Infinite Generalized Gaussian Mixture Modeling and Applications

Tarek Elguebaly and Nizar Bouguila

Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Canada, Qc, H3G 2W1
t_elgue@encs.concordia.ca, bouguila@ciise.concordia.ca

Abstract. A fully Bayesian approach to analyze infinite multidimensional generalized Gaussian mixture models (IGGM) is developed in this paper. The Bayesian framework is used to avoid model overfitting and the infinite assumption is adopted to avoid the difficult problem of finding the right number of mixture components. The utility of the proposed approach is demonstrated by applying it on texture classification and infrared face recognition, while comparing it to different other approaches.

1 Introduction

Over the last decade, technological advances have brought an explosion of data generation not only in size but also in dimension. These data pose a challenge to standard statistical methods and have received much attention recently. The importance of finding a way to model and analyze multidimensional data lie in their usefulness in wide range of applications such as image processing and computer vision. In recent years a lot of different learning algorithms were developed to recognize complex patterns, and to produce intelligent decisions based on observed data. Mixture models are one of the machine learning techniques receiving considerable attention in different applications. Mixture models are normally used to model complex data sets by assuming that each observation has arisen from one of the different groups or components [1]. In most of the applications, the Gaussian density is used in data analysis. However, many signal processing systems often operate in environments characterized by non-Gaussian and highly peaked sources [2]. Generalized Gaussian distribution (GGD) is considered as a good alternative to the Gaussian due to its shape flexibility which allows it to model a large number of non-Gaussian signals (see, for instance, [3,4,5,2]).

In the recent past, some deterministic approaches have been proposed for the estimation of generalized Gaussian mixture (GGM) models parameters (see, for instance, [6,7,4,8]). Despite the fact that deterministic approaches have dominated mixture models estimation due to their small computational time, many works have demonstrated that these methods have severe problems such as convergence to local maxima, and their tendency to overfitt the data [9] especially when data are sparse or noisy. Moreover, another important issue is the difficulty

M. Kamel and A. Campilho (Eds.): ICIAR 2011, Part I, LNCS 6753, pp. 201–210, 2011.

of getting reliable estimates in case of high dimensional data. With the computational tools evolution, researchers were encouraged to implement and use Bayesian MCMC methods and techniques as an alternative approach. Bayesian methods consider parameters to be random, and to follow different prior distributions (probability distributions). These distributions are used to describe our knowledge before considering the data, as for updating our prior beliefs the likelihood is used. Please refer to [9] for interesting and in depth discussions about the general Bayesian theory. One of the most challenging aspects, when using mixture models, is the estimation of the number of clusters that best describes the data without over or under fitting it. For this purpose, many approaches have been suggested, which can be classified from computational point of view into two groups: deterministic, and Bayesian methods. In this paper we use a Bayesian non-parametric approach based on allowing the number of components to increase to infinity as new data arrive. We describe a Bayesian algorithm for learning IGGM, and demonstrate its effectiveness by applying it to two real applications namely image texture classification and infrared face recognition.

The remainder of this paper is organized as follows. The next section describes our Bayesian learning approach. Section 3 presents the complete algorithm used for learning the model parameters. In section 4, we assess the performance of our model on different applications while comparing it to other models. Our last section is devoted to the conclusion.

2 Learning of the IGGM Model

2.1 The Mixture Model

If a d-dimensional $\boldsymbol{X} = (X_1, ..., X_d)$ follows a GGD, then:

$$P(\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{k=1}^{d} \frac{\beta_k \alpha_k}{2\Gamma(1/\beta_k)} e^{-(\alpha_k |X_k - \mu_k|)^{\beta_k}} \tag{1}$$

where $\boldsymbol{\mu}$= $(\mu_1, \ldots, \mu_d)$, $\boldsymbol{\alpha}$= $(\alpha_1, \ldots, \alpha_d)$, and $\boldsymbol{\beta}$= $(\beta_1, \ldots, \beta_d)$ are the mean, the inverse scale, and the shape parameters. Let $\mathcal{X}$= $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N)$ be a set of N iid vectors assumed to arise from a GGM with M components:

$$P(\mathcal{X}|\Theta) = \sum_{j=1}^{M} P(\boldsymbol{X}|\boldsymbol{\mu}_j, \boldsymbol{\alpha}_j, \boldsymbol{\beta}_j) p_j \tag{2}$$

