa](mailto:s.bourouis@tu.edu.sa)

<sup>4</sup> The Concordia Institute for Information Systems Engineering (CIISE),  
Concordia University, Montreal, QC H3G 1T7, Canada  
[nizar.bouguila@concordia.ca](mailto:nizar.bouguila@concordia.ca)

**Abstract.** In this paper we propose to improve image and video sequences segmentation through the integration of feature selection process into an unsupervised learning approach based on a finite mixture of bounded generalized Gaussian distributions (BGGMD). The proposed algorithm is less sensitive to over-segmentation, more flexible to data modeling and leading to better characterization and localization of object of interest in high-dimensional spaces since it is able to automatically reject irrelevant visual features. In order to determine adequately and automatically the number of regions in each image or frame, spatial information is incorporated as a prior information between neighboring pixels. Experimental results which are performed on a several real world images and videos demonstrate the effectiveness of the proposed framework with respect to other conventional Gaussian-based mixture models.

**Keywords:** Image/video segmentation · Visual features selection  
Mixture of bounded generalized Gaussian distributions  
Spatial information · Minimum message length

## 1 Introduction

Accurate images and video sequences segmentation is always a crucial step in many applications such as object detection and tracking. To deal with this problem, we proceed constantly by involving more and more information (visual features) in the whole segmentation process. Visual features (color, texture, shape, etc.) can be divided into two sub-sets: relevant (informative) and irrelevant (uninformative) features. The irrelevant features can be just noise which make

the identification of the real regions too difficult. In addition, the presence of self-shadow or nonuniform illumination can form new false clusters and all these issues lead to the so-called over-segmentation. Therefore, it is important to avoid irrelevant features during the segmentation process to better recognize objects in a given image or frame. In the state of the art, feature extraction and selection steps have been widely used in the field of computer vision and image processing. For example, an adaptive Mean-Shift Blob tracking technique and graph cuts theory are used in conjunction with a step of shape features extraction in [14] to better characterize the ROI. In [25], an active feature selection approach based on Fisher information criterion is developed. It employs an online boosting feature selection mechanism by optimizing the Fisher information criterion which is validated with success for real-time object tracking. Later, this work has been extended in [24] by integrating a prior information to handle efficiently visual drift. Another effective feature selection method for color image segmentation is developed in [9] which selects a group of mixed color features from different color spaces based on the principle of the least entropy of pixels frequency histogram distribution. Finite mixture models have been also widely used for image and video segmentation [1, 3, 5, 15–17, 19]. For example in [19], Gaussian mixture model (GMM) has been used with a feature extraction method for conditional random fields (CRF) to extract objects and background from images. In [1], authors investigate the benefits of the generalized Gaussian mixture model as a flexible model to segment images. In [20], authors introduced a method for change detection in videos of crowded scenes by combining an extended mixture of Gaussian Switch Model and a Flux Tensor pre-segmentation. Mixture models have been also used in [23], in which an ontology-based semantic image segmentation approach that jointly models image segmentation and object detection is proposed. It is important to note that conventional Gaussian-based models have an issue related to their distributions which are unbounded. In fact, in many applications, the observed data are digitalized and have bounded support and this statement can be exploited to select the appropriate model shape. Thus, the bounded generalized Gaussian mixture model has been developed in [11, 18] as a generalized model, which includes Gaussian model (GMM), Laplacian model (LMM) and generalized Gaussian model (GGMM) as special cases. This new model has the advantage to fit different shapes of the observed data such as bounded support and non-Gaussian data. Moreover, it is possible to model the observed data in each component of the model with different bounded support regions. The bounded Generalized Gaussian mixture model (BGGMM) have been well used and outperforms other classical Gaussian-based mixture models especially in the case of speech modelling [11] and image denoising [6]. Motivated by the aforementioned observations, we introduce in this work a feature selection approach for the finite bounded generalized Gaussian mixture model (BGGMM + FS) for image and video segmentation, which includes the GMM, LMM, GGMM, and BGMM, as special cases. On the other hand, we associate a relevance weight for each feature which measures the degree of its dependence on class labels. In order to select the optimal number of components, we derive

also a Minimum Message Length criterion [22] specific to the proposed mixture model and we integrate the spatial information as a prior information between neighboring pixels through the well known EM-algorithm. Finally, we validate our framework for both real world images and video sequences segmentation. The remainder of this paper is organized as follows. Section 2 outlines the proposed unsupervised feature selection model based on a bounded generalized Gaussian mixture model with feature selection mechanism and spatial information. Then, in Sect. 3, obtained results and a comparative study are presented. Finally, we end with a conclusion and some future works in Sect. 4.

