2018 IEEE Confs on Internet of Things, Green Computing and Communications, Cyber, Physical and Social Computing, Smart Data, Blockchain, Computer and Information Technology, Congress on Cybermatics

A Finite multi-dimensional generalized Gamma Mixture Model

Basim Alghabashi
CIISE, Concordia University
Montreal, QC, Canada H3G 1M8
b_algh@encs.concordia.ca

Nizar Bouguila
CIISE, Concordia University
Montreal, QC, Canada H3G 1M8
nizar.bouguila@concordia.ca

***Abstract*—Over the last two decades, statistical mixture models have been widely exploited to tackle the issue of data modeling. Examples of statistical mixture models' applications in data modeling include object recognition, speech recognition, information retrieval, and intrusion detection. In this paper, an unsupervised learning algorithm, based on a finite multi-dimensional generalized Gamma mixture model (GGMM) is presented for the purpose of positive vectors clustering. Maximum likelihood (ML) is a well-known method conducted via expectation maximization algorithm (EM) and used for estimating the parameters of the proposed model. Newton Raphson's optimization algorithm was also utilized to solve the problem (obstacle) of the non-existence of closed form. Experiments are conducted using both synthetic data and a real data set of images representing shapes to test the performance of the proposed model. Moreover, we compared the performance of the generalized Gamma mixture model with Gamma and Gaussian mixture models.**

***Index Terms*—Generalized Gamma, mixture models, maximum likelihood, clustering, expectation maximization, shape modeling**

I. Introduction

The availability of vast amounts of data sets that are dynamically growing in size in the last decade have increased the demand on analyzing and processing these data sets in order to extract useful and valuable knowledge [1]–[5]. Consequently, data analysts response to the above was to start designing statistical models using diverse data mining and machine learning techniques, such as, clustering, which received a lot of attention due to its importance [6], [7]. Machine learning techniques can be categorized into three families of approaches; supervised learning, unsupervised learning and semi-supervised learning [8]. Model-based approaches and in particular finite mixture models are widely used for unsupervised learning [9], [10] and they are considered to be a powerful and flexible tool for modeling univariate and multivariate data [11]. Finite mixture models have shown effective results in different fields such as biology, medicine, astronomy, engineering and marketing [9], [12], [13]. As a result, the number of related research works in the field has increased [14], [15]. The separation of a multi-dimensional heterogeneous large group of data into homogeneous clusters using mixture models is a crucial issue in image processing, computer vision and pattern recognition fields. Gaussian mixture model has been widely used and has shown appealing results, for applications such as texture classification and infrared face recognition [16], [17], etc. On the other hand, non-Gaussian mixture models have recently shown good results that outperform Gaussian distributions [18]–[27] and this is because of the fact that the shape of Gaussian model is not flexible which could affect the accuracy of the data modeling [28], [29]. For instance, Gamma mixture models for shape clustering have provided results with higher accuracy compared to Gaussian Models [28], [30], [31]. Also, It is noteworthy to highlight the merit GGMM has, which is the flexibility of its shape compared to Gamma and Gaussian models. Truly, the challenging task in unsupervised machine learning is to draw inferences from unlabeled observations to perform the clustering analysis by means of those statistical-based approaches, such as finite mixture models. This challenging task will become more complex when dealing with multi-dimensional data. For example, these multi-dimensional data occur when representing an image by a vector of multi-dimensional features [32].

To the best of our knowledge, generalized Gamma mixture models (GGMM) for multi-dimensional data clustering have never been considered in the past. The three-parameters generalized Gamma distribution was first introduced by Stacty [33] in 1962 which was achieved by introducing a positive exponential parameter in Gamma distribution. In addition, generalized Gamma distribution comes from the exponential family, which includes but not limited to Bernoulli, Dirichlet, Gamma, Laplace, and Weibull distributions. Maximum Likelihood (ML) is one of the most dominant and efficient techniques adopted to estimate the model parameters by finding the parameter values which maximize the likelihood [34] via expectation maximization (EM) [35]. In this paper, we present a statistical model based on finite mixtures of generalized Gamma distribution to cluster multi-dimensional positive vectors using a real-life application namely object shape modeling and clustering. Furthermore, the performance of the proposed algorithm has been investigated and compared against Gamma and Gaussian mixture models.

While the model we have proposed supports multi-dimensional data, there is a small number of related works that have proposed unsupervised learning of GGMM with different applications which consider only one dimension. For example, the authors in [36] have proposed GGMM with its application in statistical modeling of high-resolution synthetic-aperture radar (SAR) images. Ultrasonic tissue characterization [37]

978-1-5386-7975-3/18/$31.00 ©2018 IEEE
DOI 10.1109/Cybermatics_2018.2018.00158

and blind signal separation (BSS) [38] are two more interesting applications based on GGMM.

The rest of this paper is organized as follows. Section 2 describes the proposed generalized Gamma mixture model. In Section 3 the parameters estimation algorithm is proposed, then Section 4 evaluates the performance of the proposed model, and shows the experimental results using both synthetic and real data. Finally, the last Section is devoted to the conclusion and future works.

II. The proposed model

Suppose we have a data set $\mathcal{X} = \{\vec{X}_1, \vec{X}_2, ..., \vec{X}_n\}$ where each $\vec{X}_i = (X_{i1}, X_{i2}, ..., X_{iD})$ is a D-dimensional positive vector that follows a mixture of multi-dimensional GG distributions. Let M denotes the number of different components, and $i = 1, ..., N$. The mixture model can be written as follows:

$$p(\vec{X}_i|\Theta_M) = \sum_{j=1}^{M} p_j p(\vec{X}_i|\theta_j) \tag{1}$$

where $\Theta_M = \{\theta_1, \theta_2, ..., \theta_M, p_1, .., p_M\}$, and θ_j is the set of parameters of the distribution of class j, where $\theta_j = \{\vec{\alpha}_j, \vec{\beta}_j, \vec{\lambda}_j\}$, with $\vec{\alpha_j} = (\alpha_{j1}, \alpha_{j2}, ..., \alpha_{jD})$, $\vec{\beta}_j = (\beta_{j1}, \beta_{j2}, ..., \beta_{jD})$, and $\vec{\lambda}_j = (\lambda_{i1}, \lambda_{i2}, ..., \lambda_{iD})$. The parameter p_j is the mixing proportion such that $0 \leqslant p_j \leqslant 1$, and $\sum_{j=1}^{M} p_j = 1$.

$$p(\vec{X}_i|\theta_j) = \prod_{d=1}^{D} p(X_{id}|\theta_{jd}) = \prod_{d=1}^{D} \frac{\frac{\lambda_{jd}}{\alpha_{jd}^{\beta_{jd}}} X_{id}^{\beta_{jd}-1} \exp - \left(\frac{X_{id}}{\alpha_{jd}}\right)^{\lambda_{jd}}}{\Gamma(\frac{\beta_{jd}}{\lambda_{jd}})} \tag{2}$$

where $\theta_{jd} = (\alpha_{jd}, \beta_{jd}, \lambda_{jd})$ and $\Gamma(.)$ denotes the Gamma function, and $d = 1, \ldots, D$. The probability of generalized Gamma mixture distribution with M components is supposed to be known in our case, can be expressed by:

$$p(\mathcal{X}|\Theta) = \prod_{i=1}^{N} \sum_{j=1}^{M} p_j p(\vec{X}_i|\theta_j) \tag{3}$$

