

Effective Frameworks Based on Infinite Mixture Model for Real-World Applications

Norah Saleh Alghamdi¹, Sami Bourouis^{2,*} and Nizar Bouguila³

¹Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O.Box 84428, Riyadh 11671, Saudi Arabia
²College of Computers and Information Technology, Taif University, Taif, 21944, Saudi Arabia
³The Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, QC H3G 1T7, Canada

*Corresponding Author: Sami Bourouis. Email: s.bourouis@tu.edu.sa

Received: 24 August 2021; Accepted: 01 December 2021

Abstract: Interest in automated data classification and identification systems has increased over the past years in conjunction with the high demand for artificial intelligence and security applications. In particular, recognizing human activities with accurate results have become a topic of high interest. Although the current tools have reached remarkable successes, it is still a challenging problem due to various uncontrolled environments and conditions. In this paper two statistical frameworks based on nonparametric hierarchical Bayesian models and Gamma distribution are proposed to solve some real-world applications. In particular, two nonparametric hierarchical Bayesian models based on Dirichlet process and Pitman-Yor process are developed. These models are then applied to address the problem of modelling grouped data where observations are organized into groups and these groups are statistically linked by sharing mixture components. The choice of the Gamma mixtures is motivated by its flexibility for modelling heavy-tailed distributions. In addition, deploying the Dirichlet process prior is justified by its advantage of automatically finding the right number of components and providing nice properties. Moreover, a learning step via variational Bayesian setting is presented in a flexible way. The priors over the parameters are selected appropriately and the posteriors are approximated effectively in a closed form. Experimental results based on a real-life applications that concerns texture classification and human actions recognition show the capabilities and effectiveness of the proposed framework.

Keywords: Infinite Gamma mixture model; variational Bayes; hierarchical Dirichlet process; Pitman-Yor process; texture classification; human action recognition



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction and Literature Review

Data clustering has been the subject of wide research to the present days [1–4]. The goal of clustering is to group the observed data into separate subgroups, such as the data in each subgroup shows some similarities to each other. Various clustering approaches have been developed in the past and some of them are based on distance metrics (such as the K-means and SOM algorithms). It is noteworthy that these algorithms are sensitive to noise, are not flexible when dealing with incomplete data, and may not definitely capture heterogeneity inherent in complex datasets. Another important alternative is the model-based clustering approach [5]. It is effective for modelling the structure of clusters since data are supposed to be produced by some distributions. In particular, finite mixture models (such as Gaussian mixture model) have shown to be effective (in terms of discovering complex patterns and grouping them into similar clusters) for many computer vision and pattern recognition applications [6]. Nevertheless, Gaussian model cannot fit well complex non-Gaussian shapes. To cope with the disadvantages related to conventional Gaussian assumption, many contributions have occurred to develop other flexible mixture models and were interested to non-Gaussian behavior of real datasets [7,8]. For instance, Gamma (GaMM) mixtures have demonstrated to offer high flexibility and ease of use than Gaussian [9] for many image processing and pattern recognition problems. This is due their compact analytical form which is able to cover long-tailed distributions and to approximate data with outliers. This mixture has been used with success for a range of interesting problems [7,10–12] especially when dealing with proportional data. An illustration of how mixture models can be applied for data modelling, classification and recognition is given in Fig. 1.

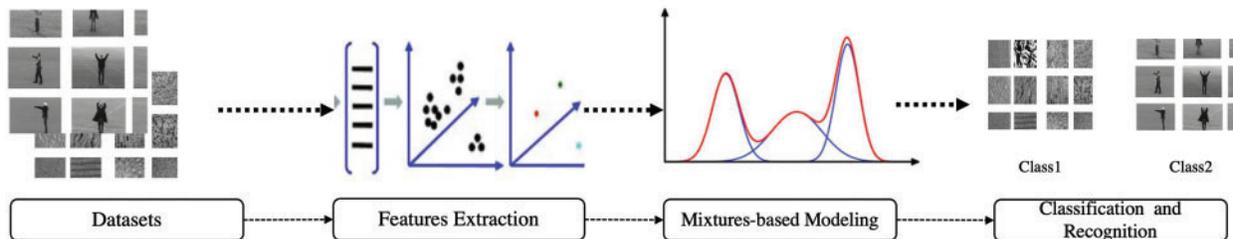


Figure 1: Block diagram of application of mixture model for data classification and recognition

Unfortunately, the inadequacy of finite mixture model has been apparent when selecting the appropriate number of mixture components. In other word, model selection (i.e., model’s complexity) is one of the difficult problems within finite mixture models. The crucial problem of “how many groups in the dataset?” still remains of great interest for various data mining fields since determining inappropriate number of clusters may conduct to poor generalization capability. This problem can be solved by considering an infinite number of components via nonparametric Bayesian methods [13,14], principally within “Dirichlet process (DP)” [15]. Indeed, a Dirichlet Process (DP) is a parameterized stochastic process characterized by a base distribution and it can be defined as a probability distribution over discrete distributions. Numerous studies have been devoted to infinite mixture models which are emerged to cope with the challenging problem of model selection.

Two sound hierarchical Bayesian alternatives to the conventional DP named as hierarchical Dirichlet process (HDP) and hierarchical Pitman-Yor process (HPYP) have exposed encouraging results especially when dealing with modeling grouped data [16,17]. Indeed, within hierarchical models, mixture components can be shared through data groups (i.e., parameters are shared among groups). In such cases, it is possible to make a Bayesian hierarchy on the DP and the base distribution of the DP is distributed according to another DP.

Thus, the main contributions of this manuscript are first to extend our previous works about the Gamma mixture by investigating two efficient nonparametric hierarchical Bayesian models based on both Dirichlet and Pitman-Yor processes mixtures of Gamma distributions. Indeed, in order to reach enhanced modeling performance, we consider Gamma distribution which is able to cover long-tailed distributions and to approximate accurately visual vectors. Another critical issue when dealing with mixture models is model parameters estimation. Accordingly, we propose to develop an effective variational Bayes learning algorithm to estimate the parameters of the implemented models. It is noteworthy to indicate that the complexity of variational inference-based algorithm still remains less than Markov Chain Monte Carlo-Based Bayesian inference and leads to faster convergence. Finally, the implemented hierarchical Bayesian models and the variational inference approaches are validated via challenging real-life problems namely human activity recognition and texture categorization.

This manuscript is organized as follows. In Section 2 we introduce the hierarchical DP and PYP mixtures of Gamma distributions which are based on stick-breaking construction. In Section 3, we describe the details of our variational Bayes learning framework. Section 4 reports the obtained results, which are based on two challenging applications, to verify the merits and effectiveness of our framework, and Section 5 is devoted to conclude the manuscript.

2 Model Specification

In this section, we start by briefly presenting finite Gamma mixture model and then we present our nonparametric frameworks based on hierarchical Dirichlet and Pitman-Yor processes mixtures.

2.1 Finite Gamma Mixture Model

If a D -dimensional random vector $\vec{Y} = (Y_1, \dots, Y_D)$ is distributed according to a multidimensional Gamma distribution, then its probability density function (pdf) is defined as:

$$Ga(\vec{Y} | \vec{\theta}) = \prod_{d=1}^D \frac{(\beta_d)^{\alpha_d} Y_d^{\alpha_d-1} e^{-\beta_d Y_d}}{\Gamma(\alpha_d)} \tag{1}$$

where $\vec{\theta} = \{\alpha_d, \beta_d\}$ is the vector of parameters of the Gamma distribution. Here α_d denotes the shape and β_d is for the location parameters. $\Gamma(\cdot)$ is the Gamma function which is defined as: $\Gamma(x) = \int_0^\infty s^{x-1} e^{-s} ds$.