## 2 Unsupervised Feature Selection Model

Let  $\mathcal{X}$  be an image composed by a  $D$ -dimensional vectors  $\mathcal{X} = \{X_1, \dots, X_N\}$ . This data set can be described using a mixture of  $K$  components. Each component is identified by a set of parameters denoted by  $\theta_j$ . The mixture of this distribution can be written as:

$$p(x_i|\Theta) = \sum_{j=1}^K \pi_j \prod_{l=1}^D f(x_l|\theta_{jl}) \quad (1)$$

where  $\pi_j$  are the mixing proportions,  $f(x_l|\theta_{jl})$  represents the probability density function of the  $l$ -th feature in component  $j$ ,  $\Theta = \{\theta_1, \dots, \theta_K, \pi_1, \dots, \pi_K\}$  is the complete set of parameters used to characterize the mixture model and  $K \geq 1$  is the number of components in the mixture model [13]. Each  $f(x_l|\theta_{jl})$  is defined as:

$$f(x_l|\theta_{jl}) = \frac{f_{ggd}(x_l|\theta_{jl})H(x_l|\Omega_{jl})}{\int_{\partial_{jl}} f_{ggd}(u|\theta_{jl})du} \quad (2)$$

Where  $H(x_l|\Omega_{jl})$  is the indicator function which defines for each component  $\Omega_{jl}$  the bounded support region  $\partial_{jl}$  in  $\mathfrak{R}$  ( with  $H(x_l|\Omega_{jl}) = 1$  if  $x_l \in \partial_{jl}$  ) and  $f_{ggd}$  represents the generalized Gaussian distribution [1, 17]. In order to provide an accurate estimation of the number of clusters, we consider for each pixel its immediate neighbors which we call peers [2]. Let  $\mathcal{X}$  and the set of peers  $\hat{\mathcal{X}} = \{\hat{X}_1, \dots, \hat{X}_N\}$  be our observed data. The set of membership indicators for all pixels  $\mathcal{Z} = \{\mathbf{Z}_1, \dots, \mathbf{Z}_N\}$  correspond to the unobserved data, where  $\mathbf{Z}_i = \{Z_{i1}, \dots, Z_{iK}\}$  is the missing membership indicator and  $Z_{ij}$  is equal to one if  $X_i$  belongs to the same cluster  $j$  as  $\hat{X}_i$  and zero otherwise. In order to estimate the model's parameters we adopt Maximum Likelihood approach within the expectation maximization (EM) algorithm [13]. The complete data likelihood function is given as:

$$p(\mathcal{X}, \hat{\mathcal{X}}, \mathcal{Z}|\Theta) = \prod_{i=1}^N \prod_{j=1}^K \left[ \pi_j f(X_i|\theta_j) \pi_j f(\hat{X}_i|\theta_j) \right]^{Z_{ij}} \quad (3)$$

Let  $\phi_l$  a binary variable, where  $\phi_l = 0$  when the  $l$ th feature is irrelevant. Thus, the distribution of each  $x_{il}$  is given as [10]:

$$p(x_{il}|\theta_{jl}^*, \varphi_l^*, \phi_l) \simeq (p(x_{il}|\theta_{jl}))^{\phi_l} (p(x_{il}|\varphi_l))^{1-\phi_l} \quad (4)$$

The unknown distribution of the  $l$ th feature is denoted by the superscript star,  $p(x_{il}|\theta_{jl})$  and  $p(x_{il}|\varphi_l)$  are univariate bounded generalized Gaussian distributions. Since the mixture of BGGDs is almost able to approximate any arbitrary distribution of irrelevant feature, we consider the irrelevant component  $p(x_{il}|\varphi_l)$  as a common mixture of BGGDs independent of the region labels. Therefore, we can deal with the case when a feature is defined by only overlapped component which make the distinction between real regions of an image more difficult to achieve. In this common mixture, we consider  $K$  as the number of components with the parameters  $\varphi_{1l}, \dots, \varphi_{Kl}$  for each feature. The outline of feature saliency technique can be described following these steps: (1)  $\phi_l$ 's are considered as missing variables, (2) we consider  $\rho_{l1} = p(\phi_l = 1)$ , the probability that the  $l$ th feature is relevant and  $\rho_{l2} = p(\phi_l = 0)$  the probability that the  $l$ th feature is irrelevant (such that  $\rho_{l1} + \rho_{l2} = 1$ ). Consequently, we derive the following model for the segmentation with feature selection:

$$p(\mathbf{x}_i|\Theta) = \sum_{j=1}^M p_j \prod_{l=1}^d \left( \rho_{l1} p(x_{il}|\theta_{jl}) + \rho_{l2} \sum_{k=1}^K \pi_{kl} p(x_{il}|\varphi_{kl}) \right) \quad (5)$$