$Z_i = (Z_{i1}, Z_{i2}, ..., Z_{iM})$ are the introduced membership vectors, which specifies the cluster j each data point (observation) X_i belongs to. This means, an observation X_i will be equal to one if it belongs to cluster j and zero, otherwise. Then, the complete-data likelihood function is given by:

$$p(\mathcal{X}, Z|\Theta) = \prod_{i=1}^{N} \prod_{j=1}^{M} p_j p(\vec{X}_i|\theta_j)^{Z_{ij}} \tag{4}$$

III. Parameter estimation

The expectation maximization (EM) algorithm is a popular method used for estimating the latent mixture parameters that maximize the log-likelihood function. EM is an iterative method and consists of two steps; the expectation step and the maximization step. The complete log-likelihood function can be expressed by:

$$\log p(\mathcal{X}, Z|\Theta) = \sum_{i=1}^{N} \sum_{j=1}^{M} Z_{ij} \log p_j p(\vec{X}_i|\theta_j) \tag{5}$$

The main goal is to maximize the log-likelihood function by taking the gradient with the respect to each parameter p_j, α_{jd}, β_{jd}, and λ_{jd}. For the estimation of the parameter p_j, Lagrange multiplier is introduced to satisfy the constraints $0 \leqslant p_j \leqslant 1$, and $\sum_{j=1}^{M} p_j = 1$. As a result we obtain the log-likelihood function given by:

$$\log p(\mathcal{X}, Z|\Theta) = \sum_{i=1}^{N} \sum_{j=1}^{M} Z_{ij} \log p_j p(\vec{X}_i|\theta_j) + \Lambda \left(1 - \sum_{j=1}^{M} P_j\right) \tag{6}$$

By taking the gradient of the complete log-likelihood with respect to the parameters $p_j, \alpha_{jd}, \beta_{jd}$ and λ_{id}, we obtain the following (see appendix A and B):

$$p_j = \frac{\sum_{i=1}^{N} p(j|\vec{X}_i)}{N} \tag{7}$$

where p_j is the prior probability.

$$\hat{Z}_{ij} = p(j|\vec{X}_i) = \frac{p_j p(\vec{X}_i|\theta_j)}{\sum_{j=1}^{M} p_j p(\vec{X}_i|\theta_j)} \tag{8}$$

$$= \frac{p_j \prod_{d=1}^{D} \frac{\frac{\lambda_{jd}}{\alpha_{jd}^{\beta_{jd}}} X_{id}^{\beta_{jd}-1} \exp - \left(\frac{X_{id}}{\alpha_{jd}}\right)^{\lambda_{jd}}}{\Gamma(\frac{\beta_{jd}}{\lambda_{jd}})}}{\sum_{j=1}^{M} p_j \prod_{d=1}^{D} \frac{\frac{\lambda_{jd}}{\alpha_{jd}^{\beta_{jd}}} X_{id}^{\beta_{jd}-1} \exp - \left(\frac{X_{id}}{\alpha_{jd}}\right)^{\lambda_{jd}}}{\Gamma(\frac{\beta_{jd}}{\lambda_{jd}})}}$$

is the posterior probability function which helps assigning a vector $\vec{X}_i$ to a specific cluster j.

$$\alpha_{jd} = \left(\frac{\lambda_{jd}}{\beta_{jd}} \frac{\sum_{i=1}^{N} p(j|\vec{X}_i) X_{id}^{\lambda_{jd}}}{N p_j}\right)^{\frac{1}{\lambda_{jd}}} \tag{9}$$

$$\beta_{jd} = \lambda_{id} \Psi^{-1} \left(\lambda_{jd} \left(\frac{\sum_{i=1}^{N} p(j|\vec{X_{id}}) \log(X_{id})}{\sum_{i=1}^{N} p(j|\vec{X_{id}})} - \log(\alpha_{jd})\right)\right) \tag{10}$$

where $\Psi^{-1}(.)$ is the inverse digamma function.

By differentiating the log-likelihood w.r.t. the parameter λ, we obtain:

$$\frac{\partial p(\mathcal{X}|\Theta)}{\partial \lambda_{jd}} = \sum_{i=1}^{N} p(j|\vec{X}_i) \left[\frac{1}{\lambda_{jd}} - \left((\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}} \log(\frac{X_{id}}{\alpha_{jd}})\right) + \frac{\Psi(\frac{\beta_{jd}}{\lambda_{jd}})\beta_{jd}}{\lambda_{jd}^2}\right] \tag{11}$$

Looking at equation (11), we can see that we do not have a closed form, and this is because of the fact that the parameter λ_{jd} is coupled with the rest of the parameters. In this case, the Newton-Raphson method in [18] is used to solve this issue and estimate λ_{jd}. The Newton-Raphsons method can be expressed as follows:

$$\lambda_{jd}^{new} = \lambda_{jd}^{old} - \gamma \frac{\partial \log p(\mathcal{X}|\Theta)}{\partial \lambda_{jd}} \left(\frac{\partial^2 \log p(\mathcal{X}|\Theta)}{\partial^2 \lambda_{jd}} \right)^{-1} \quad (12)$$

Where γ is the constant step size added to Newton's method. After computing $\frac{\partial^2 \log p(\mathcal{X}|\Theta)}{\partial^2 \lambda_{jd}}$ we get:

$$\frac{\partial p(\mathcal{X}|\Theta)}{\partial \lambda_{jd}^2} = \sum_{i=1}^{N} P(j|\vec{X}_i) \Bigg[-\frac{1}{\lambda_{jd}^2} - (\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}} \log(\frac{X_{id}}{\alpha_{jd}}) \log(\frac{X_{id}}{\alpha_{jd}}) \quad (13)$$
$$- \left(\frac{\psi'(\frac{\beta_{jd}}{\lambda_{jd}})\beta^2}{\lambda_{jd}^4} \right) - \left(\frac{\beta_{jd} 2\psi(\frac{\beta_{jd}}{\lambda_{jd}})}{\lambda_{jd}^3} \right) \Bigg]$$

where $\psi'(.)$ is the trigamma function.

K-Means and Method of Moments (MoM) are the two techniques used to initialize the EM algorithm.