where $\{p_j\}$ are the mixing proportions which must be positive and sum to one. The set of parameters of the mixture with M components is defined by $\Theta = (\{\boldsymbol{\mu}_j\}, \{\boldsymbol{\mu}_j\}, \{\boldsymbol{\alpha}_j\}, \{\boldsymbol{\beta}_j\}, \{p_j\})$. We introduce stochastic indicator variables, Z_i, one for each observation, whose role is to encode to which component the observation belongs. In other words, Z_{ij}, the unobserved or missing vector, equals 1 if $\boldsymbol{X}_i$ belongs to class j and 0, otherwise. The complete-data likelihood for this case is then:

$$P(\mathcal{X}, Z|\Theta) = \prod_{i=1}^{N} \prod_{j=1}^{M} (P(\boldsymbol{X}_i|\xi_j) p_j)^{Z_{ij}} \tag{3}$$

where $Z = \{Z_1, Z_2, ..., Z_N\}$, and $\xi_j = (\boldsymbol{\mu}_j, \boldsymbol{\alpha}_j, \boldsymbol{\beta}_j)$. Bayesian MCMC simulation methods are based on the well-known Bayesian formulae:

$$\pi(\Theta|\mathcal{X}, Z) = \frac{\pi(\Theta)P(\mathcal{X}, Z|\Theta)}{\int \pi(\Theta)P(\mathcal{X}, Z|\Theta)} \propto \pi(\Theta)P(\mathcal{X}, Z|\Theta) \tag{4}$$

where $(\mathcal{X}, Z)$, $\pi(\Theta)$ and $\pi(\Theta|\mathcal{X}, Z)$ are the complete data, the prior information about the parameters and the posterior distribution, respectively. Having $\pi(\Theta|\mathcal{X}, Z)$ we can simulate our model parameters Θ, rather than computing them.

For the $\{p_j\}$, we know that ($0 \leq p_j \leq 1$ and $\sum_{j=1}^{M} p_j = 1$), then the typical choice, as a prior, is a symmetric Dirichlet distribution with parameter η/M. As for $\pi(Z|p)$ we have:

$$\pi(Z|p) = \prod_{j=1}^{M} \pi(Z_i|p) = \prod_{i=1}^{N}\prod_{j=1}^{M} p_j^{Z_{ij}} = \prod_{j=1}^{M} p_j^{n_j} \tag{5}$$

where $n_j = \sum_{i=1}^{N} \mathbf{I}_{Z_{ij=1}}$. Then using the standard Dirichlet integral, we may integrate out the mixing proportions and write the prior directly in terms of the indicators:

$$\pi(Z|\eta) = \frac{\Gamma(\eta)}{\Gamma(\eta + N)} \prod_{j=1}^{M} \frac{\Gamma(\eta/M + n_j)}{\Gamma(\eta/M)} \tag{6}$$

In order to be able to use Gibbs sampling for the missing vector, Z, we need the conditional prior for a single indicator given all the others; this can be easily obtained from Eq. 6 by keeping all but a single indicator fixed:

$$\pi(Z_i = j|\eta, Z\_i) = \frac{n_{-ij} + \eta/M}{N - 1 + \eta} \tag{7}$$

where the subscript $-i$ indicates all indexes except i. Note that n_{-ij} is the number of observations, excluding $\boldsymbol{X_i}$, in cluster j . For the parameters ξ, we assign independent Normal prior with δ, ε^2 as the mean and variance for the distributions means ($\boldsymbol{\mu}_j$), respectively. Independent Gamma prior with ι, ρ as the shape and rate parameters, respectively, is assigned for the inverse scale $\boldsymbol{\alpha}_j$. For the shape parameter, $\boldsymbol{\beta}_j$, we used independent Gamma prior with κ, ς as the shape and rate parameters, respectively [10]. Thus, the posterior distributions for $\boldsymbol{\mu}_j$, $\boldsymbol{\alpha}_j$, and $\boldsymbol{\beta}_j$ are given by:

$$P(\boldsymbol{\mu}_j|Z, \mathcal{X}) \propto \prod_{k=1}^{d} \frac{1}{\varepsilon} \mathbf{e}^{\frac{-(\mu_{jk}-\delta)^2}{2\varepsilon^2}} \times \prod_{k=1}^{d} \mathbf{e}^{\sum z_{ij=1}(-\alpha_{jk}|X_{ik}-\mu_{jk}|)^{\beta_{jk}}} \tag{8}$$