The prior probability  $\pi_{kl}$  denotes that  $x_{il}$  is generated by the  $k$ th component of the common mixture, given that the  $l$ th feature is irrelevant ( $\phi_l = 0$ ) where  $\sum_{k=1}^K \pi_{kl} = 1$ . Thus, the segmentation approach by integration a feature selection mechanism can be performed by optimizing an objective functional that estimates a set of all model's parameters which we will note  $\Theta = (p, \theta_{jl}, \varphi_{kl}, \pi_l)$  where  $p = (p_1, \dots, p_M)$ ,  $\theta_{jl} = (\lambda_{jl}, \mu_{jl}, \sigma_{jl})$  and  $\pi_l = (\pi_{l1}, \dots, \pi_{lK})$ . In order to perform feature selection, we use the minimum message length (MML) criterion as proposed in [22]. Within the following constraints:  $0 < p_j \leq 1$ ,  $0 < \rho_{l1} \leq 1$ ,  $0 < \pi_{kl} \leq 1$  and  $\sum_{j=1}^M p_j = 1$ ,  $\sum_{k=1}^K \pi_{kl} = 1$ , the final MML objective of a data set  $X$  is given as:

$$\begin{aligned} MML = & -\log p(X|\Theta) + \frac{c}{2} \log N + \frac{3d}{2} \sum_{j=1}^M \log p_j + \frac{3}{2} \sum_{l=1}^d \sum_{k=1}^K \log \pi_{kl} \\ & + \frac{c}{2} (1 + \log \frac{1}{12}) + \frac{3M}{2} \sum_{l=1}^d \log \rho_{l1} + \frac{3K}{2} \sum_{l=1}^d \log \rho_{l2} \end{aligned} \quad (6)$$

### 3 Experiments and Discussions

In this work, we conducted a series of experiments on many examples of real world image and video segmentation. We compared the performance of the proposed segmentation model for feature selection in bounded generalized Gaussian mixtures (BGGMM + FS) with the Gaussian mixture (GMM), a Gaussian mixture with feature selection (GMM + FS), the generalized Gaussian mixture (GGMM), a generalized Gaussian mixture with feature selection (GGMM + FS), and the bounded generalized Gaussian mixture (BGGMM).

### 3.1 Experiment 1: Color-Texture Image Segmentation

In this application, each pixel is represented by a feature vector  $\mathbf{x}(i, j)$  that contains color and texture information. 19 visual features are calculated and used to describe images as follows: 16 texture features are calculated from the color correlogram (CC) [3, 8] and 3 color features are obtained using the RGB color space. Performances are determined w.r.t the ground truth. To this end, we perform tests on the database provided by the Berkeley Benchmark [12]. Quantitative measures are done on the basis of the accuracy and the boundary displacement error (BDE) [12] as depicted in Table 1. It should be emphasized that better results are obtained when the accuracy value is high and the BDE is low.

*Accuracy*: it measures the proportion of correctly labelled pixels over all available pixels and it is calculated as:  $Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$ .

*Boundary displacement error (BDE)*: it measures the average displacement error of one boundary pixels and its closest boundary pixels in the other segmentation [7].

**Table 1.** Average metrics for all images in the Berkeley benchmark dataset produced by different algorithms

<i>Algorithm</i>	GMM	GMM + FS	GGMM	GGMM + FS	BGGMM	BGGMM + FS
Accuracy	0.9210	0.9326	0.9387	0.9503	0.9498	0.9668
BDE	0.324	0.276	0.291	0.218	0.223	0.186

A summary of the segmentation accuracy are shown in Table 1. According to these results, we can conclude that both BGGMM + FS, BGGMM and GGMM + FS are able to provide high performances. More precisely, the BGGMM + FS outperforms the other models. Indeed, the accuracy value is about 96.68% for BGGMM + FS, against 94.98% for BGGMM and 92.10% for GMM. It is noteworthy that the consideration of the spatial information within the proposed model allows the fusing of small regions into the main ones and permits to have more accurate number of clusters and segmentation results. In addition, the feature selection step has a positive impact and enhanced the segmentation quality since BGGMM + FS outperforms BGGMM (i.e. without feature selection).