A. *Algorithm*

Algorithm 1 Multi-dimensional generalized Gamma Mixture Parameters Estimation

```
procedure
INPUT: D-Dimensional data set X = {X_1, X_2, ...., X_n}
OUTPUT: Θ
Initialization algorithm:
    applying K-Means on N D-dimensional vectors
    computing p_j
    obtaining α_j and β_j by applying the method of moments
    obtaining λ_j by using random positive values.
EM algorithm:
    while relative change in log-likelihood ≥ t_min do
        E-Step:
        compute the posterior probabilities P(j|X_i) usig Eq. (8)
        M-Step:
        update paramers p_j using Eq. (7)
        update the parameter α_j using Eq. (9)
        update the parameter β_j using Eq. (10)
        start Newton's Raphson Algorithm
        for all 1 ≤ j ≤ M do
        end for
        update the parameter λ_j using Eq. (12)
        end Newton's Raphson Algorithm
    end while
end procedure
```

TABLE I
REAL AND ESTIMATED PARAMETERS OF ONE-DIMENSIONAL DATA SET GENERATED FROM TWO GENERALIZED GAMMA DENSITIES

	real parameters	estimated parameters
Component 1	$\alpha_{11} = 1$	$\hat{\alpha}_{11} = 1.01$
	$\beta_{11} = 5$	$\hat{\beta}_{11} = 4.93$
	$\lambda_{11} = 5$	$\hat{\lambda}_{11} = 4.88$
component 2	$\alpha_{11} = 2$	$\hat{\alpha}_{22} = 1.96$
	$\beta_{21} = 6$	$\hat{\beta}_{22} = 5.51$
	$\lambda_{21} = 6$	$\hat{\lambda}_{22} = 5.795$

TABLE II
REAL AND ESTIMATED PARAMETERS OF ONE-DIMENSIONAL DATA SET GENERATED FROM THREE GENERALIZED GAMMA DENSITIES

	real parameters	estimated parameters
Component 1	$\alpha_{11} = 1$	$\hat{\alpha}_{11} = 1.01$
	$\beta_{11} = 5$	$\hat{\beta}_{11} = 4.93$
	$\lambda_{11} = 5$	$\hat{\lambda}_{11} = 4.88$
component 2	$\alpha_{11} = 2$	$\hat{\alpha}_{22} = 1.96$
	$\beta_{21} = 6$	$\hat{\beta}_{22} = 5.51$
	$\lambda_{21} = 6$	$\hat{\lambda}_{22} = 5.795$
component 3	$\alpha_{11} = 2$	$\hat{\alpha}_{22} = 1.96$
	$\beta_{21} = 6$	$\hat{\beta}_{22} = 5.51$
	$\lambda_{21} = 6$	$\hat{\lambda}_{22} = 5.795$

IV. EXPERIMENTAL RESULTS

A. *Synthetic data*

The first two experiments were conducted on one-dimensional ($D = 1$) data sets generated from two and three generalized Gamma densities ($M = 2, N = 100$) as shown in figs. 1 and 2 respectively. Furthermore, tables 1 and 2 display the real and estimated parameters.
The third experiment was conducted on a data set generated from generalized Gamma which includes one component and 2 dimensions ($M = 1, D = 2, N = 100$) as displayed in fig. 3, and the results of the estimated and real values of the parameters are shown in table III. Fig. 4 shows an example when considering 2-dimensional data set generated from a 2-components generalized Gamma mixture, where its real and estimated parameters are displayed in table IV. Finally, fig. 5 shows an example of generalized Gamma when generating 2-dimensional data from three densities, and table V shows the results of the estimated and real values. As we can see, the obtained results prove that the model is capable of performing well and able to estimate Accurately the model parameters.

B. *Real Data*

In this section, experiments are conducted to evaluate the performance of the proposed model using a well-known real data set to tackle the problem of object shape clustering [39]. To test the effectiveness of the GGMM model, the modeling capability of GGMM is compared against Gamma and Gaussian mixture models. Moreover, the provided results are the average over ten runs of the proposed algorithm. We have used 5 classes out of 7 from the data set MPEG-7

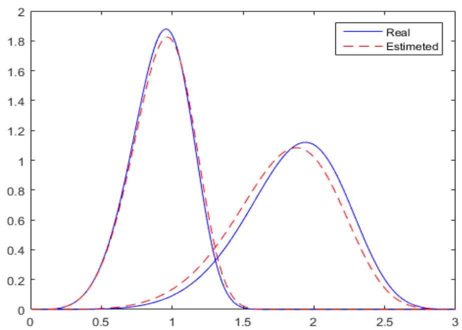

Fig. 1. Probability density functions of generalized Gamma distributions

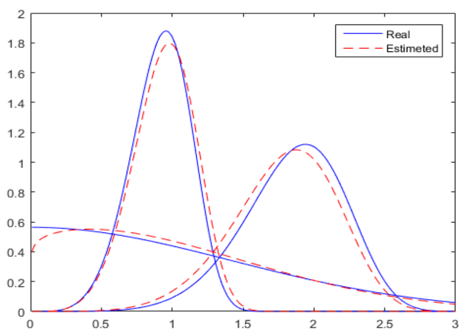

Fig. 2. Probability density functions of generalized Gamma distributions

CE Shape-1 Part-B [1], where each class contains 20 samples. For each image, a 36-dimensional positive vector features is extracted using Zernike Moments Magnitudes. Zernike Moments Magnitudes is an effective method used to extract features from shape images [28]. Examples of its applications include Invariant Image Recognition [40], and region-based shape modeling [41]. Fig. 6 shows some examples from the data set. Furthermore, the vector of features is computed based on the method proposed in [42], which imposes the constraints such that $n = 0, 1, 2, ..., \infty$, $|m| \leq n$, and $n - |m|$ is even, where n is the order of the Zernike moments and m is the Zernike moment repetition number. We performed shape clustering using generalized Gamma, Gamma and Gaussian mixture models on a multi-dimensional real data set. The confusion matrix of the proposed model is shown in table VI which is based on 10 runs. As we can see, 3 heart shape images are miss-classified which represents 97% clustering accuracy. On the other hand, table VII illustrates the result of clustering using Gamma mixture and displays 5 miss-classified images which are 2 hummers, 2 hearts and one key which represents 95% accuracy. Last experiment was conducted by applying Gaussian mixture on shape modeling and as shown in table VIII, there are 11 miss-classified objects (2 bones, 3 forks, 2 hummers, 2 hearts, and 2 keys) which give 89% accuracy.

[1]http://www.dabi.temple.edu/ shape/MPEG7/dataset.html

TABLE III
REAL AND ESTIMATED PARAMETERS OF TWO-DIMENSIONAL DATA SET GENERATED FROM ONE GENERALIZED GAMMA COMPONENT

	real parameters		estimated parameters	
Component 1	$\alpha_{11} = 2$	$\alpha_{12} = 4$	$\hat{\alpha}_{11} = 1.95$	$\hat{\alpha}_{12} = 3.3$
	$\beta_{11} = 2.5$	$\beta_{12} = 10$	$\hat{\beta}_{11} = 2.67$	$\hat{\beta}_{12} = 9.8$
	$\lambda_{11} = 1.5$	$\lambda_{12} = 5$	$\hat{\lambda}_{11} = 1.51$	$\hat{\lambda}_{12} = 5$