$$P(\boldsymbol{\alpha}_j|Z, \mathcal{X}) \propto \prod_{k=1}^{d} \frac{\alpha_{jk}^{\iota-1}\rho^{\iota}\mathbf{e}^{-\rho\alpha_{jk}}}{\Gamma(\iota)} \times \prod_{k=1}^{d} \left[\alpha_{jk}\right]^{n_j} \mathbf{e}^{\sum z_{ij=1}(-\alpha_{jk}|X_{ik}-\mu_{jk}|)^{\beta_{jk}}} \tag{9}$$

$$P(\boldsymbol{\beta}_j|Z,\mathcal{X}) \propto \prod_{k=1}^{d} \frac{\beta_{jk}^{\kappa-1}\varsigma^{\kappa}\mathbf{e}^{-\varsigma\beta_{jk}}}{\Gamma(\kappa)} \times \prod_{k=1}^{d}\left[\frac{\beta_{jk}}{\Gamma(1/\beta_{jk})}\right]^{n_j} \mathbf{e}^{\sum_{Z_{ij=1}}(-\alpha_{jk}|X_{ik}-\mu_{jk}|)^{\beta_{jk}}} \tag{10}$$

In order to have a more flexible model, we introduce an additional hierarchical level by allowing the hyperparameters to follow some selected distributions. The hyperparameters, δ and ε^2 associated with the $\boldsymbol{\mu}_j$ are given Normal and Inverse Gamma priors with parameters (ϵ,χ^2) and (φ,ϱ), respectively. Thus,

$$P(\delta|\ldots) \propto P(\delta|\epsilon,\chi^2)\prod_{j=1}^{M} P(\boldsymbol{\mu}_j|\delta,\varepsilon^2) \propto \mathbf{e}^{\frac{-(\delta-\epsilon)^2}{2\chi^2}} \times \prod_{j=1}^{M}\prod_{k=1}^{d}\mathbf{e}^{\frac{-(\mu_{jk}-\delta)^2}{2\varepsilon^2}} \tag{11}$$

$$P(\varepsilon^2|\ldots) \propto P(\varepsilon^2|\varphi,\varrho)\prod_{j=1}^{M} P(\boldsymbol{\mu}_j|\delta,\varepsilon^2) \propto \frac{\exp(-\varrho\varepsilon^2)}{\varepsilon^{2(\varphi+1)}}\left[\frac{1}{\varepsilon}\right]^{Md} \times \prod_{j=1}^{M}\prod_{k=1}^{d}\mathbf{e}^{\frac{-(\mu_{jk}-\delta)^2}{2\varepsilon^2}} \tag{12}$$

The hyperparameters ι and ρ associated with the $\boldsymbol{\alpha}_j$ are given inverse Gamma and Gamma priors with parameters (ϑ,ϖ) and (τ,ω), respectively. Thus,

$$P(\iota|\ldots) \propto P(\boldsymbol{\alpha}_{\alpha}|\vartheta,\varpi)\prod_{j=1}^{M} P(\boldsymbol{\alpha}_j|\iota,\rho) \propto \frac{\exp(-\varpi/\iota)}{\iota^{\vartheta+1}}\left[\frac{\rho^{\iota}}{\Gamma(\iota)}\right]^{Md} \times \prod_{j=1}^{M}\prod_{k=1}^{d}\alpha_{jk}^{\iota-1}\mathbf{e}^{-\rho\alpha_{jk}} \tag{13}$$

$$P(\rho|\ldots) \propto P(\boldsymbol{\beta}_{\alpha}|\tau,\omega)\prod_{j=1}^{M} P(\boldsymbol{\alpha}_j|\iota,\rho) \propto \rho^{\tau-1}\mathbf{e}^{-\omega\rho}\left[\rho^{\iota}\right]^{Md} \times \prod_{j=1}^{M}\prod_{k=1}^{d}\alpha_{jk}^{\iota-1}\mathbf{e}^{-\rho\alpha_{jk}} \tag{14}$$

