Chapter 5
Online Recognition via a Finite Mixture of Multivariate Generalized Gaussian Distributions

Fatma Najar, Sami Bourouis, Rula Al-Azawi, and Ali Al-Badi

Abstract The huge amount of data expanding day by day entail creating powerful real-time algorithms. Such algorithms allow a reactive processing between the input multimedia data and the system. In particular, we are mainly concerned with active learning and clustering images and videos for the purpose of pattern recognition. In this paper, we propose a novel online recognition algorithm based on multivariate generalized Gaussian distributions. We estimate at first the generative model's parameters within a discriminative framework (fixed-point, Riemannian averaged fixed-point, and Fisher scoring). Then, we propose an online recognition algorithm in accordance with those algorithms. Finally, we applied our proposed framework on three challenging problems, namely: human action recognition, facial expression recognition, and pedestrian detection from infrared images. Experiments demonstrate the robustness of our approach by comparing with the state-of-the art algorithms and offline learning techniques.

5.1 Introduction

Online, real-time sequential arrival of data has increased the computer science community efforts to analyze, understand, and extract information. Despite the

F. Najar (✉)
Laboratoire RISC Robotique Informatique et Systèmes Complexes, Université de Tunis El Manar, ENIT, Tunis, Tunisie
e-mail: fatma.najjar@enit.utm.tn

S. Bourouis
Taif University, Taif, Saudi Arabia

Université de Tunis El Manar, LR-SITI Laboratoire Signal, Image et Technologies de l'Information, Tunis, Tunisie
e-mail: s.bourouis@tu.edu.sa

R. Al-Azawi · A. Al-Badi
Gulf College, Al Maabelah, Muscat, Oman
e-mail: aalbadi@gulfcollege.edu.om

fact that information is continuously changing in real time and cannot be available at once, the traditional learning approach remains constant. In fact, when data is generated in a function of time, we need to incrementally assemble data as long as they arrive in a time sequence. Besides, when the size of data is out of the memory limits, it will be computationally infeasible to train over the entire dataset. In order to meet these necessities, online learning has been emerged to deal with data in an incremental process, react to new data, and predict the future coming inputs. As the notation suggests, online learning is an online method that processes information at a time. The core idea of this learning algorithm is to generate a model from training on a stored dataset and then using an iterative algorithm like stochastic gradient descent and recursive least squares to learn new data introduced dynamically to the model.

Researchers have made interesting progress in developing online approaches in several research fields including machine learning, pattern recognition, computer vision, game theory, and information theory. Online learning is important for various applications such as faster clustering, forecasting times, catastrophic interference, spam filtering, pattern recognition, and online tracking. Among the extensive related work in this field, we cite the most interesting approaches proposed in literature.

Recognizing human activity is an active research topic where the need to identify real-time moves and actions continuously over the time remains a challenging problem. Authors in [46] propose an online method to recognize human gestures through discriminative key poses and speed-aware action graphs. In [50], a hidden Markov model with modified short-time Viterbi algorithm was proposed for online recognizing human daily activity. In order to deal with the problem of clustering parallel data streams, the authors in [5] develop an online version of the classical K-means clustering algorithm. The idea of this method is based on an incremental computation of distances between streams of data using a DFT approximation. A probabilistic model was proposed for online clustering in [48] to detect the novel objects from sequences of data. They used a non-parametric Dirichlet process for modeling documents in an online fashion and an empirical Bayes method to estimate model hyperparameters. When features are expensive, authors in [39] proposed a novel online feature selection allowing the feature to be only available one at a time. This online framework was based on grafting approach that combines the speed of filters with the accuracy of wrappers. Applied to spam filtering, an online model has been presented to filter a sequence of emails using distance-based kernels and string kernels in [2]. In this paper, we consider particularly the problem of online recognition which is one of the most important problems that arises in computer vision, image analysis, information retrieval, data compression, and pattern recognition. Online cluster analysis is the task of grouping data into homogeneous clusters as long as they arrive in a temporal sequence. Finite mixture model is among the most applied approaches in the context of machine learning applications [11–13, 19, 33], especially for online clustering. In [41], an online approach was introduced based on a stochastic approximation of the Expectation-Maximization algorithm for the normalized Gaussian network. Experiments results showed that this online EM-algorithm for the NGnet is able to manage dynamic