TABLE IV
REAL AND ESTIMATED PARAMETERS OF TWO-DIMENSIONAL DATA SET GENERATED FROM TWO GENERALIZED GAMMA COMPONENTS

	real parameters		estimated parameters	
Component 1	$\alpha_{11} = 2$	$\alpha_{12} = 4$	$\hat{\alpha}_{11} = 1.85$	$\hat{\alpha}_{12} = 3.091$
	$\beta_{11} = 2.5$	$\beta_{12} = 10$	$\hat{\beta}_{11} = 2.58$	$\hat{\beta}_{12} = 10.67$
	$\lambda_{11} = 1.5$	$\lambda_{12} = 5$	$\hat{\lambda}_{11} = 1.5$	$\hat{\lambda}_{12} = 5$
component 2	$\alpha_{21} = 5$	$\alpha_{22} = 1.2$	$\hat{\alpha}_{21} = 5.16$	$\hat{\alpha}_{22} = 1.31$
	$\beta_{21} = 7$	$\beta_{22} = 8$	$\hat{\beta}_{21} = 6.25$	$\hat{\beta}_{22} = 6.83$
	$\lambda_{21} = 10$	$\lambda_{22} = 2$	$\hat{\lambda}_{21} = 9.8$	$\hat{\lambda}_{22} = 1.9$

TABLE V
REAL AND ESTIMATED PARAMETERS OF TWO-DIMENSIONAL DATA SET GENERATED FROM THREE GENERALIZED GAMMA COMPONENTS

	real parameters		estimated parameters	
Component 1	$\alpha_{11} = 2$	$\alpha_{12} = 4$	$\hat{\alpha}_{11} = 1.75$	$\hat{\alpha}_{12} = 3.09$
	$\beta_{11} = 2.5$	$\beta_{12} = 10$	$\hat{\beta}_{11} = 2.7$	$\hat{\beta}_{12} = 10.5$
	$\lambda_{11} = 1.5$	$\lambda_{12} = 5$	$\hat{\lambda}_{11} = 1.48$	$\hat{\lambda}_{12} = 5$
component 2	$\alpha_{21} = 5$	$\alpha_{22} = 1.2$	$\hat{\alpha}_{21} = 5.16$	$\hat{\alpha}_{22} = 1.33$
	$\beta_{21} = 7$	$\beta_{22} = 8$	$\hat{\beta}_{21} = 6.2$	$\hat{\beta}_{22} = 6.7$
	$\lambda_{21} = 10$	$\lambda_{22} = 2$	$\hat{\lambda}_{21} = 9.5$	$\hat{\lambda}_{22} = 2.7$
component 3	$\alpha_{31} = 9.4$	$\alpha_{32} = 6$	$\hat{\alpha}_{31} = 9.34$	$\hat{\alpha}_{32} = 6.1$
	$\beta_{31} = 4.2$	$\beta_{32} = 11.6$	$\hat{\beta}_{31} = 4.67$	$\hat{\beta}_{32} = 10.95$
	$\lambda_{31} = 15.7$	$\lambda_{32} = 10$	$\hat{\lambda}_{31} = 16.5$	$\hat{\lambda}_{32} = 9.1$

TABLE VI
CONFUSION MATRIX FOR GENERALIZED GAMMA MIXTURE MODEL

	Bones	Hearts	Keyes	Fountains	Forks
Bones	20	0	0	0	0
Hearts	0	20	0	0	0
Keyes	0	0	20	0	0
Fountains	3	0	0	17	0
Forks	0	0	0	0	20

TABLE VII
CONFUSION MATRIX FOR GAMMA MIXTURE MODEL

	Bones	Hearts	Keyes	Fountains	Forks
Bones	20	0	0	0	0
Hearts	0	20	0	0	0
Keyes	0	2	18	0	0
Fountains	1	0	1	18	0
Forks	0	0	1	0	19

TABLE VIII
CONFUSION MATRIX FOR GAUSSIAN MIXTURE MODEL

	Bones	Hearts	Keyes	Fountains	Forks
Bones	18	2	0	0	0
Hearts	0	17	1	2	0
Keyes	2	0	18	0	0
Fountains	1	2	0	17	0
Forks	0	2	0	0	18

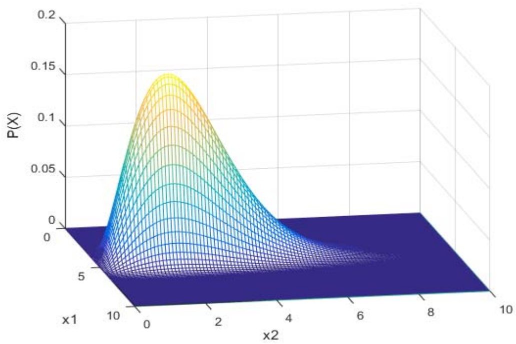

Fig. 3. One component of two-dimensional generalized Gamma mixture model

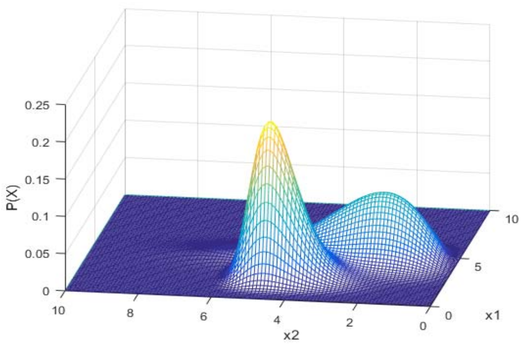

Fig. 4. two components of two-dimensional generalized Gamma mixture models

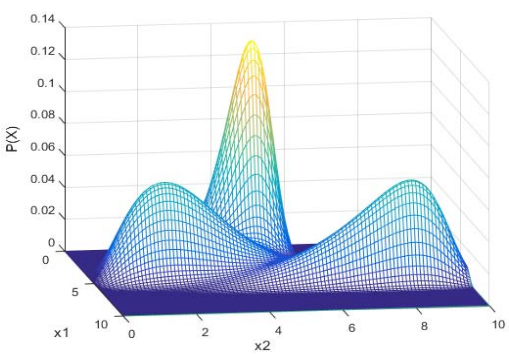

Fig. 5. Three components of two-dimensional generalized Gamma mixture models

Fig. 6. Samples from MPEG-7 CE Shape-1 Part-B

V. Conclusion

Recently, mixture models have shown promising results for clustering multi-dimensional data. The main goal of this contribution was to design a powerful and flexible statistical mixture model capable of providing an efficient modeling of multi-dimensional unlabeled positive vectors. One of the outstanding properties the generalized Gamma mixture model owns is the flexibility of its shape, which is the main reason why it outperforms Gaussian and Gamma mixture models. The model parameters have been estimated by conducting ML approach via EM. Experimental results are obtained by using a real-life application, namely shapes clustering. Consequently, these results have shown the high efficiency and capability that the proposed model possesses. Future work will be devoted to extend the proposed model to develop a method to tackle the issue of selecting the optimal number of components, which represent the data [43]–[46], or to feature selection [47]–[51].

Appendix A

As mentioned before, we have used the EM algorithm to estimate the parameters which maximize the log-likelihood computed in the E-step, the first derivative of $\log p(\mathcal{X}|\Theta)$ W.R.T the parameters $p_j, \alpha_{jd}, \beta_{jd}$ and λ_{id} has to be calculated.