The hyperparameters κ and ς associated with the $\boldsymbol{\beta}_j$ are given inverse Gamma and Gamma priors with parameters (λ,ϕ) and (ν,ψ), respectively. Thus,

$$P(\kappa|\ldots) \propto P(\kappa|\lambda,\phi)\prod_{j=1}^{M} P(\boldsymbol{\beta}_j|\kappa,\varsigma) \propto \frac{\exp(-\phi/\kappa)}{\kappa^{\lambda+1}}\left[\frac{\varsigma^{\kappa}}{\Gamma(\kappa)}\right]^{Md} \times \prod_{j=1}^{M}\prod_{k=1}^{d}\beta_{jk}^{\kappa-1}\mathbf{e}^{-\varsigma\beta_{jk}} \tag{15}$$

$$P(\varsigma|\ldots) \propto P(\varsigma|\nu,\psi)\prod_{j=1}^{M} P(\boldsymbol{\beta}_j|\kappa,\varsigma) \propto \varsigma^{\nu-1}\mathbf{e}^{-\psi\varsigma}\left[\varsigma^{\kappa}\right]^{Md} \times \prod_{j=1}^{M}\prod_{k=1}^{d}\beta_{jk}^{\kappa-1}\mathbf{e}^{-\varsigma\beta_{jk}} \tag{16}$$

2.2 The IGGM Model

So far, we have considered M to be a fixed quantity. In this section, we overcome this obstacle by assuming that $M \to \infty$ in Eq.7 which gives us

$$\pi(Z_i = j|\eta, Z_{-i}) = \begin{cases} \frac{n_{-ij}}{N-1+\eta}; \text{ if } n_{-ij} > 0 \text{ (cluster } j \in \mathcal{R}) \\ \frac{\eta}{N-1+\eta}; \text{ if } n_{-ij} = 0 \text{ (cluster } j \in \mathcal{U}) \end{cases} \tag{17}$$

where $\mathcal{R}$ and $\mathcal{U}$ are the sets of represented and unrepresented clusters, respectively. Thus, the conditional posterior is obtained by combining this prior with the likelihood of the data:

$$\pi(Zi = j|\eta, \boldsymbol{\mu_j}, \boldsymbol{\alpha_j}, \boldsymbol{\beta_j}, Z_{-i}, \boldsymbol{X_i}) =$$

$$\begin{cases} \frac{n_{-ij}}{N-1+\eta} p(\boldsymbol{X_i}|\boldsymbol{\mu_j}, \boldsymbol{\alpha_j}, \boldsymbol{\beta_j}); & \text{if } j \in \mathcal{R} \\ \int \frac{\eta}{N-1+\eta} p(\boldsymbol{X_i}|\boldsymbol{\mu_j}, \boldsymbol{\alpha_j}, \boldsymbol{\beta_j}) p(\boldsymbol{\mu_j}|\delta, \varepsilon^2) p(\boldsymbol{\alpha_j}|\iota, \rho) p(\boldsymbol{\beta_j}|\kappa, \varsigma) d\boldsymbol{\mu_j} d\boldsymbol{\alpha_j} d\boldsymbol{\beta_j} & \text{if } j \in \mathcal{U} \end{cases} \quad (18)$$

The choice of the concentration parameter η is of high importance as it controls the generation frequency of new clusters. We decided to use an Inverse Gamma prior for η:

$$P(\eta|\upsilon, \gamma) \sim \frac{\gamma^{\upsilon} \exp(-\gamma/\eta)}{\Gamma(\upsilon)\eta^{\upsilon+1}} \quad (19)$$

Using the above equation with Eq. 17 we reach the following posterior:

$$P(\eta|...) \propto \frac{\gamma^{\upsilon} \exp(-\gamma/\eta)}{\Gamma(\upsilon)\eta^{\upsilon+1}} \eta^M \prod_{j=1}^{N} \frac{1}{i-1+\eta} \propto \frac{\gamma^{\upsilon} \exp(-\gamma/\eta)}{\Gamma(\upsilon)\eta^{\upsilon+1}} \frac{\eta^M \Gamma(\eta)}{\Gamma(N+\eta)} \quad (20)$$

Our hierarchical model can be displayed as a directed acyclic graph (DAG) as shown in Fig. 1.

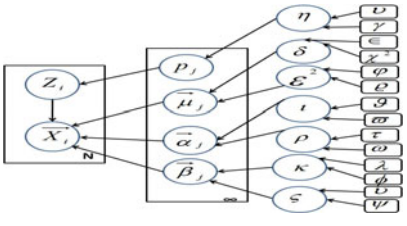

Fig. 1. Graphical Model representation of the Bayesian hierarchical IGGM model. Nodes in this graph represent random variables, rounded boxes are fixed hyperparameters, boxes indicate repetition (with the number of repetitions in the lower right) and arcs describe conditional dependencies between variables.