### 3.2 Experiment 2: Video Segmentation

Video segmentation is an important task for a variety of applications, such as object retrieval and tracking which require an accurate segmentation approach. It is noteworthy that most of existing video are contaminated with irrelevant data such as noise, self shadowing, etc., which can decrease the accuracy of their segmentation. Let  $F$  be the number of frames in a video sequence composed by

R regions. In order to provide a good segmentation result, we have to discard irrelevant frames from the video sequence. Thus, the frames in which objects are not well separated will be automatically rejected. In this case, the proposed video segmentation model with feature selection is formulated as follow [21]:

$$p(\mathbf{x}_{if}|\Theta) = \sum_{j=1}^R p_j \prod_{f=1}^F \left( \rho_{f1} p(x_{if}|\theta_j) + \rho_{f2} \sum_{k=1}^K \pi_{kl} p(x_{if}|\varphi_k) \right) \quad (7)$$

In this manner, frames become the features among which selection is performed, where  $f$  is the frame number,  $\mathbf{x}_{if}$  is the value of the pixel  $i$  in the frame  $f$  and  $\rho_{f1}, \rho_{f2}$  designate the saliency of the frame  $f$ . As it has been described above, the EM algorithm is used to estimate the model parameters. Therefore, the distributions of the irrelevant frames will be forced to zero and thus not considered as informative data for the segmentation task. Figure 1 shows the segmentation results of four different video sequences (Akiyo, Gandma, Claire, Mother-Daughter) using the tested models. A quantitative comparative study between different methods is given also in Table 2. From these measurements, we notice that the results are improved using our model BGGMM + FS. Also, the integration of the spatial information in the proposed model helps to overcome the problem of over segmentation and to find the accurate number of regions as illustrated in the two last rows of the Table 2. All these results confirm clearly the enhancement brought by the proposed model against the compared ones.



**Fig. 1.** Video Segmentation Results; Column1: Original frame, Column2: GMM, Column3: GMM + FS, Column4: GGMM, Column5: GGMM + FS, Column6: BGGMM, Column7: BGGMM + FS

**Table 2.** Average metrics and estimated number of regions for different video segmentation models. Values between parentheses denote the real number of regions for each video

	GMM	GMM + FS	GGMM	GGMM + FS	BGGMM	BGGMM + FS
Accuracy	0.941	0.944	0.949	0.956	0.962	0.966
BDE	0.354	0.320	0.317	0.269	0.246	0.207
$\hat{M}$ [Akiyo (10)]	14	12	14	13	13	12
$\hat{M}$ [Claire (7)]	9	8	9	8	8	7

## 4 Conclusion

In this paper, we have presented a spatially constrained unsupervised learning statistical model with feature selection approach for image and video segmentation. This model is based on combining a bounded generalized Gaussian mixture model with spatial information and a feature selection operation in order to enhance the segmentation performances of both color-texture images and videos. The proposed approach has the ability to automatically detect the appropriate number of clusters thanks to the spatial information and the MML criterion. Obtained results show the merits of our proposed framework which outperforms conventional gaussian-based models (with and without feature selection). As a future work, we plan to use an enhanced extension of the EM algorithm which is the ECM method [4] to avoid some problems related to EM.

## References

1. Allili, M.S., Ziou, D., Bouguila, N., Boutemedjet, S.: Image and video segmentation by combining unsupervised generalized gaussian mixture modeling and feature selection. *IEEE Trans. Circuits Syst. Video Technol.* **20**(10), 1373–1377 (2010)
2. Bouguila, N.: Spatial color image databases summarization. In: *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*, vol. 1, pp. I–953. *IEEE* (2007)
3. Bouguila, N., ElGuebaly, W.: Integrating spatial and color information in images using a statistical framework. *Expert Syst. Appl.* **37**(2), 1542–1549 (2010)
4. Bouguila, N., Ziou, D.: On fitting finite dirichlet mixture using ECM and MML. In: Singh, S., Singh, M., Apte, C., Perner, P. (eds.) *ICAPR 2005*. LNCS, vol. 3686, pp. 172–182. Springer, Heidelberg (2005). [https://doi.org/10.1007/11551188\\_19](https://doi.org/10.1007/11551188_19)
5. Channoufi, I., Bourouis, S., Bouguila, N., Hamrouni, K.: Color image segmentation with bounded generalized gaussian mixture model and feature selection. In: *4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP 2018)* (2018)
6. Channoufi, I., Bourouis, S., Bouguila, N., Hamrouni, K.: Image and video denoising by combining unsupervised bounded generalized gaussian mixture modeling and spatial information. *Multimed. Tools Appl.* (2018). <https://doi.org/10.1007/s11042-018-5808-9>