By computing the derivative of $\log p(\mathcal{X}|\Theta)$ w.r.t p_j, we obtain:

$$\frac{\partial}{\partial p_j}\left[\sum_{i=1}^{N}\log\sum_{j=1}^{M}p_j\prod_{i=1}^{D}(\frac{\lambda_{jd}}{\alpha_{jd}^{\beta_{jd}}})+\log(X_{id}^{\beta_{jd}-1})+\right.$$

$$\left.\log\left(\exp-(\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\right)-\log\left(\Gamma(\frac{\beta_{jd}}{\lambda_{jd}})\right)\right]+\frac{\partial\Lambda(1-\sum_{i=1}^{M}p_j)}{\partial p_j}=0$$

$$\sum_{i=1}^{N}\frac{\prod_{i=1}^{D}(\frac{\lambda_{jd}}{\alpha_{jd}^{\beta_{jd}}})+\log(X_{id}^{\beta_{jd}-1})+\log\left(\exp-(\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\right)-\log\left(\Gamma(\frac{\beta_{jd}}{\lambda_{jd}})\right)}{\sum_{j=1}^{M}p_j\prod_{i=1}^{D}(\frac{\lambda_{jd}}{\alpha_{jd}^{\beta_{jd}}})+\log(X_{id}^{\beta_{jd}-1})+\log\left(\exp-(\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\right)-\log\left(\Gamma(\frac{\beta_{jd}}{\lambda_{jd}})\right)}-\Lambda$$

$$\frac{\sum_{i=1}^{N}p(j|\vec{X}_i)}{p_j}-\Lambda=0$$

$$p_j=\frac{\sum_{i=1}^{N}p(j|\vec{X}_i)}{\Lambda}$$

Computing the derivative w.r.t Λ we obtain:

$$1-\sum_{j=1}^{M}p_j=0$$

which gives us $\sum_{j=1}^{M}p_j=1$. Thus,

$$\sum_{j=1}^{M}\frac{\sum_{i=1}^{N}p(j|\vec{X}_i)}{\Lambda}=\frac{\sum_{i=1}^{N}\sum_{j=1}^{M}p(j|\vec{X}_i)}{\Lambda}=1$$

Since $\sum_{j=1}^{M}=1$, we obtain $\Lambda=N$, then

$$p_j=\frac{\sum_{i=1}^{N}p(j|\vec{X}_i)}{N}$$

By computing the first derivative of $\log p(\chi|\Theta)$ W.R.T α_{jd}, we obtain:

$$\sum_{i=1}^{N} p(j|\vec{X}_i)\frac{\partial}{\partial\alpha_{jd}}\left[\log(\frac{\lambda_{jd}}{\alpha_{jd}^{\beta_{jd}}})+\log(X_{id}^{\beta_{jd}-1})\right.$$
$$\left.+\log\left(\exp-(\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\right)-\log\left(\Gamma(\frac{\beta_{jd}}{\lambda_{jd}})\right)\right]$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\frac{\partial}{\partial\alpha_{jd}}\left[\log(\frac{\lambda_{jd}}{\alpha_{jd}^{\beta_{jd}}})+\log\left(\exp-(\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\right)\right.$$
$$\left.-\log\left(\Gamma(\frac{\beta_{jd}}{\lambda_{jd}})\right)\right]$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\frac{\partial}{\partial\alpha_{jd}}\left[\log(\lambda_{jd}\alpha_{jd}^{-\beta_{jd}})-(\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\right]$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\left[\frac{\alpha_{jd}^{\beta_{jd}}}{\lambda_{jd}}\frac{\partial}{\partial\alpha_{jd}}(\lambda_{jd}\alpha_{jd}^{-\beta_{jd}})-\lambda_{jd}\left(-(\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}-1}\right)\right.$$
$$\left.\times\frac{\partial}{\partial\alpha}(\frac{X_{id}}{\alpha_{jd}})\right]$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\left[\left(\frac{\alpha_{jd}^{\beta_{jd}}}{\lambda_{jd}}-\beta_{jd}\lambda_{jd}\alpha_{jd}^{-\beta_{jd}-1}\right)-\lambda_{jd}\left(-(\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}-1}\right)\right.$$
$$\left.\times\frac{\partial}{\partial\alpha}(X_{id}\alpha_{jd}{}^{-1})\right]$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\left(-\frac{\beta_{jd}}{\alpha_{jd}}-\lambda_{jd}\left(-(\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}-1}\right)(X_{id}\alpha_{jd}{}^{-2})\right)$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\left[-\frac{\beta_{jd}}{\alpha_{jd}}+\lambda_{jd}\left((\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}-1}\right)(\frac{X_{id}}{\alpha_{jd}{}^{2}})\right]$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\left[-\frac{\beta_{jd}}{\alpha_{jd}}+\left(\frac{\lambda_{jd}X_{id}^{\lambda_{jd}-1}}{\alpha_{jd}^{\lambda-1}}\right)(\frac{X_{id}}{\alpha_{jd}^{2}})\right]$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\left[-\frac{\beta_{jd}}{\alpha_{jd}}+\left(\frac{\lambda_{jd}X_{id}^{\lambda_{jd}}}{\alpha_{jd}^{\lambda_{jd}+1}}\right)\right]$$

then by setting the derivative of $\log p(\chi|\Theta)$ W.R.T α_{jd} equal to zero, we obtain:

$$\frac{\partial p(\mathcal{X},Z|\Theta)}{\partial\alpha_{jd}}=0$$
$$\sum_{i=1}^{N} p(j|\vec{X}_i)\left[-\frac{\beta_{jd}}{\alpha_{jd}}+\left(\frac{\lambda_{jd}X_{id}^{\lambda_{jd}}}{\alpha_{jd}^{\lambda_{jd}+1}}\right)\right]=0$$
$$\frac{\beta_{jd}}{\lambda_{jd}}\sum_{i=1}^{N} p(j|\vec{X}_i)=\frac{\lambda_{jd}}{\alpha_{jd}^{\lambda_{jd}+1}}\sum_{i=1}^{N} p(j|\vec{X}_i)X_{id}^{\lambda_{jd}}$$
$$\alpha_{jd}^{\lambda_{jd}}=\frac{\lambda_{jd}}{\beta_{jd}}\frac{\sum_{i=1}^{N} p(j|\vec{X}_i)X_{id}^{\lambda_{jd}}}{\sum_{i=1}^{N} P(j|\vec{X}_i)}$$
$$\alpha_{jd}=\left(\frac{\lambda_{jd}}{\beta_{jd}}\frac{\sum_{i=1}^{N} p(j|\vec{X}_i)X_{id}^{\lambda_{jd}}}{\sum_{i=1}^{N} p(j|\vec{X}_i)}\right)^{\frac{1}{\lambda_{jd}}}$$

which gives us the estimated value of α_{jd},

$$\alpha_{jd}=\left(\frac{\lambda_{jd}}{\beta_{jd}}\frac{\sum_{i=1}^{N} p(j|\vec{X}_i)X_{id}^{\lambda_{jd}}}{NP_j}\right)^{\frac{1}{\lambda_{jd}}}$$