3 The Complete Algorithm

Having all the conditional posteriors, we can employ a Gibbs sampler with the following steps:

1. Generate Z_i from Eq. 18 then update n_j.
2. Update the number of represented components M.
3. Update the mixing parameters for the represented components by $p_j = \frac{n_j}{N+\eta}$ for $j = 1, \ldots, M$, and for the unrepresented components by $p_U = \frac{\eta}{N+\eta}$.
4. Generate the mixture parameters $\boldsymbol{\mu}_j$, $\boldsymbol{\alpha}_j$, and $\boldsymbol{\beta}_j$ from Eqs. 8, 9 and 10.
5. Update the hyperparameters δ, ε^2, ι, ρ, κ, ς, and η from Eqs. 11, 12, 13, 14, 15, 16,20, respectively.

Note that, for the initialization step we started by assuming that all the vectors are in the same cluster, and we generated the parameters by sampling from their prior distributions. It is quite easy to notice that we cannot simulate directly from these posterior distributions because they are not in well known forms. To solve this problem we applied the well known Metropolis-Hastings (M-H) algorithm given in [11].

4 Experimental Results

In the following applications, we use 5000 iterations for our Metropolis-within-Gibbs sampler (we discarded the first 800 iterations as "burn-in" and kept the rest), and our specific choices for the hypeparameters are

$$(\upsilon, \gamma, \epsilon, \chi^2, \varphi, \varrho, \vartheta, \varpi, \tau, \omega, \lambda, \phi, \nu, \psi) = (2, 0.2, 1, 0.5, 2, 5, 2, 5, 2, 0.2, 2, 5, 2, 0.2)$$

4.1 Categorization of Texture Images

In this application we are interested by the categorization of texture images which is important in the case of content-based image retrieval, for instance. In order to determine the vector of characteristics for a given texture, we use set of features derived from the image correlogram [12], by considering four neighborhoods and directions: $(1; 0)$, $(1, \pi/4)$, $(1, \pi/2)$, and $(1, 3\pi/4)$, from which we derive eight features: mean, variance, energy, correlation, entropy, contrast, homogeneity, and cluster prominence [13]. Thus, each image is represented by a 32-dimensional vector. Finally, we apply two methods, our IGGM and the infinite Gaussian mixture (IGM) [14], in order to categorize the images.

We perform our experiments using the Vistex texture data set [1]. Six homogeneous texture groups (Bark, Fabric, Food, Metal, Water, and Sand) are considered. We use four 512×512 images from each of the Bark, Fabric, and Metal texture groups, and six 512×512 from each of the Food, Water, and Sand texture groups, then we divide each image into sixty four 64×64 subimages. Now, we have a total of $1,920$ sub-images: 256 sub-images for each class in the first three groups, and 384 sub-images for each class in the second three groups. Examples of images from each of the six categories are shown in figure 2.

The IGGM mixture favored 6 categories which is the case here. The IGM, however, classified the texture images into 7 clusters where the 7^{th} component had a very small probability of 0.0276. In order to be able to compare both methods, we supposed that we obtained the right number of clusters in the case of the IGM. The confusion matrices for both methods are given in tables 1.a and 1.b. As shown the total number of misclassified images in the case of IGGM is 36 which identifies a high accuracy of 98.12%. The accuracy of the IGM was 93.80%, as it misclassified 119 images.

[1] MIT Vision and Modeling Group (http://vismod.www.media.mit.edu)

(a) (b) (c) (d) (e) (f)

Fig. 2. Sample images from each group. (a) Bark, (b) Fabric, (c) Food, (d) Metal, (e) Sand, (f) Water.