environments and to deal efficiently with the robot dynamics problem. In video surveillance application, an adaptive Gaussian mixture model [29] has been used to model real video data with an incremental EM for the learning update. In [51], Gaussian mixture models have been proposed in an online fashion based on description length reducing prior and a MAP estimation procedure for an up-to-date description of the data. Despite the adoption of this model to various online clustering because of its simplicity, real-world applications cannot be considered by the Gaussian assumption which fails to fit the shape of the data. For instance, recent works have shown that other non-Gaussian models such as the Dirichlet, the generalized Gaussian, and the Beta Liouville mixtures provide better clustering results in several applications. In [8], a finite mixture of Dirichlet and a stochastic approach was proposed in the light of online clustering application, namely the dynamic summarization of image databases. A more general distribution has been applied to this type of non-Gaussian data is the generalized Dirichlet. The authors in [18] have proposed an online variational learning of generalized Dirichlet mixture models with feature selection to challenging problems, namely text clustering and image clustering using the bag-of-visual-words representation. Another approach that can control how the system should perceive new coming data over time is based on the generalized inverted Dirichlet [4]. For recognizing human facial expression, an online variational learning based on Beta-Liouville mixture model was proposed in [17]. A novel approach based on spherical mixtures has been proposed in [3] to tackle the problem of tracking and detecting news topic trend. The model in [1], besides, proposes a flexible online clustering algorithm in order to accurately approximate the non-Gaussian data. This online technique, based on finite mixture of generalized Gaussian distribution, has been applied to video foreground segmentation. In fact, generalized Gaussian mixture models have been the subject of wide applications [16, 26, 40]. However, in many multivariate statistical processes, generalized Gaussian distribution fails to be as accurate as the multivariate generalized Gaussian mixture as shown in previous works [32, 34, 35]. In fact, authors have proved that this multivariate mixture model is able to efficiently recognize human activity. Based on these studies, it is concluded that it is interesting to build our online framework based on the multivariate generalized Gaussian mixture model. One of the fundamental tasks of finite mixture model is parameter estimation, usually related to optimization problem.

The question to ask then is this: how to recursively estimate the parameters of the mixture of multivariate generalized Gaussian distributions and how to simultaneously select the number of components? In this paper, we seek to answer this question by improving our previous deterministic approaches proposed in [34, 35] and presenting a novel online recognition algorithm based on multivariate generalized Gaussian mixture model suitable for various applications. We are mainly interested by recognizing the human actions, facial expression from videos and detecting pedestrian from infrared images.

This paper is organized as follows: Sect. 5.2 proposes the deterministic framework based on multivariate generalized Gaussian mixture model. In Sect. 5.3, we introduce our novel online learning algorithm. Next, we applied the proposed algo-

rithms to the problem of human action recognition, facial expression recognition, and pedestrian detection from infrared images. Finally, we conclude this paper with a summary and potential future works in Sect. 5.4.

5.2 Multivariate Generalized Gaussian Mixture Model

5.2.1 Multivariate Generalized Gaussian Distribution

Multivariate generalized Gaussian distribution (MGGD) is defined by the probability density functions [25] as follows:

$$p(\mathbf{X}|\Sigma; \beta; \boldsymbol{\mu}) = \frac{\Gamma(\frac{d}{2})}{\pi^{\frac{d}{2}}\Gamma(\frac{d}{2\beta})2^{\frac{d}{2\beta}}} \frac{\beta}{|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}((\mathbf{X}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu}))^\beta\right] \quad (5.1)$$

where $\mathbf{X} \in R^d$, $\Sigma = mM$ is a $d \times d$ symmetric positive definite matrix, called the dispersion matrix, $\boldsymbol{\mu}$ is a d-dimensional mean vector, and $\beta > 0$ is the shape parameter that we assumed to be the same for all the dimensions of the data. Noting now that if $\beta = 1$, the MGGD is equivalent to the multivariate Gaussian distribution. The shape parameter β controls the peakedness and the spread of the distribution. The smaller the beta, the more peaked for the probability distribution function (pdf), and the larger the beta, the flatter will be the pdf just as exposed in Fig. 5.1b. Positive shape parameter values produce skewed distributions to the left and bounded to the right. In contrast, negative shape parameter values produce skewed distributions to the right and bounded to the left.