By computing the first derivative of $\log p(\chi|\Theta)$ W.R.T β_{jd}, we obtain:

$$\frac{\partial p(\mathcal{X}|\Theta)}{\partial\beta_{jd}}=\sum_{i=1}^{N} p(j|\vec{X}_i)\frac{\partial}{\partial\beta_{jd}}\left[\log(\frac{\lambda_{jd}}{\alpha_{jd}^{\beta_{jd}}})+\log(X_{id}^{\beta_{jd}-1})\right.$$
$$\left.+\log\left(\exp-(\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\right)-\log\left(\Gamma(\frac{\beta_{jd}}{\lambda_{jd}})\right)\right]$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\frac{\partial}{\partial\beta_{jd}}\left[\log(\frac{\lambda_{jd}}{\alpha_{jd}^{\beta_{jd}}})+\log(X_{id}^{\beta_{jd}-1})\right.$$
$$\left.-\log\left(\Gamma(\frac{\beta_{jd}}{\lambda_{jd}})\right)\right]$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\left[(\frac{\alpha_{jd}^{\beta_{jd}}}{\lambda_{jd}})\frac{\partial}{\partial\beta_{jd}}(\frac{\lambda_{jd}}{\alpha_{jd}^{\beta_{jd}}})+\log(X_{id})-\frac{\Psi(\frac{\beta_{jd}}{\lambda_{jd}})}{\lambda_{jd}}\right]$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\left[(\frac{\alpha_{jd}^{\beta_{jd}}}{\lambda_{jd}})\lambda_{jd}\frac{\partial}{\partial\beta_{jd}}(\frac{1}{\alpha_{jd}^{\beta_{jd}}})+\log(X_{id})-\frac{\Psi(\frac{\beta_{jd}}{\lambda_{jd}})}{\lambda_{jd}}\right]$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\left[(\frac{\alpha_{jd}^{\beta_{jd}}}{\lambda_{jd}})\lambda_{jd}\left(\log(\alpha_{jd})\right)(\alpha_{jd}^{-\beta_{jd}})\frac{\partial}{\partial\beta_{jd}}(-\beta_{jd})\right.$$
$$\left.+\log(X_{id}^{\beta_{jd}-1})-\frac{\Psi(\frac{\beta_{jd}}{\lambda_{jd}})}{\lambda_{jd}}\right]$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\left[(\alpha_{jd}{}^{\beta_{jd}})\left(\log(\alpha_{jd})\right)(\alpha_{jd}^{-\beta_{jd}})(-1)\right.$$
$$\left.+\log(X_{id})-\frac{\Psi(\frac{\beta_{jd}}{\lambda_{jd}})}{\lambda_{jd}}\right]$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\left[-\log(\alpha_{jd})+\log(X_{id})-\frac{\Psi(\frac{\beta_{jd}}{\lambda_{jd}})}{\lambda_{jd}}\right]$$

then by setting the derivative of $\log p(\chi|\Theta)$ W.R.T β_{jd} equal to zero, we obtain:

$$\frac{\partial p(\mathcal{X}|\Theta)}{\partial\beta_{jd}}=0$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\left[-\log(\alpha_{jd})+\log(X_{id})-\frac{\Psi(\frac{\beta_{jd}}{\lambda_{jd}})}{\lambda_{jd}}\right]=0$$
$$=\sum_{i=1}^{N} p(j|\vec{X}_i)\left(-\log(\alpha_{jd})-\frac{\Psi(\frac{\beta jd}{\lambda_{jd}})}{\lambda_{jd}}\right)+\sum_{i=1}^{N} p(j|\vec{X}_i)\log(X_{id})$$
$$\sum_{i=1}^{N} p(j|\vec{X}_i)\left(\log(\alpha_{jd})+\frac{\Psi(\frac{\beta jd}{\lambda_{jd}})}{\lambda_{jd}}\right)=\sum_{i=1}^{N} p(j|\vec{X}_i)\log(X_{id})$$
$$\frac{\Psi(\frac{\beta_{jd}}{\lambda_{jd}})}{\lambda_{jd}}=\frac{\sum_{i=1}^{N} p(j|\vec{X}_i)\log(X_{id})}{\sum_{i=1}^{N} p(j|\vec{X}_i)}-\log(\alpha_{jd})$$
$$\Psi(\frac{\beta jd}{\lambda_{jd}})=\lambda_{jd}\left(\frac{\sum_{i=1}^{N} p(j|\vec{X}_i)\log(X_{id})}{\sum_{i=1}^{N} p(j|\vec{X}_i)}-\log(\alpha_{jd})\right)$$

which gives us the estimated value of β_{jd},

$$\beta_{jd} = \lambda_{id}\Psi^{-1}\left(\lambda_{jd}\left(\frac{\sum_{i=1}^{N} p(j|\vec{X_{id}})\log(X_{id})}{\sum_{i=1}^{N} p(j|\vec{X_{id}})} - \log(\alpha_{jd})\right)\right)$$

By computing the first derivative of $\log p(\chi|\Theta)$ W.R.T λ_{jd}, we obtain:

$$\frac{\partial p(\mathcal{X}|\Theta)}{\partial \lambda_{jd}} = \sum_{i=1}^{N} p(j|\vec{X}_i)\frac{\partial}{\partial \lambda_{jd}}\left[\log(\frac{\lambda_{jd}}{\alpha_{jd}^{\beta_{jd}}}) + \log(X_{id}^{\beta_{jd}-1}) \right.$$
$$\left. + \log\left(\exp -(\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\right) - \log\left(\Gamma(\frac{\beta_{jd}}{\lambda_{jd}})\right)\right]$$
$$= \sum_{i=1}^{N} p(j|\vec{X}_i)\left[\log(\frac{\lambda_{jd}}{\alpha_{jd}^{\beta_{jd}}}) + \log\left(\exp -(\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\right)\right.$$
$$\left. - \log\left(\Gamma(\frac{\beta_{jd}}{\lambda_{jd}})\right)\right]$$
$$= \sum_{i=1}^{N} p(j|\vec{X}_i)\left[(\frac{\alpha_{jd}^{\beta_{jd}}}{\lambda_{jd}})\frac{\partial}{\partial \lambda_{jd}}(\lambda_{jd}\alpha_{jd}^{-\beta_{jd}})\right.$$
$$\left. -(\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\log(\frac{X_{id}}{\alpha_{jd}}) - \frac{1}{\Gamma\left(\frac{\beta_{jd}}{\lambda_{jd}}\right)}\frac{\partial}{\partial \lambda_{jd}}\left(\Gamma\left(\frac{\beta_{jd}}{\lambda_{id}}\right)\right)\right]$$
$$= \sum_{i=1}^{N} p(j|\vec{X}_i)\left[(\frac{\alpha_{jd}^{\beta_{jd}}}{\lambda_{jd}})(\alpha_{jd}^{-\beta_{jd}}) - (\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\log(\frac{X_{id}}{\alpha_{jd}})\right.$$
$$\left. -\beta_{jd}\frac{\partial}{\partial \lambda_{jd}}\left(\frac{1}{\lambda_j}\right) \times \psi\left(\frac{\beta_{jd}}{\lambda_{id}}\right)\right]$$
$$= \sum_{i=1}^{N} p(j|\vec{X}_i)\left[\frac{1}{\lambda_{jd}} - \left((\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\log(\frac{X_{id}}{\alpha_{jd}})\right)\right.$$
$$\left. + \frac{\frac{\partial}{\partial \lambda_{jd}}\lambda jd}{\lambda_{jd}^2}\beta_{jd}\psi\left(\frac{\beta_{jd}}{\lambda_{id}}\right)\right]$$
$$= \sum_{i=1}^{N} p(j|\vec{X}_i)\left[\frac{1}{\lambda_{jd}} - \left((\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\log(\frac{X_{id}}{\alpha_{jd}})\right) + \frac{\Psi(\frac{\beta_{jd}}{\lambda_{jd}})\beta_{jd}}{\lambda_{jd}^2}\right]$$