Table 1. Confusion matrix for texture categorization using (a) IGGM and (b) IGM

	Bark	Fabric	Metal	Food	Sand	Water
Bark	255	0	0	0	1	0
Fabric	0	248	0	8	0	0
Metal	0	0	252	0	0	4
Food	0	6	0	378	0	0
Sand	3	0	0	0	380	1
Water	3	2	2	4	2	371

(a)

	Bark	Fabric	Metal	Food	Sand	Water
Bark	241	0	0	1	6	5
Fabric	2	238	0	6	2	2
Metal	0	2	237	3	0	4
Food	0	7	3	362	0	2
Sand	5	0	2	0	363	3
Water	4	1	2	3	1	360

(b)

4.2 Infrared Face Recognition

Recently, different studies have shown that thermal IR offers a promising alternative to visible imagery for handling variations in face appearance [15]. Figure 3 shows visual and thermal image characteristics of faces with variations in illumination and facial expression. Although illumination and facial expression significantly change the visual appearance of the face, thermal characteristics of the face remain nearly invariant. Several approaches have been proposed to analyze and recognize infrared faces and can be divided into two main groups: appearance-based and feature-based methods. While appearance-based methods focus on the global properties of the face, feature-based methods explore the facial features (ex. eyes, mouth) statistical and geometrical properties [16]. Many of these approaches, however, suppose that the extracted infrared face features are Gaussian which is not generally an appropriate assumption. We propose then, in this section, an appearance-based approach using IGGM. We are

Fig. 3. Visual and thermal image characteristics of faces with variations in illumination

considering face recognition as an image classification problem by trying to classify to which person this image belongs. For feature extraction step we have employed both the edge orientation histograms [17] and the co-occurrence matrices which capture the local spatial relationships between gray levels [18]. Figure 4 shows 3 face images, where the first two images are for the same person taken from different poses, and the third image is for another person. It is quite clear that the first two images have very close edge-orientation histograms compared to the third image. In our experiments, we have considered the following four co-occurrence matrices: $(1;0)$, $(1,\pi/4)$, $(1,\pi/2)$, and $(1,3\pi/4)$, respectively [19]. For each co-occurrence matrix we derived the following features: mean, variance, energy, correlation, entropy, contrast, homogeneity, and cluster prominence [19]. Besides, the edge directions are quantized into 72 bins of 5° each. Using the co-occurrence matrices and the histogram of edge directions each image was represented by a 104-dimensional vector.

In our experiments, we performed face recognition using images from the Iris thermal face database which is a subset of the Object Tracking and Classification Beyond the Visible Spectrum (OTCBVS) database. First we used 1320 images of fifteen persons not wearing glasses. Knowing that in IR imaging thermal radiation cannot transmit through glasses because glasses severely attenuate electromagnetic wave radiation beyond 2.4 mm, we decided to investigate if our algorithm will be capable to identify persons with glasses, so we added 880 images of eight persons with glasses. For both experiments we used 11 images

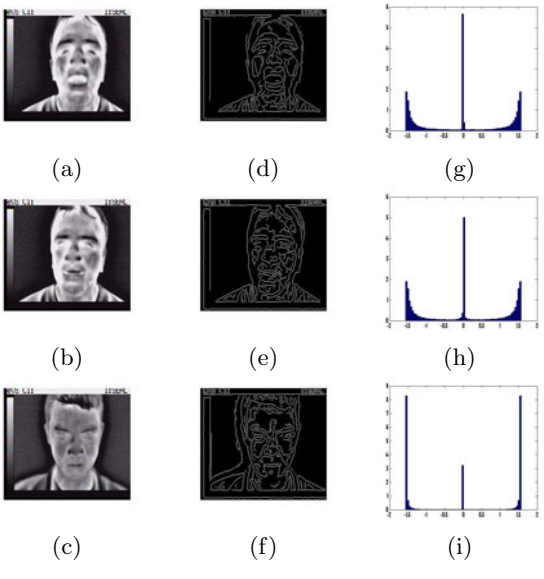

Fig. 4. 3 different images of two different persons with their corresponding shape images and corresponding shape histograms, (a)-(c) show three database images, (d)-(f) show the corresponding edge images, (g)-(i) show the corresponding shape histograms.

Table 2. Accuracies for The seven different methods

	IGGM	EMGGM	IGM	EMGM	PCA	HICA	LICA
Data 1	97.02%	94.20%	86.67%	85.89%	95.58%	95.32%	94.46%
Data 2	96.33%	92.40%	82.54%	82.18%	94.35%	93.99%	92.81%

for each person as training set and the rest as testing set. This gave us 165 and 1155 images for training and testing, respectively, in the first data. The second data set was composed of 253 and 1947 images for training and testing, respectively. In order to validate our algorithm (IGGM) we have compared it with the expectation maximization (EM) one (EMGGM). We also compared it to six other methods namely principal component analysis (PCA) with cosine distance, localized independent component analysis (LICA) with cosine distance, holistic ICA (HICA) with cosine distance as implemented by FastICA [20], IGM and Gaussian mixture models learned with EM (EMGM). Table 2 shows the accuracies for the seven different methods. According to this table it is clear that the IGMM outperforms all other methods which can be explained by its ability to incorporate prior information during classes learning and modeling.