For a random vector \vec{Y} which is produced from a finite Gamma mixture model with M components, we have the following:

$$p(\vec{Y} | \vec{\pi}, \vec{\theta}) = p(\vec{Y} | \vec{\pi}, \vec{\alpha}, \vec{\beta}) = \sum_{j=1}^M \pi_j Ga(\vec{Y} | \vec{\theta}_j) \tag{2}$$

where $\vec{\theta}_j = \{\alpha_j, \beta_j\}$ is the set of parameters of the Gamma distribution $Ga(\cdot)$ corresponding to the j^{th} mixture component. The vector of mixing weights $\vec{\pi} = (\pi_1, \dots, \pi_M)$ is subject to $0 \leq \pi_j \leq 1$, and $\sum_{j=1}^M \pi_j = 1$.

2.2 Hierarchical Dirichlet Process Mixture Model

The hierarchical Dirichlet process (HDP) is an effective nonparametric Bayesian method to modelling grouped data, which allows the mixture models to share components. Here, observed data are arranged into groups (i.e., mixture model) that we want to make them statistically linked. HDP is built on the Dirichlet process (DP) as well described in [17] for each group data. It is noteworthy that the DP has acquired popularity in machine learning to handle nonparametric problems [18]. The DP was presented as a prior on probability distributions and this makes it extremely appropriate for specifying infinite mixture models thanks to the use of the stick-breaking process [19]. In the case of hierarchical Dirichlet process (HDP), the DPs for all groups share a base distribution which is itself distributed according to a Dirichlet process. Let's assume that we have a grouped data set \mathcal{Y} separated into M groups, such that each group is associated with a DP G_j , thus the HDP takes part an indexed set of DPs G_j that share a global (or base) distribution G_0 which is itself distributed as a DP with base distribution H and concentration parameter γ :

$$G_0 \sim DP(\gamma, H) G_j \sim DP(\lambda, G_0) \quad \text{for each } j, \quad j \in \{1, \dots, M\} \quad (3)$$

Here, the hierarchical Dirichlet process is represented using the stick-breaking construction [19,20]. The global measure G_0 is distributed according to $DP(\gamma, H)$ and it can be expressed as

$$G_0 = \sum_{k=1}^{\infty} \psi_k \delta_{\Omega_k} \quad \psi_k = \psi'_k \prod_{s=1}^{k-1} (1 - \psi'_s) \quad \psi'_s \sim \text{Beta}(1, \gamma) \quad \Omega_k \sim H \quad (4)$$

where δ_{Ω_k} represents an atom concentrated at Ω_k , and $\{\Omega_k\}$ is a set of independent random variables drawn from H . The variable $\{\psi_k\}$ are denoted as stick-breaking that verify $\sum_{k=1}^{\infty} \psi_k = 1$. As G_0 is defined as the base distribution of the DP G_j and has the stick-breaking representation as shown in Eq. (3), then G_j includes all the atoms Ω_k with distinct weights (by following the property of Dirichlet process [18]). On the other side, we carry out another stick-breaking process to construct each group-level DP G_j according to [21] as

$$G_j = \sum_{t=1}^{\infty} \pi_{jt} \delta_{\omega_{jt}} \quad \pi_{jt} = \pi'_{jt} \prod_{s=1}^{t-1} (1 - \pi'_{js}) \quad \pi'_{js} \sim \text{Beta}(1, \lambda) \quad \omega_{jt} \sim G_0 \quad (5)$$

where $\{\pi_{jt}\}$ is a set of stick-breaking weights which shall be positive and sum to one and $\delta_{\omega_{jt}}$ are group-level Dirac delta atoms at ω_{jt} . As ω_{jt} (group-level atom) is distributed according to G_0 , thus it will take onto Ω_k (base-level atoms) with probability ψ_k .

Next, we introduce a latent indicator W_{jtk} as an indicator variable, such that $W_{jtk} \in \{0, 1\}$ (in order to indicate which group-level atom maps to). $W_{jtk} = 1$ if ω_{jt} (group-level atom) maps to the Ω_k (global-level atom) that is indexed by k ; otherwise, $W_{jtk} = 0$. Accordingly, we can have $\omega_{jt} = \sum_k \Omega_k^{W_{jtk}}$. By this way, there is no need to keep a representation for ω_{jt} . The indicator variable $\vec{W} = (W_{j11}, W_{j12}, \dots)$ is distributed as:

$$p(\vec{W} | \vec{\psi}) = \prod_{j=1}^M \prod_{t=1}^{\infty} \prod_{k=1}^{\infty} \psi_k^{W_{jtk}} \quad (6)$$

Given that ψ is a function of ψ' according to Eq. (4), it is possible to rewrite the indicator variable $p(\vec{W})$ as:

$$p(\vec{W} | \vec{\psi}') = \prod_{j=1}^M \prod_{t=1}^{\infty} \prod_{k=1}^{\infty} \left[\psi'_k \prod_{s=1}^{k-1} (1 - \psi'_s) \right]^{W_{jtk}} \tag{7}$$

According to Eq. (4) the stick lengths ψ' are drawn from a specific Beta distribution and their realization is determined as

$$p(\vec{\psi}') = \prod_{k=1}^{\infty} \text{Beta}(1, \gamma_k) = \prod_{k=1}^{\infty} \gamma_k (1 - \psi'_k)^{\gamma_k - 1} \tag{8}$$

To complete the description of the HDP mixture model, given a grouped observation (data) \mathcal{Y} , we associate each data point \vec{Y}_{ji} with a factor θ_{ji} (here i indexes the data in each group j) such that \vec{Y}_{ji} and $\vec{\theta}_j = (\theta_{j1}, \theta_{j2}, \dots)$ are distributed according to $F(\theta_{ji})$ and G_j , respectively. In this case the likelihood function can be written as:

$$\theta_{ji} | G_j \sim G_j \quad Y_{ji} | \theta_{ji} \sim F(\theta_{ji}) \tag{9}$$

where $F(\theta_{ji})$ is the probability distribution of Y_{ji} given θ_{ji} . The base distribution H provides the prior distribution for θ_{ji} . This setting (i.e., hierarchical Dirichlet process (HDP) mixture model) plays an important role and ensures that each group is associated with a mixture model, and the components of the mixture are shared across different groups.

As each θ_{ji} is distributed according to G_j (see Eq. (9)), it takes the value ω_{jt} with probability π_{jt} . We then introduce another indicator variable $Z_{jit} \in \{0, 1\}$ for θ_{ji} as

$$p(\vec{Z} | \vec{\pi}) = \prod_{j=1}^M \prod_{i=1}^N \prod_{t=1}^{\infty} \pi_{jt}^{Z_{jit}} \tag{10}$$

That is, the indicator Z_{jit} is used to indicate which component θ_{ji} belongs to. In particular, Z_{jit} is equal to 1 if θ_{ji} is associated with component t (also maps to the group-level atom ω_{jt}); else, $Z_{jit} = 0$. Thus, we can write $\theta_{ji} = \omega_{jt}^{Z_{jit}}$. As ω_{jt} maps to Ω_k , we then can write $\theta_{ji} = \omega_{jt}^{Z_{jit}} = \Omega_k^{W_{jtk} Z_{jit}}$.