5.2.2 Finite Mixture and Deterministic Learning

The finite mixture of K multivariate generalized Gaussian distributions is given by:

$$p(\mathbf{X}|\Theta) = \sum_{j=1}^{K} p_j p(\mathbf{X}|\Theta_j), \quad (5.2)$$

where $p(\mathbf{X}|\Theta_j)$ is known as the jth component of the mixture defined with its parameters $\Theta_j = (\boldsymbol{\mu}_j, \Sigma_j, \beta_j)$. The parameter p_j is called a mixing weight parameter and must satisfy $0 \leq p_j \leq 1$ together with $\sum_{j=1}^{K} p_j = 1$.

The main purpose of deterministic techniques is maximizing the likelihood function with respect to model's parameters. One of the standard inferential methods and the powerful tool used to fit Gaussian based-mixture model to an observed data is the Expectation-Maximization (EM) algorithm [15]. Its aim is to optimize the
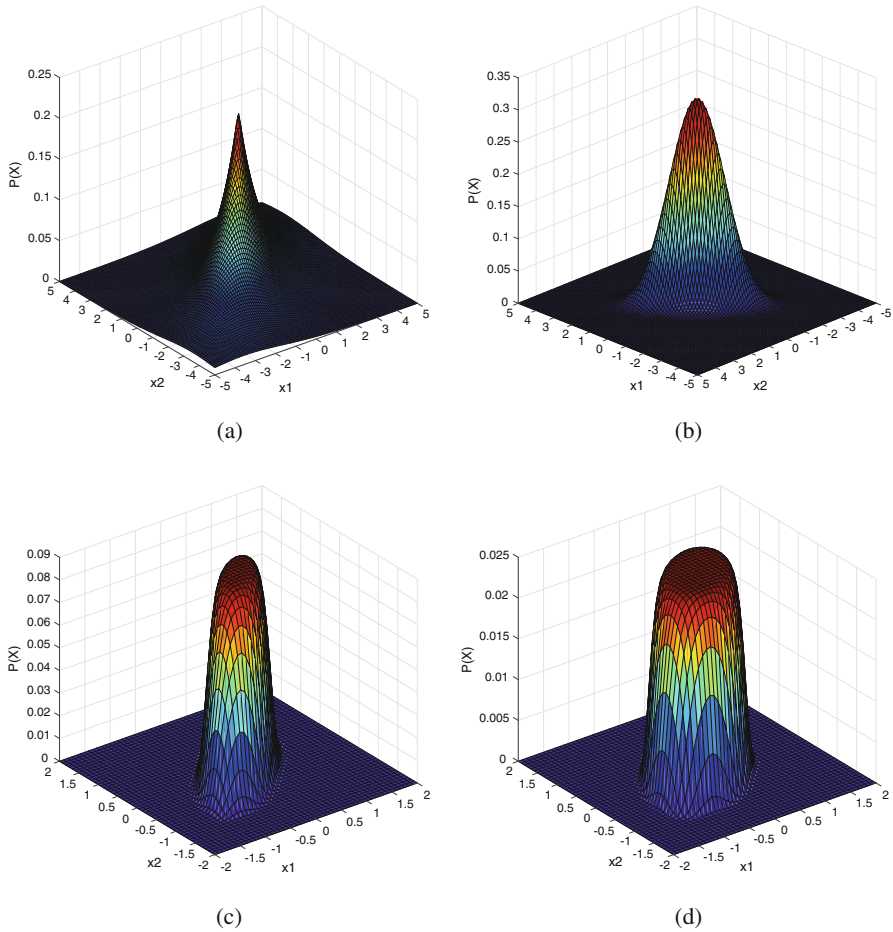
Fig. 5.1 Multivariate generalized Gaussian distributions with different shape parameters. (**a**) $\beta = 0.5$. (**b**) $\beta = 1$. (**c**) $\beta = 3$. (**d**) $\beta = 5$

likelihood function in regard to the model's parameters. The EM algorithm starts with initializing parameters Θ_0. Then, it iterates between two steps: the expectation and the maximization and converges to the maximum. In the expectation step (E-step), the expected likelihood is estimated given the current estimated parameters. For that purpose, the following posterior probability, named also responsibilities, for the j-th component of the mixture is computed:

$$p(j|\mathbf{X}) = \frac{p_j\, p(\mathbf{X}|\Theta_j)}{\sum_{m=1}^{K} p_m\, p(\mathbf{X}|\Theta_m)} \tag{5.3}$$