5 Conclusion

We have described and illustrated a Bayesian nonparametric approach based on infinite generalized Gaussian mixtures. We proposed an MCMC algorithm to learn the parameters of this mixture. The effectiveness of the proposed approach has been shown using two important applications namely texture images classification and Infrared face recognition.

Acknowledgment

The completion of this research was made possible thanks to the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. McLachlan, G.J., Peel, D.: Finite Mixture Models. Wiley, New York (2000)
2. Elguebaly, T., Bouguila, N.: Bayesian Learning of Finite Generalized Gaussian Mixture Models on Images. Signal Processing 91(4), 801–820 (2011)
3. Meignen, S., Meignen, H.: On the Modeling of Small Sample Distributions with Generalized Gaussian Density in a Maximum Likelihood Framework. IEEE Transactions on Image Processing 15(6), 1647–1652 (2006)
4. Allili, M.S., Bouguila, N., Ziou, D.: Finite General Gaussian Mixture Modeling and Application to Image and Video Foreground Segmentation. Journal of Electronic Imaging 17(1), 1–13 (2008)

5. Elguebaly, T., Bouguila, N.: Bayesian learning of generalized gaussian mixture models on biomedical images. In: Schwenker, F., El Gayar, N. (eds.) ANNPR 2010. LNCS, vol. 5998, pp. 207–218. Springer, Heidelberg (2010)
6. Allili, M.S., Bouguila, N., Ziou, D.: Finite generalized gaussian mixture modeling and applications to image and video foreground segmentation. In: Proc. of the Canadian Conference on Robot and Vision (CRV), pp. 183–190 (2007)
7. Allili, M.S., Bouguila, N., Ziou, D.: A robust video foreground segmentation by using generalized gaussian mixture modeling. In: Proc. of the Canadian Conference on Robot and Vision (CRV), pp. 503–509 (2007)
8. Fan, S.-K.S., Lin, Y.: A Fast Estimation Method for the Generalized Gaussian Mixture Distribution on Complex Images. Computer Vision and Image Understanding 113(7), 839–853 (2009)
9. Robert, C.P.: The Bayesian Choice From Decision-Theoretic Foundations to Computational Implementation, 2nd edn. Springer, Heidelberg (2007)
10. Robert, C.P., Casella, G.: Monte Carlo Statistical Methods, 2nd edn. Springer, Heidelberg (2004)
11. Lewis, S.M., Raftery, A.E.: Estimating Bayes Factors via Posterior Simulation with the Laplace-Metropolis Estimator. Journal of the American Statistical Association 90, 648–655 (1997)
12. Huang, J., Kumar, S.R., Mitra, M., Zhu, W.-J., Zabih, R.: Image Indexing Using Color Correlograms. In: Proc. of the IEEE Conference Computer Vision and Pattern Recognition, p. 762 (1997)
13. Randen, T., Husoy, J.H.: Sum and Difference Histograms for Texture Classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(4), 291–310 (1999)
14. Rasmussen, C.E.: The Infinite Gaussian Mixture Model. In: Advances in Neural Information Processing Systems (NIPS), pp. 554–560 (2000)
15. Han, X., Koelling, K.W., Tomasko, D.L., Lee, L.J.: A comparative analysis of face recognition performance with visible and thermal infrared imagery. In: Proc. of the International Conference on Pattern Recognition (ICPR), pp. 217–222 (2002)
16. Arandjelovic, O., Hammoud, R., Cipolla, R.: Multi-sensory face biometric fusion (for personal identification). In: Proc. of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications, CVBVS (2006)
17. Jain, A.K., Vailaya, A.: Image retrieval using color and shape. Pattern Recognition 29, 1233–1244 (1996)
18. Shanmugam, K., Haralickand, R.M., Dinstein, I.: Texture features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, 610–621 (1973)
19. Unser, M.: Filtering for Texture Classification: A Comparative Study. IEEE Transactions on Pattern Analysis and Machine Intelligence 8(1), 118–125 (1986)
20. Hyvrinen, A.: The fixed-point algorithm and maximum likelihood estimation for independent component analysis. Neural Processing Letters 10, 1–5 (1999)