Based on the stick-breaking construction of the DP (see Eq. (5)), we notice that $\vec{\pi}$ is a function of $\vec{\pi}'$, therefore the previous equation becomes

$$p(\vec{Z} | \vec{\pi}') = \prod_{j=1}^M \prod_{i=1}^N \prod_{t=1}^{\infty} \left[\pi'_{jt} \prod_{s=1}^{t-1} (1 - \pi'_{js}) \right]^{Z_{jit}} \tag{11}$$

Finally, according to Eq. (5), the prior distribution of π' is a Beta and it is given as follows:

$$p(\vec{\pi}') = \prod_{j=1}^M \prod_{t=1}^{\infty} \text{Beta}(1, \lambda_{jt}) = \prod_{j=1}^M \prod_{k=1}^{\infty} \lambda_{jt} (1 - \pi'_{jt})^{\lambda_{jt} - 1} \tag{12}$$

2.3 Hierarchical Pitman-Yor Process Mixture Model

The Pitman-Yor process (PYP) [22] is a two-parameter extension to the DP (i.e., is a generalization of DP) that permits modelling heavier-tailed distributions. It can be applied to build hierarchical models. It offers a sophisticated way to cluster data such that the number of clusters is unknown. It is characterized by an additional discount parameter γ_a in addition to the concentration parameter γ_b , that satisfying the $0 < \gamma_a < 1, \gamma_b > -\gamma_a$. Similar to DP, the sample drawn from PYP also associated a probability measure H [23]. Here, a hierarchical Pitman–Yor process (HPYP) is introduced where the base measure for a PYP is itself a draw from a PYP. Specifically, HPYP defines the global-level measure G_0 and group-level distribution G_j (that is the indexed G_j shares a same base G_0 which itself follows a PYP). This behavior makes the HPYP especially suitable for complex visual data modeling and classification. We can use the HPYP to cluster data by applying the stick-breaking construction that defines the base measure as follows:

$$G_0 = \sum_{k=1}^{\infty} \eta_k \delta_{\Lambda_k} \quad \eta_k = \eta'_k \prod_{l=1}^{k-1} (1 - \eta'_l) \quad \eta'_k \sim \text{Beta}(1 - \gamma_a, \gamma_b + k\gamma_a) \quad \Lambda_k \sim H \quad (13)$$

where $\{\Lambda_k\}$ is a set of independent random variables drawn from H and δ_{Λ_k} is an atom (probability mass) at Λ_k . The random variables η_k represent the stick-breaking weights that satisfying $\sum_{k=1}^{\infty} \eta_k = 1$.

The stick-breaking representation for the group-level PYP G_j is expressed as:

$$G_j = \sum_{t=1}^{\infty} p_{jt} \delta_{\psi_{jt}} \quad p_{jt} = p'_{jt} \prod_{s=1}^{t-1} (1 - p'_{js}) \quad p'_{js} \sim \text{Beta}(1 - \beta_a, \beta_b + t\beta_a) \quad \psi_{jt} \sim G_0 \quad (14)$$

where $\{p_{jt}\}$ are the stick-breaking weights ($\sum_{t=1}^{\infty} p_{jt} = 1$). ψ_{jt} is the atom of second-level PYP that is distributed according to G_0 . Then, a global-level indicator variables I and a group-level indicator variables C are introduced. Here, C is used to map θ_{ji} to group-level atom ψ_{jt} and the indicator I is used to map the atom θ_{ji} to base-level atom Λ_k .

$$p(\vec{I} | \vec{\eta}') = \prod_{j=1}^M \prod_{t=1}^{\infty} \prod_{k=1}^{\infty} \eta_k^{I_{jtk}} = \prod_{j=1}^M \prod_{t=1}^T \prod_{k=1}^K \left[\eta'_k \prod_{s=1}^{k-1} (1 - \eta'_s) \right]^{I_{jtk}} \quad (15)$$

$$p(\vec{C} | \vec{p}') = \prod_{j=1}^M \prod_{i=1}^N \prod_{t=1}^{\infty} p_{jt}^{C_{jit}} = \prod_{j=1}^M \prod_{i=1}^N \prod_{t=1}^T \left[p'_{jt} \prod_{s=1}^{t-1} (1 - p'_{js}) \right]^{C_{jit}} \quad (16)$$

2.4 Hierarchical Infinite Gamma Mixture Model

In this subsection, we introduce two hierarchical infinite mixture models with Gamma distributions. In this case, each vector $\vec{Y}_{ji} = (Y_{ji1}, \dots, Y_{jiD})$, from the grouped data, is drawn from a hierarchical infinite Gamma mixture model. Then, the two likelihood functions of these hierarchical models given the unknown parameters of Gamma and latent variables can be expressed as follows:

$$p(\vec{Y} | \vec{Z}, \vec{W}, \vec{\theta}) = p(\vec{Y} | \vec{Z}, \vec{W}, \vec{\alpha}, \vec{\beta}) = \prod_{j=1}^M \prod_{i=1}^N \prod_{t=1}^{\infty} \prod_{k=1}^{\infty} Ga(\vec{Y}_{ji} | \vec{\theta}_k)^{Z_{jit} W_{jtk}}$$

$$p(\vec{Y} | \vec{C}, \vec{I}, \vec{\theta}) = p(\vec{Y} | \vec{C}, \vec{I}, \vec{\alpha}, \vec{\beta}) = \prod_{j=1}^M \prod_{i=1}^N \prod_{t=1}^{\infty} \prod_{k=1}^{\infty} Ga(\vec{Y}_{ji} | \vec{\theta}_k)^{C_{jit} I_{jtk}} \tag{17}$$

where $\vec{\theta}_k = (\vec{\alpha}_k, \vec{\beta}_k)$ are the full parameters of the Gamma distribution. The above two functions are HDP Gamma mixture (denoted by HDPGaM) and HPYP Gamma mixture (denoted by HPYPGaM), respectively.

Next, we have to place conjugate distributions over the unknown parameters α and β . As α and β are positive, then it is convenient that they follow Gamma distributions $\mathcal{G}(\cdot)$. Thus, we have

$$p(\vec{\alpha}) = \mathcal{G}(\vec{\alpha} | \vec{u}, \vec{v}) = \prod_{k=1}^{\infty} \prod_{l=1}^D \mathcal{G}(\alpha_{jd} | u_{jd}, v_{jd}) = \prod_{k=1}^{\infty} \prod_{l=1}^D \frac{v_{kl}^{u_{kl}}}{\Gamma(u_{kl})} \alpha_{kl}^{u_{kl}-1} e^{-v_{kl} \alpha_{kl}} \tag{18}$$

$$p(\vec{\beta}) = \mathcal{G}(\vec{\beta} | \vec{g}, \vec{h}) = \prod_{k=1}^{\infty} \prod_{l=1}^D \mathcal{G}(\beta_{jd} | g_{jd}, h_{jd}) = \prod_{k=1}^{\infty} \prod_{l=1}^D \frac{h_{kl}^{g_{kl}}}{\Gamma(g_{kl})} \beta_{kl}^{g_{kl}-1} e^{-h_{kl} \beta_{kl}} \tag{19}$$