During the maximization step (M-step), the model's parameters are updated using the current responsibilities. In order to maximize the likelihood function, the log-likelihood function is maximized instead with respect to parameters as it is a monotone function. Then applying the logarithm to the likelihood function, it follows for $\mathcal{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_N)$ that :

$$\mathcal{L}(\mathcal{X}|\Theta) = \sum_{i=1}^{N} \log p(\mathbf{X}_i|\Theta) = \sum_{i=1}^{N} \log \left(\sum_{j=1}^{K} p_j p(\mathbf{X}_i|\Theta_j) \right) \tag{5.4}$$

The M-step can be formally described as solving directly the following equation:

$$\frac{\partial \mathcal{L}(\mathcal{X}|\Theta)}{\partial \Theta_j} = 0 \tag{5.5}$$

Given the multivariate generalized Gaussian distribution $p(\mathbf{X}|\boldsymbol{\mu}_j, \Sigma_j, \beta_j)$, we obtain the following estimates :

- Mixing parameter

$$\hat{p}_j = \frac{1}{N} \sum_{i=1}^{N} p(j|\mathbf{X}_i) \tag{5.6}$$

- Mean parameter

$$\hat{\boldsymbol{\mu}}_j = \frac{\sum_{i=1}^{N} p(j|X_i)|X_i - \boldsymbol{\mu}_j|^{\beta_j - 1} X_i}{\sum_{i=1}^{N} p(j|X_i)|X_i - \boldsymbol{\mu}_j|^{\beta_j - 1}} \tag{5.7}$$

As there is no closed form for the covariance matrix and the likelihood estimation for this parameter is indistinct, the authors in [9, 10, 38] have proven the existence of the covariance matrix estimator up to certain conditions. We present those mentioned works in the following section.

5.2.2.1 Fixed-Point Estimation Method

One of the above-mentioned parameters estimation techniques of the MGGD is the so-called fixed-point method [38]. Indeed, this method guarantees the existence and uniqueness of the MLE of the covariance matrix for each shape parameter belonging to [0,1]. The existence was proved by showing that the profile likelihood is positive, bounded in the set of symmetric positive definite matrices and equals to zero on the boundary of this set. Regarding the uniqueness, it was proved that for any initial symmetric positive definite matrix, the sequence of matrices satisfying a fixed point equation converges to the unique maximum of this profile likelihood.

Let $(X_1, X_2, \ldots, X_N)$ be a random sample of N observation vectors of dimension d, drawn from a zero-mean MGGD with scatter matrix $C = m\Sigma$; m is the scale parameter, and β is the shape parameter. The MLE of m, β, and Σ are found by solving the maximum likelihood equations defined as follows:

$$\hat{m} = \left[\frac{1}{N} \sum_{i=1}^{N} (u_i)^\beta \right]^{\frac{1}{\beta}}, \tag{5.8}$$

where $u_i = X_i^T \Sigma^{-1} X_i$.

Assuming first that β is known. By differentiating the likelihood function with respect to Σ, the MLE of the covariance matrix satisfies the following fixed point (FP) equation:

$$f(\Sigma) = \sum_{i=1}^{N} \frac{d}{u_i + u_i^{1-\beta} \sum_{i \neq j} u_j^\beta} X_i X_i^T, \tag{5.9}$$

In other words, the fixed point equation can be written as:

$$\hat{\Sigma}_{k+1} = f(\Sigma_k) \tag{5.10}$$

Indeed, mathematically the solution of fixed point equation is settled using an iterative proceeding until Σ stabilize (i.e., there is no sensible difference between Σ_k and Σ_{k+1}).