7. Freixenet, J., Muñoz, X., Raba, D., Martí, J., Cufí, X.: Yet another survey on image segmentation: region and boundary information integration. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) *ECCV 2002*. LNCS, vol. 2352, pp. 408–422. Springer, Heidelberg (2002). [https://doi.org/10.1007/3-540-47977-5\\_27](https://doi.org/10.1007/3-540-47977-5_27)
8. Huang, J., Kumar, S.R., Mitra, M., Zhu, W.J., Zabih, R.: Image indexing using color correlograms. In: *1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997*. Proceedings, pp. 762–768. IEEE (1997)
9. Junfeng, L., Jinwen, M.: Effective selection of mixed color features for image segmentation. In: *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, pp. 794–798. IEEE (2016)
10. Law, M.H., Figueiredo, M.A., Jain, A.K.: Simultaneous feature selection and clustering using mixture models. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(9), 1154–1166 (2004)
11. Lindblom, J., Samuelsson, J.: Bounded support gaussian mixture modeling of speech spectra. *IEEE Trans. Speech Audio Process.* **11**(1), 88–99 (2003)
12. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Proceedings Eighth IEEE International Conference on Computer Vision, 2001. ICCV 2001*, vol. 2, pp. 416–423. IEEE (2001)
13. McLachlan, G., Peel, D.: *Finite Mixture Models*. Wiley, Hoboken (2004)
14. Mohcine, B., Benayad, N.: Object detection and segmentation using adaptive meanshift blob tracking algorithm and graph cuts theory. In: Choras, R.S. (ed.) *Image Processing and Communications Challenges 5. Advances in Intelligent Systems and Computing*, vol. 233, pp. 143–151. Springer, Heidelberg (2014). [https://doi.org/10.1007/978-3-319-01622-1\\_17](https://doi.org/10.1007/978-3-319-01622-1_17)
15. Nacereddine, N., Tabbone, S., Ziou, D., Hamami, L.: Asymmetric generalized gaussian mixture models and EM algorithm for image segmentation. In: *2010 20th International Conference on Pattern Recognition (ICPR)*, pp. 4557–4560 (2010)
16. Najar, F., Bourouis, S., Bouguila, N., Belghith, S.: A fixed-point estimation algorithm for learning the multivariate GGMM: application to human action recognition. In: *Accepted in the 31st IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2018)* (2018)
17. Najar, F., Bourouis, S., Bouguila, N., Belguith, S.: A comparison between different gaussian-based mixture models. In: *14th IEEE International Conference on Computer Systems and Applications, Tunisia*. IEEE (2017)
18. Nguyen, T.M., Wu, Q.J., Zhang, H.: Bounded generalized gaussian mixture model. *Pattern Recogn.* **47**(9), 3132–3142 (2014)
19. Qi, Y., Zhang, G., Li, Y.: Object segmentation based on gaussian mixture model and conditional random fields. In: *2016 IEEE International Conference on Information and Automation (ICIA)*, pp. 900–904. IEEE (2016)
20. Radolko, M., Farhadifard, F., von Lukas, U.: Change detection in crowded underwater scenes-via an extended gaussian switch model combined with a flux tensor pre-segmentation. In: *VISIGRAPP (4: VISAPP)*, pp. 405–415 (2017)
21. Song, X., Fan, G.: Selecting salient frames for spatiotemporal video modeling and segmentation. *IEEE Trans. Image Process.* **16**(12), 3035–3046 (2007)
22. Wallace, C.S.: *Statistical and Inductive Inference by Minimum Message Length*. Springer, New York (2005). <https://doi.org/10.1007/0-387-27656-4>



23. Zand, M., Doraisamy, S., Halin, A.A., Mustafa, M.R.: Ontology-based semantic image segmentation using mixture models and multiple CRFs. *IEEE Trans. Image Process.* **25**(7), 3233–3248 (2016)
24. Zhang, K., Zhang, L., Yang, M.H.: Real-time object tracking via online discriminative feature selection. *IEEE Trans. Image Process.* **22**(12), 4664–4677 (2013)
25. Zhang, K., Zhang, L., Yang, M.H., Hu, Q.: Robust object tracking via active feature selection. *IEEE Trans. Circuits Syst. Video Technol.* **23**(11), 1957–1967 (2013)