3 Model Learning via Variational Bayes

Variational inference [3,24] is a well-defined method deterministic approximation method that is used in order to approximate posterior probability via an optimization process. In this section, we propose to develop a variational learning framework of our hierarchical infinite Gamma mixture models. Here, $\Theta = \{Z, W, \vec{\psi}', \vec{\pi}', \vec{\theta}\}$ represents both unknown and the latent variables. Our objective is to estimate a suitable approximation $q(\Theta)$ for the true posterior distribution $p(\Theta | \mathcal{Y})$ via a process of maximizing the lower bound of $\ln p(\mathcal{Y})$ given as

$$\mathcal{L}(Q) = \int Q(\Theta) \ln \left(\frac{p(\mathcal{Y}, \Theta)}{Q(\Theta)} \right) d\Theta \tag{20}$$

In particular, we adopt one of the most successfully variational inference techniques namely the factorial approximation (or mean fields approximation) [3], which is able to offer effective updates. Thus, we apply this method to fully factorize $q(\Theta)$ of on HDPGaM and HPYPGaM mixtures into disjoint factors. Then, we apply a truncation method as previously applied in [20] to truncate the variational approximations into global truncation level K and group truncation level T as follow:

$$\psi'_K = 1, \quad \sum_{k=1}^K \psi_k = 1, \quad \psi_k = 0 \quad \text{when } k > K \tag{21}$$

$$\eta'_K = 1, \quad \sum_{k=1}^K \eta_k = 1, \quad \eta_k = 0 \quad \text{when } k > K \tag{22}$$

$$\pi'_{jT} = 1, \quad \sum_{t=1}^T \pi_{jt} = 1, \quad \pi_{jt} = 0 \quad \text{when } t > T \tag{23}$$

$$p'_{jT} = 1, \quad \sum_{t=1}^T p_{jt} = 1, \quad p_{jt} = 0 \quad \text{when } t > T \tag{24}$$

where the truncation levels K and T will be optimized over the learning procedure. The approximated posterior distribution is then factorized as

$$q(\Theta) = q(\vec{W}, \vec{\psi}', \vec{Z}, \vec{\pi}', \vec{\alpha}, \vec{\beta}) = q(\vec{W})q(\vec{\psi}')q(\vec{Z})q(\vec{\pi}')q(\vec{\alpha})q(\vec{\beta})$$

$$q(\Theta) = q(\vec{I}, \vec{\eta}', \vec{C}, \vec{p}', \vec{\alpha}, \vec{\beta}) = q(\vec{I})q(\vec{\eta}')q(\vec{C})q(\vec{p}')q(\vec{\alpha})q(\vec{\beta}) \quad (25)$$

For a specific variational factor $q_s(\Theta_s)$, the general equation of the optimal solution is expressed as [3]:

$$q_s(\Theta_s) = \frac{\exp\langle \log p(\mathcal{Y}, \Theta) \rangle_{i \neq s}}{\int \exp\langle \log p(\mathcal{Y}, \Theta) \rangle_{i \neq s} d\Theta} \quad (26)$$

where $\langle \cdot \rangle_{i \neq s}$ denotes an expectation with respect to all the distributions of $q_i(\Theta_i)$ except for $i = s$. The parametric forms for the variational posteriors (for each factor) are determined on the basis of Eq. (26) as

$$q(\vec{W}) = \prod_{j=1}^M \prod_{t=1}^T \prod_{k=1}^K \vartheta_{jtk}^{W_{jtk}} \quad (27)$$

$$q(\vec{Z}) = \prod_{i=1}^N \prod_{j=1}^M \prod_{t=1}^T \rho_{ijt}^{Z_{ijt}} \quad (28)$$

$$q(\vec{\psi}') = \prod_{k=1}^K \text{Beta}(\psi'_k | c_k, d_k) \quad (29)$$

$$q(\vec{\pi}') = \prod_{j=1}^M \prod_{t=1}^T \text{Beta}(\pi'_{jt} | a_{jt}, b_{jt}) \quad (30)$$

$$q(\vec{I}) = \prod_{j=1}^M \prod_{t=1}^T \prod_{k=1}^K \varphi_{jtk}^{I_{jtk}} \quad (31)$$

$$q(\vec{C}) = \prod_{i=1}^N \prod_{j=1}^M \prod_{t=1}^T \varrho_{ijt}^{C_{ijt}} \quad (32)$$

$$q(\vec{\eta}') = \prod_{k=1}^K \text{Beta}(\eta'_k | e_k, f_k) \quad (33)$$

$$q(\vec{p}') = \prod_{j=1}^M \prod_{t=1}^T \text{Beta}(p'_{jt} | m_{jt}, n_{jt}) \quad (34)$$

$$q(\vec{\alpha}) = \prod_{k=1}^K \prod_{d=1}^D \mathcal{G}(\alpha_{kd} | u_{kd}^*, v_{kd}^*) \quad (35)$$

$$q(\vec{\beta}) = \prod_{k=1}^K \prod_{d=1}^D \mathcal{G}(\beta_{kd} \mid s_{kd}^*, t_{kd}^*) \quad (36)$$

where the corresponding hyperparameters in the above equations can be calculated as a similar way in [25,26]. The complete variational Bayes inference algorithms of both HDPGaM and HPYPGaM are summarized in Algorithm 1 and Algorithm 2, respectively.

Algorithm 1: HDPGaM: Proposed Hierarchical Dirichlet process gamma mixture algorithm

- 1 Pick out initial truncation levels K and T .
 - 2 Set initial values for hyperparameters u, v, s, t, a, b, c, d .
 - 3 **repeat**
 - 4 **Variational E-step :**
 - 5 Estimate the expected values following the same strategy in [26].
 - 6 **Variational M-step :**
 - 7 Update the variational solution for the factor $Q(Z)$ using Eq. (28).
 - 8 Update the variational solution for the factor $Q(\psi)$ using Eq. (29).
 - 9 Update the variational solution for the factor $Q(\alpha)$ using Eq. (35).
 - 10 Update the variational solution for the factor $Q(\beta)$ using Eq. (36).
 - 11 **until** *Until convergence is reached*
 - 12 The convergence is achieved when the difference between the current value and previous values is less than 10^{-2} or the epochs number exceeds 300.
 - 13 Return the optimal number of components by eliminating the components with small mixing coefficients close to 0.
-

Algorithm 2: HPYPGaM: Proposed Hierarchical Pitman-Yor process gamma mixture algorithm

- 1 Pick out initial truncation levels K and T .
 - 2 Set initial values for hyperparameters u, v, s, t, e, f, m, n .
 - 3 **repeat**
 - 4 **Variational E-step :**
 - 5 Estimate the expected values following the same strategy in [25].
 - 6 **Variational M-step :**
 - 7 Update the variational solution for the factor $Q(C)$ using Eq. (32).
 - 8 Update the variational solution for the factor $Q(\eta)$ using Eq. (33).
 - 9 Update the variational solution for the factor $Q(\alpha)$ using Eq. (35).
 - 10 Update the variational solution for the factor $Q(\beta)$ using Eq. (36).
 - 11 **until** *Until convergence is reached*
 - 12 The convergence is achieved when the difference between the current value and previous values is less than 10^{-2} or the epochs number exceeds 300.
 - 13 Return the optimal number of components by eliminating the components with small mixing coefficients close to 0.
-

4 Experimental Results

The principal purpose of the experiment section is to investigate the performance of the developed two frameworks based on HDP mixture and HPYP mixture model with Gamma distributions. Hence, we propose to compare them with other statistical models using two challenging applications: Texture categorization and human action categorization. In all these experiments, the global truncation level K and the group level truncation level T are both initialized to 120 and 60, respectively. For HDP mixture, we set the hyperparameters of the stick lengths γ and λ as (0.25, 0.25). The parameters of HPYP mixture $\gamma_a, \gamma_b, \beta_a$ and β_b are initialized to (0.5, 0.25, 0.5, 0.25). The hyperparameters of Gamma base distribution are initialized by sampling from priors.