APPENDIX B

By computing the second derivative of $\log p(\chi|\Theta)$ W.R.T λ_{jd}, we obtain:

$$\frac{\partial^2 \lambda_{jd}}{\partial^2 \lambda_{jd}}\sum_{i=1}^{N} p(j|\vec{X}_i)\left[\frac{1}{\lambda_{jd}} - \left((\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\log(\frac{X_{id}}{\alpha_{jd}})\right) + \frac{\Psi(\frac{\beta_{jd}}{\lambda_{jd}})\beta_{jd}}{\lambda_{jd}^2}\right]$$
$$= \sum_{i=1}^{N} p(j|\vec{X}_i)\left[-\frac{1}{\lambda_{id}^2} - \log\left(\frac{x_{id}}{\alpha_{jd}}\right)\frac{\partial \lambda_{jd}}{\partial \lambda_{jd}}\left[\left(\frac{x_{id}}{\alpha_{jd}}\right)^{\lambda_{jd}}\right]\right.$$
$$\left. +\beta_{jd}\frac{\partial \lambda_{jd}}{\partial \lambda_{jd}}\left(\frac{\psi(\frac{\beta_{jd}}{\lambda_{jd}})}{\lambda_{id}^2}\right)\right]$$
$$= \sum_{i=1}^{N} p(j|\vec{X}_i)\left[-\frac{1}{\lambda_{id}^2} - \log\left(\frac{x_{id}}{\alpha_{jd}}\right)\left(\frac{x_{id}}{\alpha_{jd}}\right)^{\lambda_{id}}\log\left(\frac{x_{id}}{\alpha_{jd}}\right)\right.$$
$$\left. +\beta_{jd}\frac{\partial \lambda_{jd}}{\partial \lambda_{jd}}\left(\frac{\psi(\frac{\beta_{jd}}{\lambda_{jd}})}{\lambda_{id}^2}\right)\right]$$
$$= \sum_{i=1}^{N} p(j|\vec{X}_i)\left[-\frac{1}{\lambda_{id}^2} - \left(\frac{x_{id}}{\alpha_{jd}}\right)^{\lambda_{jd}}\log^2\left(\frac{x_{id}}{\alpha_{jd}}\right)\right.$$
$$\left. +\beta_{jd}\frac{\partial \lambda_{jd}}{\partial \lambda_{jd}}\left(\frac{\psi(\frac{\beta_{jd}}{\lambda_{jd}})}{\lambda_{id}^2}\right)\right]$$
$$= \sum_{i=1}^{N} p(j|\vec{X}_i)\left[-\frac{1}{\lambda_{jd}^2} - (\frac{X_{id}}{\alpha_{jd}})^{\lambda_{jd}}\log(\frac{X_{id}}{\alpha_{jd}})\log(\frac{X_{id}}{\alpha_{jd}})\right.$$
$$\left. -\left(\frac{\psi'(\frac{\beta_{jd}}{\lambda_{jd}})\beta^2}{\lambda^4}\right) - \left(\frac{\beta_{jd}2\psi(\frac{\beta_{jd}}{\lambda_{jd}})}{\lambda_{jd}^3}\right)\right]$$

REFERENCES

[1] Z. Tufekci, *Big Questions for social media big data: Representativeness, validity and other methodological pitfalls*. The AAAI Press, 2014, pp. 505–514.

[2] N. Bouguila, "Spatial color image databases summarization," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 1, 2007, pp. I–953–I–956.

[3] ——, "Count data modeling and classification using finite mixtures of distributions," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 186–198, 2011.

[4] N. Bouguila and W. ElGuebaly, "A generative model for spatial color image databases categorization," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 821–824.

[5] ——, "On discrete data clustering," in *Advances in Knowledge Discovery and Data Mining, 12th Pacific-Asia Conference, PAKDD 2008, Osaka, Japan, May 20-23, 2008 Proceedings*, ser. Lecture Notes in Computer Science, T. Washio, E. Suzuki, K. M. Ting, and A. Inokuchi, Eds., vol. 5012. Springer, 2008, pp. 503–510. [Online]. Available: https://doi.org/10.1007/978-3-540-68125-0˙44

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

[7] L. Portnoy, E. Eskin, and S. Stolfo, "Intrusion detection with unlabeled data using clustering," in *Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001*, 2001, pp. 5–8.

[8] X. Zhu, "Semi-supervised learning literature survey," Computer Sciences, University of Wisconsin-Madison, Tech. Rep. 1530, 2005.

[9] G. McLachlan and D. Peel, *Finite Mixture Models*, ser. Wiley Series in Probability and Statistics. Wiley, 2004.

[10] O. Amayri and N. Bouguila, "Infinite langevin mixture modeling and feature selection," in *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Oct 2016, pp. 149–155.

[11] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, March 2002.

[12] W. Pan, J. Lin, and C. T. Le, "A mixture model approach to detecting differentially expressed genes with microarray data," *Functional & Integrative Genomics*, vol. 3, no. 3, pp. 117–124, Jul 2003.

[13] N. A. Ming-Hsuan Yang, "Gaussian mixture model for human skin color and its applications in image and video databases," *Proc. SPIE*, vol. 3656, p. 3656, 1998.

[14] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 2, Aug 2004, pp. 28–31 Vol.2.

[15] D. Peel and G. J. McLachlan, "Robust mixture modelling using the t distribution," *Statistics and Computing*, vol. 10, no. 4, pp. 339–348, Oct 2000.

[16] T. Elguebaly and N. Bouguila, "Infinite generalized gaussian mixture modeling and applications," in *Image Analysis and Recognition*, M. Kamel and A. Campilho, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 201–210.

[17] M. Azam and N. Bouguila, "Unsupervised keyword spotting using bounded generalized gaussian mixture model with ica," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2015, pp. 1150–1154.

[18] T. Bdiri and N. Bouguila, "Positive vectors clustering using inverted Dirichlet finite mixture models," *Expert Systems with Applications*, vol. 39, no. 2, pp. 1869 – 1882, 2012.

[19] F. R. Al-Osaim and N. Bouguila, "A finite gamma mixture model-based discriminative learning frameworks," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, Dec 2015, pp. 819–824.

[20] D. Ziou and N. Bouguila, "Unsupervised learning of a finite gamma mixture using mml: application to sar image analysis," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 2, 2004, pp. 68–71 Vol.2.