4.1 Texture Classification

In this work we are primarily motivated by the problem of modeling and classifying texture images. Contrary to natural images which include certain objects and structures, texture images are very special case of images that do not include a well-defined shape. Texture pattern is one of the most important elements in visual multimedia content analysis. It forms the basis for solving complex machine learning and computer vision tasks. In particular, texture classification supports a wide range of applications, including information retrieval, image categorization [27–30], image segmentation [27,31,32], material classification [33], facial expression recognition [29], and object detection [28,34]. The goal here to classify texture images using the two hierarchical infinite mixtures and also by incorporating three different representations (to extract relevant features from images) from the literature, namely Local Binary Pattern (LBP) [35], Local Binary Pattern (LBP) [35], scale-invariant feature transform (SIFT) [36], and dense micro-block difference (DMD) [37]. A deep review for these methods is outside the scope of current work. Instead, we focus on some powerful feature extraction methods that have shown interesting state-of-the-art results.

4.1.1 Methodology

For this application, we start by extracting features from input images and then we model them using the proposed HDPGaM and HPYPGaM. Each image I_j is considered as a group and is related to a infinite mixture model G_j . Next, every vector Y_{ji} of I_j is assumed generated from G_j , where G_j represents visual words. The next step is to generate a global vocabulary to share it among all groups via G_0 (global infinite model). It is noteworthy that the the building of the visual vocabulary, here, is part of our hierarchical models and therefore, the size of the vocabulary (i.e., number of components) is inferred automatically from the data thanks to the characteristic of nonparametric Bayesian models. Regarding SIFT features, the bag-of-visual-words model is adopted here to calculate the histogram of visual words from each input image. Regarding the set of DMD descriptors, these are obtained after extracting DMD features and then encoding them though th Fisher vector method as proposed in [37]. The resulting descriptors are able to attain good discrimination thanks to their invariance with respect to scale, resolution, and orientation. Finally, each image is represented by a multidimensional vector of high-order statistics¹.

4.1.2 Dataset and Results

We conducted our experiments of texture classification using the proposed hierarchical HDP Gamma mixture (referred to as HDPGaM) and HPYP Gamma mixture (referred to as HPYPGaM) on three publicly available databases. The first one namely UIUCTex [38] contains 25 texture classes and each class has 40 images. The second dataset namely UMD [39] contains 25 textures classes containing each one 40 images. The third dataset namely KTH-TIPS [33] includes 10 classes and each one contains 81 images. Some sample texture images from each class and each dataset are shown in Fig. 2. We use 10-fold cross-validation technique to partition these databases and to study the performance. In addition, the evaluation process and the obtained results are based on 30 runs.

In order to quantify the performance of the proposed frameworks (HDPGaM and HPYPGaM), we proceed by evaluating and comparing the obtained results with seven other methods namely infinite mixture of Gaussian distribution (inGM), infinite mixture of generalized Gaussian distribution (inGGM), infinite mixture of multivariate generalized Gaussian distribution (inMGGM),

¹The Matlab code for the features is available at <http://www.cs.tut.fi/~x00A0/mehta/texturedmd>.

Hierarchical Dirichlet Process mixture of Gaussian distribution (HDPGM), hierarchical Pitman-Yor process mixture of Gaussian distribution (HPYPGM), Hierarchical Dirichlet Process mixture of generalized Gaussian distribution (HDPGGM), and hierarchical Pitman-Yor process mixture of generalized Gaussian distribution (HPYPGGM).

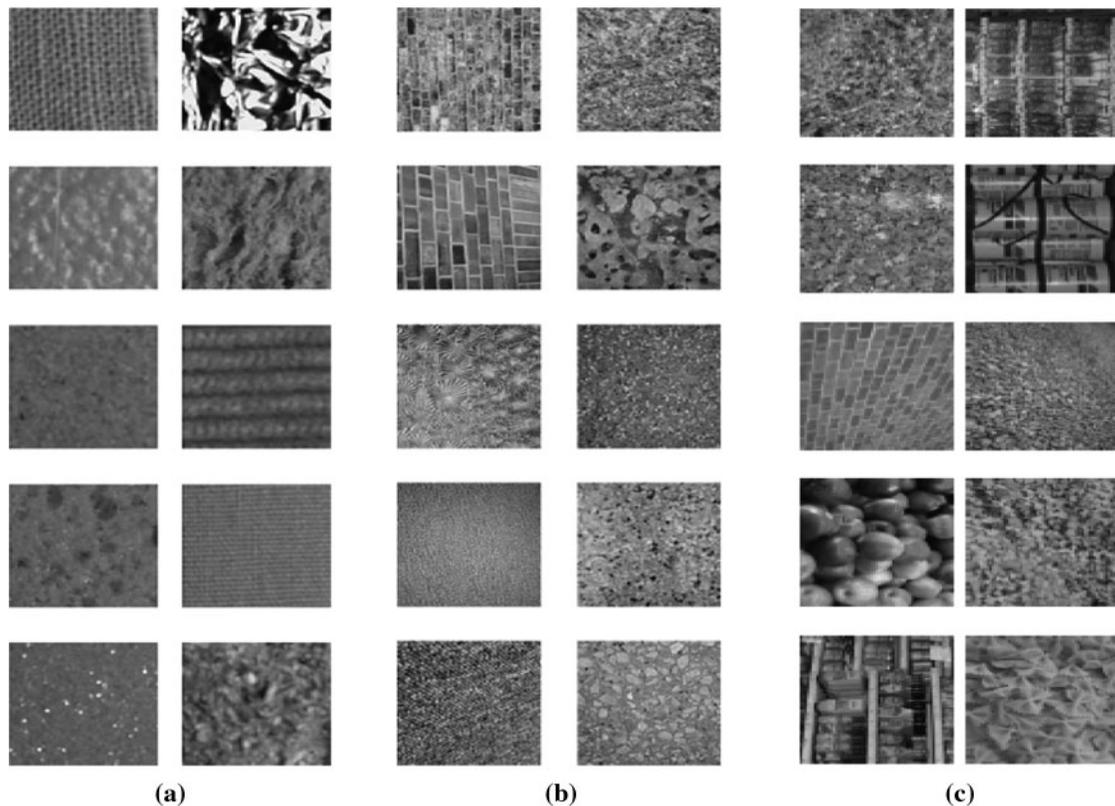


Figure 2: Texture samples in different categories for different datasets (a) KTH-TIPS, (b) UIUCTex, and (c) UMD dataset

We run all methods 30 times and calculate the average classification accuracy which are depicted in [Tabs. 1–3](#) respectively. According to these results, we can notice that HDPGaM and HPYPGaM have the highest achieved accuracy for the three databases in terms of the texture classification accuracy rate. It is noted that when comparing these results by considering the Student's t-test, the differences in performance are statistically significant between our frameworks and the rest of methods. In particular, results indicate the benefits of our proposed models in terms of texture modeling and classification capabilities which surpass those obtained by HDPGM, HDPGGM, HPYPGM, and HPYPGGM. By contrast, the worst performance is obtained within the infinite Gaussian mixture models. It should be noted that the proposed frameworks outperform the other methods and that the three adopted feature extraction methods (SIFT, LBP and DMD). Thus, these results confirm the merit of the proposed methods. Due to the effectiveness of DMD descriptor for describing and modelling observed texture images, we also find that DMD achieves better accuracy compared with both SIFT and LBP. It shows the merits of DMD which is able to consider all possible fine details images at different resolutions. We also note that with HPYP mixture we can reach better results compared to HDP mixture and this is for all tested distributions. This can be explained by the fact that HPYP mixture model has better

generalization capability and better capacity to model heavier-tailed distribution (PYP prior can lead to better modeling ability).

Table 1: The average accuracy results of texture classification using different algorithms for the 3 texture-datasets using SIFT features

Method	KTH-TIPS	UIUCTex	UMD
inGM	80.94	85.33	84.14
inGGM	83.30	87.21	87.17
inMGGM	85.91	89.03	89.10
HDPGM	85.92	89.10	89.17
HPYPGM	85.95	89.13	89.20
HDPGGM	85.95	89.21	89.26
HPYPGGM	85.97	89.26	89.28
HDPGaM (our method)	86.03	89.77	89.62
HPYPGaM (our method)	86.12	89.82	89.70

Table 2: The average accuracy results of texture classification using different algorithms for the 3 texture-datasets using LBP features

Method	KTH-TIPS	UIUCTex	UMD
inGM	81.24	86.03	84.84
inGGM	83.94	87.84	87.86
inMGGM	86.29	89.81	89.97
HDPGM	86.32	89.85	90.01
HPYPGM	86.35	89.88	90.05
HDPGGM	86.41	89.93	90.07
HPYPGGM	86.44	89.98	90.09
HDPGaM (our method)	86.46	90.05	90.13
HPYPGaM (our method)	86.50	90.15	90.18

Table 3: The average accuracy results of texture classification using different algorithms for the 3 texture-datasets using DMD features

Method	KTH-TIPS	UIUCTex	UMD
inGMM	84.13	89.77	87.12
inGGMM	86.88	90.67	90.20
inMGGM	88.91	92.11	92.12
HDPGM	88.93	92.11	92.11
HPYPGM	88.96	92.13	92.17
HDPGGM	88.99	92.15	92.19
HPYPGGM	88.99	92.17	92.20

(Continued)

Table 3: Continued

Method	KTH-TIPS	UIUCTex	UMD
HDPGaM (our method)	88.99	92.19	92.21
HPYPGaM (our method)	89.05	92.23	92.30

4.2 Human Actions Categorization

Visual multimedia recognition has been a challenging research topic which could attract many applications such as actions recognition [40,41], image categorization [42,43], biomedical image recognition [44], and facial expressions [29,30]. In this work, we are focusing on a particular problem that has received a lot of attention namely Human actions recognition (HAR) through sequence of videos. Indeed, the intention of recognizing activities is to identify and analyze various human actions. At present, HAR is one the hot computer vision topics not only in research but also in industries where automatic identification of any activity can be useful, for instance, for monitoring, healthcare, robotics, and security-based applications [45]. Recognizing manually activities is very challenging and time consuming. This issue has been addressed and so various tools have been implemented such as in [40,45–47]. However, precise recognition of actions is still required using advanced and efficient algorithms in order to deal with complex situations such as noise, occlusions, and lighting.

We perform here the recognition of Human activities using the proposed frameworks HDPGaM and HPYPGaM. Our methodology is outlined as following: First, we extract and normalize SIFT3D descriptor [40] from observed images. These features are then quantized as visual words via bag-of-words (BOW) model and K-means algorithm [48]. Then, these features are quantized as visual words via K-means algorithm [48]. Then, a probabilistic Latent Semantic Analysis (pLSA) [49] is adopted to construct a d -dimensional vector. In particular, each image I_j is considered as a “group” and is associated with an infinite mixture model G_j . Thus, we suppose that Each SIFT3D feature vector is drawn from the infinite mixture model G_j and “visual words” denote mixture components of G_j .

On the other hand, a global vocabulary is generated and shared between all groups via the global-model G_0 of the proposed hierarchical models. This setting agrees with the purpose behind the hierarchical process mixture model. It is also noted that the building of the visual vocabulary is part of the hierarchical process mixture models and this step is not carried out separately via the k-means algorithm as many other approaches do. It is also noted that the generation of the visual vocabulary is part of our hierarchical process mixture models and this process is not performed separately via k-means as many other approaches did. In fact, it is due to the property of the nonparametric Bayesian model that the number of components in the global-level mixture model can be deduced from the data. We conducted our experiments of actions recognition using a publicly dataset known as KTH human action dataset [46]. This database contains 2391 sequences of different actions grouped into 6 classes. It also represents four scenarios (outdoors (s1), outdoors with scale variation (s2), outdoors with different clothes (s3) and indoors (s4)). Some scenarios from this dataset are given in Fig. 3. We randomly divided this dataset into 2 subsets to train the developed frameworks and to evaluate its robustness.

Our purpose through this application is to show the advantages of investigating our proposed hierarchical models HDPGaM and HPYPGaM over other conventional hierarchical mixtures and other methods from the state of the art. Therefore, we focused first on evaluating the performance of HDPGaM and HPYPGaM over Hierarchical Dirichlet Process mixture of Gaussian distribution

(HDPGM), hierarchical Pitman-Yor process mixture of Gaussian distribution (HPYPGM), Hierarchical Dirichlet Process mixture of generalized Gaussian distribution (HDPGGM), and hierarchical Pitman-Yor process mixture of generalized Gaussian distribution (HPYPGGM). It is noted that we learned all the implemented models using variational Bayes. The average recognition performances of our frameworks and models based on HDP mixture and HPYP mixture are depicted in [Tab. 4](#).

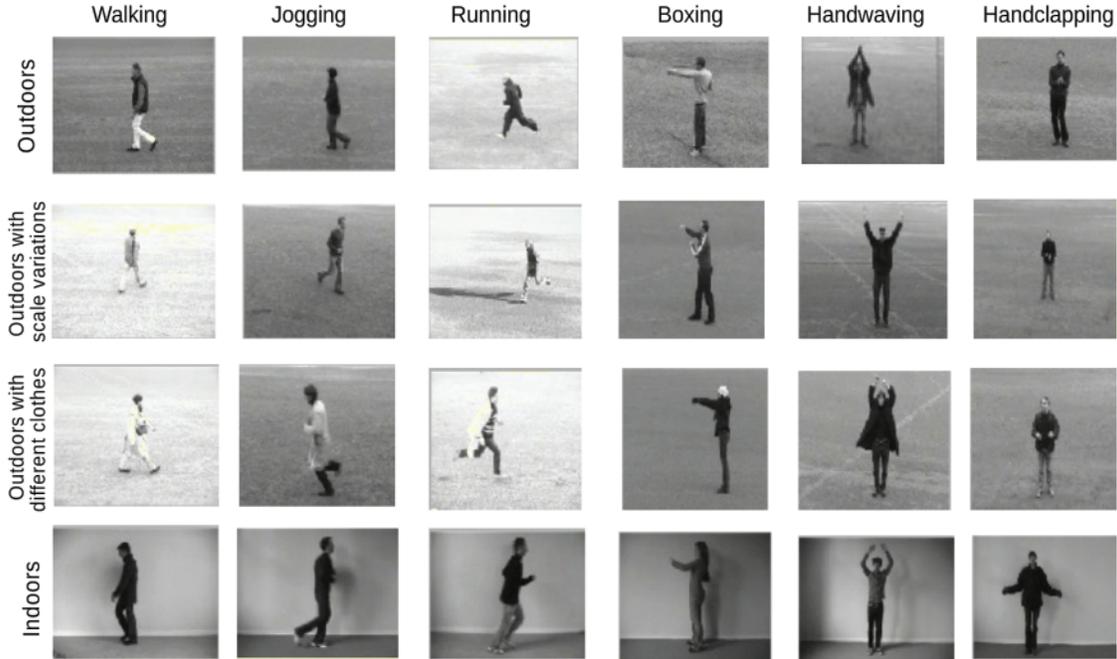


Figure 3: Sample frames of the KTH dataset actions with different scenarios

Table 4: Average recognition performance (%) obtained using our frameworks and other models based on HDP mixture and HPYP mixture for KTH database