[21] T. Elguebaly and N. Bouguila, "Finite asymmetric generalized gaussian mixture models learning for infrared object detection," *Computer Vision and Image Understanding*, vol. 117, no. 12, pp. 1659 – 1671, 2013.

[22] W. Fan and N. Bouguila, "Variational learning of a dirichlet process of generalized dirichlet distributions for simultaneous clustering and feature selection," *Pattern Recognition*, vol. 46, no. 10, pp. 2754–2769, 2013.

[23] T. Elguebaly and N. Bouguila, "Background subtraction using finite mixtures of asymmetric gaussian distributions and shadow detection," *Mach. Vis. Appl.*, vol. 25, no. 5, pp. 1145–1162, 2014.

[24] S. Boutemedjet, D. Ziou, and N. Bouguila, "Model-based subspace clustering of non-gaussian data," *Neurocomputing*, vol. 73, no. 10-12, pp. 1730–1739, 2010.

[25] T. Elguebaly and N. Bouguila, "Bayesian learning of generalized gaussian mixture models on biomedical images," in *Artificial Neural Networks in Pattern Recognition, 4th IAPR TC3 Workshop, ANNPR 2010, Cairo, Egypt, April 11-13, 2010. Proceedings*, ser. Lecture Notes in Computer Science, F. Schwenker and N. E. Gayar, Eds., vol. 5998. Springer, 2010, pp. 207–218. [Online]. Available: https://doi.org/10.1007/978-3-642-12159-3˙19

[26] S. Boutemedjet, N. Bouguila, and D. Ziou, "Feature selection for non gaussian mixture models," in *2007 IEEE Workshop on Machine Learning for Signal Processing*, 2007, pp. 69–74.

[27] A. Sefidpour and N. Bouguila, "Spatial color image segmentation based on finite non-gaussian mixture models," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 8993–9001, 2012. [Online]. Available: https://doi.org/10.1016/j.eswa.2012.02.024

[28] N. Bouguila, K. Almakadmeh, and S. Boutemedjet, "A finite mixture model for simultaneous high-dimensional clustering, localized feature selection and outlier rejection," *Expert Syst. Appl.*, vol. 39, no. 7, pp. 6641–6656, 2012.

[29] N. Bouguila and D. Ziou, "A probabilistic approach for shadows modeling and detection," in *Proceedings of the 2005 International Conference on Image Processing, ICIP 2005, Genoa, Italy, September 11-14, 2005*, 2005, pp. 329–332.

[30] T. Elguebaly and N. Bouguila, "A hierarchical nonparametric bayesian approach for medical images and gene expressions classification," *Soft Comput.*, vol. 19, no. 1, pp. 189–204, 2015.

[31] D. Ziou, N. Bouguila, M. S. Allili, and A. ElZaart, "Finite gamma mixture modelling using minimum message length inference: application to sar image analysis," *International Journal of Remote Sensing*, vol. 30, no. 3, pp. 771–792, 2009.

[32] G. V. Trunk, "A problem of dimensionality: A simple example," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 3, pp. 306–307, July 1979.

[33] E. W. Stacy, "A Generalization of the Gamma Distribution," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1187–1192, 09 1962.

[34] S. Ganesalingam, "Classification and mixture approaches to clustering via maximum likelihood," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 38, no. 3, pp. 455–466, 1989.

[35] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, ser. Wiley Series in Probability and Statistics. Wiley, 1996.

[36] H. C. Li, V. A. Krylov, P. Z. Fan, J. Zerubia, and W. J. Emery, "Unsupervised learning of generalized gamma mixture model with application in statistical modeling of high-resolution sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 2153–2170, April 2016.

[37] G. Vegas-Sanchez-Ferrero, S. Aja-Fernandez, C. Palencia, and M. Martin-Fernandez, "A generalized gamma mixture model for ultrasonic tissue characterization," *Computational and Mathematical Methods in Medicine*, vol. 2012, p. 125, 2012.

[38] M. E.-S. Waheed, O. A. Mohamed, and M. E. A. El-Aziz, "Mixture of generalized gamma density-based score function for fastica," *Mathematical Problems in Engineering*, vol. 2011, pp. 1–14, 2011.

[39] N. Bouguila, "Shape modeling and categorization using fisher kernels," in *2011 International Conference on Multimedia Computing and Systems*, 2011, pp. 1–6.

[40] A. Khotanzad and Y. H. Hong, "Invariant image recognition by zernike moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 489–497, May 1990.

[41] W.-Y. Kim and Y.-S. Kim, "A region-based shape descriptor using zernike moments," *Signal Processing: Image Communication*, vol. 16, no. 1, pp. 95 – 102, 2000.

[42] Y. Kim and W. Kim, "Content-based trademark retrieval system using a visually salient feature," *Image and Vision Computing*, vol. 16, no. 12, pp. 931 – 939, 1998.

[43] N. Bouguila and D. Ziou, "Mml-based approach for high-dimensional unsupervised learning using the generalized dirichlet mixture," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2005, San Diego, CA, USA, 21-23 September, 2005*, 2005, p. 53.

[44] ——, "On fitting finite dirichlet mixture using ECM and MML," in *Pattern Recognition and Data Mining, Third International Conference on Advances in Pattern Recognition, ICAPR 2005, Bath, UK, August 22-25, 2005, Proceedings, Part I*, ser. Lecture Notes in Computer Science, P. Wang, M. Singh, C. Apté, and P. Perner, Eds., vol. 3686. Springer, 2005, pp. 172–182. [Online]. Available: https://doi.org/10.1007/11551188˙19

[45] ——, "Unsupervised selection of a finite dirichlet mixture model: An mml-based approach," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 8, pp. 993–1009, 2006.

[46] ——, "Mml-based approach for finite dirichlet mixture estimation and selection," in *Machine Learning and Data Mining in Pattern Recognition, 4th International Conference, MLDM 2005, Leipzig, Germany, July 9-11, 2005, Proceedings*, ser. Lecture Notes in Computer Science, P. Perner and A. Imiya, Eds., vol. 3587. Springer, 2005, pp. 42–51. [Online]. Available: https://doi.org/10.1007/11510888˙5

[47] N. Bouguila, "A model-based approach for discrete data clustering and feature weighting using MAP and stochastic complexity," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 12, pp. 1649–1664, 2009. [Online]. Available: https://doi.org/10.1109/TKDE.2009.42

[48] N. Bouguila and D. Ziou, "A countably infinite mixture model for clustering and feature selection," *Knowl. Inf. Syst.*, vol. 33, no. 2, pp. 351–370, 2012.

[49] T. Elguebaly and N. Bouguila, "Generalized gaussian mixture models as a nonparametric bayesian approach for clustering using class-specific visual features," *J. Visual Communication and Image Representation*, vol. 23, no. 8, pp. 1199–1212, 2012.

[50] W. Fan, N. Bouguila, and D. Ziou, "Unsupervised anomaly intrusion detection via localized bayesian feature selection," in *2011 IEEE 11th International Conference on Data Mining*, Dec 2011, pp. 1032–1037.

[51] N. Bouguila, "On multivariate binary data clustering and feature weighting," *Computational Statistics & Data Analysis*, vol. 54, no. 1, pp. 120–134, 2010.