Method	Recognition rate (%)
HDPGM	80.12
HPYPGM	80.19
HDPGGM	81.33
HPYPGGM	81.39
HDPGaM (our method)	82.13
HPYPGaM (our method)	82.27

As we can see in this table, the proposed frameworks were able to offer the highest recognition rates (82.27% for HPYPGaM and 82.13% for HDPGaM) among all tested models. For different runs, we have p-values < 0.05 and therefore, the differences in accuracy between our frameworks and other models are statistically significant according to Student's t-test. Next, we compared our models against other mixture models (here finite Gaussian mixture (GMM) and finite generalized Gaussian mixture (GGMM)) and methods from the literature. The obtained results are given in [Tab. 5](#).

Table 5: Average recognition performance (%) obtained using our frameworks and other methods from the literature for KTH database

Method	Recognition rate (%)
GMM	72.51
GGMM	73.89
SVM-linear	73.24
SVM-polynomial	71.44
SVM-radial basis	79.33
Wong et al. [50]	73.24
Fan et al. [51]	79.33
Schuldt et al. [46]	71.71
Dollár et al. [52]	78.50
MGGMM-FS [47]	82.01
HDPGaM (our method)	82.13
HPYPGaM (our method)	82.27

Accordingly, we can observe that models again are able to provide higher discrimination rate than the other methods. Clearly, these results confirm the effectiveness of our frameworks for activities modeling and recognition compared to other conventional Dirichlet and Pitman-Yor processes based on Gaussian distribution. Another remark is that our model HPYPGaM outperforms our second model HDPGaM for this specific application and this demonstrates the advantages of using hierarchical Pitman-Yor process over Dirichlet process which is flexible enough to be used for such recognition problem.

5 Conclusions

In this paper two non-parametric Bayesian frameworks based on both hierarchical Dirichlet and Pitman-Yor processes and Gamma distribution are proposed. The Gamma distribution is considered because of its flexibility for semi-bounded data modelling. Both frameworks are learned using variational inference which has certain advantages such as easy assessment of convergence and easy optimization by offering a trade-off between frequentist techniques and MCMC-based ones. An important property of our approach is that it does not need the specification of the number of mixture components in advance. We carried out experiments on texture categorization and human action recognition to demonstrate the performance of our models which can be used further for a variety of other computer vision and pattern recognition applications.

Acknowledgement: The authors would like to thank Taif University Researchers Supporting Project number (TURSP-2020/26), Taif University, Taif, Saudi Arabia. They would like also to thank Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R40), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Funding Statement: The authors would like to thank Taif University Researchers Supporting Project number (TURSP-2020/26), Taif University, Taif, Saudi Arabia. They would like also to thank Princess

Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R40), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] J. L. Andrews, P. D. McNicholas and S. Subedi, "Model-based classification via mixtures of multivariate t-distributions," *Computational Statistics & Data Analysis*, vol. 55, no. 1, pp. 520–529, 2011.
- [2] J. W. Lau and P. J. Green, "Bayesian model-based clustering procedures," *Journal of Computational and Graphical Statistics*, vol. 16, no. 3, pp. 526–558, 2007.
- [3] T. Singh, N. Saxena, M. Khurana, D. Singh, M. Abdalla *et al.*, "Data clustering using moth-flame optimization algorithm," *Sensors*, vol. 21, no. 12, pp. 4086, 2021.
- [4] W. Fan, N. Bouguila, J. -X. Du and X. Liu, "Axially symmetric data clustering through Dirichlet process mixture models of watson distributions," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 6, pp. 1683–1694, 2019.
- [5] P. D. McNicholas, "Model-based clustering," *Journal of Classification*, vol. 33, no. 3, pp. 331–373, 2016.
- [6] X. Liu, H. Fu and Y. Jia, "Gaussian mixture modeling and learning of neighboring characters for multilingual text extraction in images," *Pattern Recognition*, vol. 41, no. 2, pp. 484–493, 2008.
- [7] V. Kumar, D. Kumar, M. Kaur, D. Singh, S. A. Idris *et al.*, "A novel binary seagull optimizer and its application to feature selection problem," *IEEE Access*, vol. 9, no. 1, pp. 103481–103496, 2021.
- [8] A. Aggarwal, V. Sharma, A. Trivedi, M. Yadav, C. Agrawal *et al.*, "Two-way feature extraction using sequential and multimodal approach for hateful meme classification," *Complexity*, vol. 2021, no. 1, pp. 1–7, 2021.
- [9] C. Beckmann, M. Woolrich and S. Smith, "Gaussian/gamma mixture modelling of ICA/GLM spatial maps," *Ninth International Conference on Functional Mapping of the Human Brain*, New York, pp. 18–22, 2003; *NeuroImage*, vol. 19, pp. S935, 2003.
- [10] A. Almulihi, F. Alharithi, S. Bourouis, R. Alroobaea, Y. Pawar *et al.*, "Oil spill detection in SAR images using online extended variational learning of Dirichlet process mixtures of gamma distributions," *Remote Sensing*, vol. 13, no. 15, pp. 2991, 2021.
- [11] 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 Transaction on Geoscience Remote Sensing*, vol. 54, no. 4, pp. 2153–2170, 2016.
- [12] S. Bourouis, H. Sallay and N. Bouguila, "A competitive generalized gamma mixture model for medical image diagnosis," *IEEE Access*, vol. 9, pp. 13727–13736, 2021.
- [13] R. Alroobaea, S. Rubaiee, S. Bourouis, N. Bouguila and A. Alsufyani, "Bayesian inference framework for bounded generalized Gaussian-based mixture model and its application to biomedical images classification," *International Journal of Imaging Systems and Technology*, vol. 30, no. 1, pp. 18–30, 2020.
- [14] S. Bourouis, R. Alroobaea, S. Rubaiee, M. Andejany, F. M. Almansour *et al.*, "Markov chain monte carlo-based Bayesian inference for learning finite and infinite inverted beta-liouville mixture models," *IEEE Access*, vol. 9, no. 1, pp. 71170–71183, 2021.
- [15] S. N. MacEachern and P. Müller, "Estimating mixture of dirichlet process models," *Journal of Computational and Graphical Statistics*, vol. 7, no. 2, pp. 223–238, 1998.
- [16] W. Fan, H. Sallay and N. Bouguila, "Online learning of hierarchical pitman-yor process mixture of generalized dirichlet distributions with feature selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 9, pp. 2048–2061, 2017.
- [17] Y. W. Teh, M. I. Jordan, M. J. Beal and D. M. Blei, "Sharing clusters among related groups: Hierarchical dirichlet processes," in *Advances in Neural Information Processing Systems 17 [Neural Information Processing Systems: NIPS 2004, December 13-18]*, Vancouver, British Columbia, Canada, pp. 1385–1392, 2004.

- [18] Y. W. Teh, M. I. Jordan, M. J. Beal and D. M. Blei, "Hierarchical dirichlet processes," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1566–1581, 2006.
- [19] J. Sethuraman, "A constructive definition of dirichlet priors," *Statistica Sinica*, vol. 4, pp. 639–650, 1994.
- [20] D. M. Blei and M. I. Jordan, "Variational inference for dirichlet process mixtures," *Bayesian Analysis*, vol. 1, no. 1, pp. 121–143, 2006.
- [21] C. Wang, J. W. Paisley, and D. M. Blei, "Online variational inference for the hierarchical dirichlet process," in *Proc. of the Fourteenth Int. Conf. on Artificial Intelligence and Statistics*, Fort Lauderdale, USA, vol. 15, pp. 752–760, 2011.
- [22] H. Ishwaran and L. F. James, "Gibbs sampling methods for stick-breaking priors," *Journal of the American Statistical Association*, vol. 96, no. 453, pp. 161–173, 2001.
- [23] J. Pitman and M. Yor, "The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator," *The Annals of Probability*, vol. 25, no. 2, pp. 855–900, 1997.
- [24] H. Attias, "A variational Bayesian framework for graphical models," *Advances in Neural Information Processing Systems*, vol. 12, no. 1–2, pp. 209–215, 2000.
- [25] Z. Song, S. Ali, N. Bouguila and W. Fan, "Nonparametric hierarchical mixture models based on asymmetric Gaussian distribution," *Digit. Signal Process*, vol. 106, no. 1, pp. 102829, 2020.
- [26] W. Fan, H. Sallay, N. Bouguila and S. Bourouis, "Variational learning of hierarchical infinite generalized dirichlet mixture models and applications," *Soft Computing*, vol. 20, no. 3, pp. 979–990, 2016.
- [27] Y. Huang, F. Zhou and J. Gilles, "Empirical curvelet based fully convolutional network for supervised texture image segmentation," *Neurocomputing*, vol. 349, no. 1, pp. 31–43, 2019.
- [28] Q. Zhang, J. Lin, Y. Tao, W. Li and Y. Shi, "Salient object detection via color and texture cues," *Neurocomputing*, vol. 243, pp. 35–48, 2017.
- [29] W. Fan and N. Bouguila, "Online facial expression recognition based on finite beta-liouville mixture models," in *2013 Int. Conf. on Computer and Robot Vision*, Regina, Saskatchewan, Canada, 2013, pp. 37–44.
- [30] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.
- [31] A. Badoual, M. Unser and A. Depeursinge, "Texture-driven parametric snakes for semi-automatic image segmentation," *Computer Vision and Image Understanding*, vol. 188, no. 3, pp. 102793, 2019.
- [32] Y. Zheng and K. Chen, "A general model for multiphase texture segmentation and its applications to retinal image analysis," *Biomedical Signal Processing and Control*, vol. 8, no. 4, pp. 374–381, 2013.
- [33] E. Hayman, B. Caputo, M. Fritz and J. -O. Eklundh, "On the significance of real-world conditions for material classification," in *Computer Vision-ECCV 2004, 8th European Conf. on Computer Vision, Czech Republic, may 11–14, Proc., Part IV*, Prague, Czech Republic, 2004, vol. 3024, pp. 253–266.
- [34] J. Zhang, M. Marszalek, S. Lazebnik and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213–238, 2007.
- [35] Z. Guo, L. Zhang and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transaction on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010.
- [36] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [37] R. Mehta and K. O. Egiazarian, "Texture classification using dense micro-block difference," *IEEE Transaction on Image Processing*, vol. 25, no. 4, pp. 1604–1616, 2016.
- [38] S. Lazebnik, C. Schmid and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1265–1278, 2005.
- [39] Y. Xu, H. Ji and C. Fermüller, "Viewpoint invariant texture description using fractal analysis," *International Journal of Computer Vision*, vol. 83, no. 1, pp. 85–100, 2009.
- [40] P. Scovanner, S. Ali and M. Shah, "A 3-dimensional sift descriptor and its application to action recognition," in *Proc. of the 15th ACM Int. Conf. on Multimedia*, Augsburg, Germany, 2007, pp. 357–360.

- [41] F. Najar, S. Bourouis, N. Bouguila and S. Belghith, “Unsupervised learning of finite full covariance multivariate generalized Gaussian mixture models for human activity recognition,” *Multimedia Tools and Applications*, vol. 78, no. 13, pp. 18669–18691, 2019.
- [42] R. Zhu, F. Dornaika and Y. Ruichek, “Learning a discriminant graph-based embedding with feature selection for image categorization,” *Neural Networks*, vol. 111, pp. 35–46, 2019.
- [43] H. Zhou and S. Zhou, “Scene categorization towards urban tunnel traffic by image quality assessment,” *Journal of Visual Communication and Image Representation*, vol. 65, pp. 102655, 2019.
- [44] F. S. Alharithi, A. H. Almulihi, S. Bourouis, R. Alroobaea and N. Bouguila, “Discriminative learning approach based on flexible mixture model for medical data categorization and recognition,” *Sensors*, vol. 21, no. 7, pp. 2450, 2021.
- [45] M. Vrigkas, C. Nikou and I. A. Kakadiaris, “A review of human activity recognition methods,” *Frontiers in Robotics and AI*, vol. 2, pp. 28, 2015.
- [46] C. Schuldt, I. Laptev and B. Caputo, “Recognizing human actions: A local SVM approach,” in *Pattern Recognition, 2004. ICPR 2004. Proc. of the 17th Int. Conf.*, Cambridge, UK, 2004, vol. 3, pp. 32–36.
- [47] F. Najar, S. Bourouis, N. Bouguila and S. Belghith, “A new hybrid discriminative/generative model using the full-covariance multivariate generalized Gaussian mixture models,” *Soft Computing*, vol. 24, no. 14, pp. 10611–10628, 2020.
- [48] G. Csurka, C. Dance, L. Fan, J. Willamowski and C. Bray, “Visual categorization with bags of keypoints,” in *Workshop on Statistical Learning in Computer Vision, ECCV*, Prague, Czech Republic, 2004, vol. 1, pp. 1–2.
- [49] A. Bosch, A. Zisserman and X. Muñoz, “Scene classification via pLSA,” *Computer Vision–ECCV*, 2006, pp. 517–530, 2006.
- [50] S. -F. Wong and R. Cipolla, “Extracting spatiotemporal interest points using global information,” in *IEEE 11th Int. Conf. on Computer Vision*, Rio de Janeiro, Brazil, pp. 1–8, 2007.
- [51] W. Fan and N. Bouguila, “Variational learning for dirichlet process mixtures of dirichlet distributions and applications,” *Multimedia Tools and Applications*, vol. 70, no. 3, pp. 1685–1702, 2014.
- [52] P. Dollár, V. Rabaud, G. Cottrell and S. Belongie, “Behavior recognition via sparse spatio-temporal features,” in *2005 IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, Beijing, China, pp. 65–72, 2005.