Afterwards, an iterative algorithm based on a Newton–Raphson technique is then applied to compute the maximum likelihood estimation of the shape parameter.

$$\hat{\beta}_{k+1} = \hat{\beta}_k - \frac{\alpha(\hat{\beta}_k)}{\alpha'(\hat{\beta}_k)} \tag{5.11}$$

where

$$\alpha(\beta) = \frac{dN}{2\sum_{i=1}^{N} u_i^\beta} \sum_{i=1}^{N} \left[u_i^\beta \log(u_i) \right] - \frac{dN}{2\beta} \left[\psi\left(\frac{d}{2\beta} \right) + \log(2) \right]$$
$$- N - \frac{dN}{2\beta} \log\left(\frac{\beta}{dN} \sum_{i=1}^{N} u_i^\beta \right) \tag{5.12}$$

where ψ is the digamma function.

5.2.2.2 Riemannian Averaged Fixed-Point Estimation Algorithm

The second developed algorithm is the named "Riemannian Averaged Fixed Point" method (RA-FP) [10]. The latter estimator is proposed as a generalization form of the previous proposed technique fixed-point (FP) estimator [38]. The basic idea of RA-FP algorithm is to implement successive Riemannian average of fixed point iterates in order to estimate the covariance matrix for any positive value of the shape parameter. This process is different from the fixed-point algorithm which estimates the covariance matrix for only the shape parameter belonging to [0, 1].

The RA-FP uses the Riemannian geometry for estimating the covariance matrix. The RA-FP based estimation of Σ is determined as follows:

For $t \in [0, 1]$, the Riemannian average of $\hat{\Sigma}_{k+1}$ is defined as:

$$\hat{\Sigma}_{k+1} = \Sigma_k \#_{t_k} f(\Sigma_k) \tag{5.13}$$
$$= \Sigma_k^{1/2} (\Sigma_k^{-1/2} f(\Sigma_k) \Sigma_k^{-1/2})^{t_k} \Sigma_k^{1/2}$$

where

$$t_k = \frac{1}{k+1}, t_k \in [0, 1], \tag{5.14}$$

and $f(\Sigma)$ is defined as (5.9).

If $t_k = 1$, the RA-FP estimator is reduced to the fixed-point estimator, and Eq. (5.13) yields to (5.10).

For computing the maximum likelihood of the shape parameter, an iterative algorithm based on a Newton–Raphson technique [10] is applied as in the fixed-point algorithm.

5.2.2.3 Fisher Scoring Algorithm

The Fisher scoring algorithm [9] is a maximum likelihood estimator based on the fixed-point technique also and followed by an optimization through the Fisher scoring method. The estimators of m, β are given by Eqs. (5.8) and (5.11) as proposed in [38]. Hence, in this work, the main purpose of the Fisher scoring algorithm is to optimize the likelihood function based on fixed-point technique and followed by an optimization iteration through the Fisher information matrix.

The likelihood function of vectors $\mathcal{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N)$ is given by:

$$\mathcal{L}(\mathcal{X}|\Theta) = \prod_{i=1}^{N} \sum_{j=1}^{K} p_j \, p(\mathbf{X}_i | \Sigma_j; \beta_j; \boldsymbol{\mu}_j) \tag{5.15}$$

The gradient of likelihood function with regard to covariance matrix is defined as:

$$\nabla\mathcal{L}(\Sigma;\Theta) = [F(\Sigma)]^{\frac{N-2}{2}} \nabla F(\Sigma),\tag{5.16}$$

with :

$$F: S_{++}^K \rightarrow R^+\{0\}\tag{5.17}$$

$$\Sigma \quad \rightarrow |\Sigma|^{-1}\Big(\sum_{i=1}^N u_i^\beta\Big)^{-\frac{d}{\beta}}$$

and the gradient of F at a point Σ is given by:

$$\nabla F(\Sigma) = F(\Sigma)\Sigma^{-1}[f(\Sigma) - \Sigma]\Sigma^{-1}\tag{5.18}$$

where $f(\Sigma)$ is defined by fixed-point algorithm in [38].

To numerically maximize the likelihood function, the Fisher-scoring iteration is given by:

$$\Sigma_{k+1} = \Sigma_k + G^{-1}\nabla L(\Sigma;\Theta)\tag{5.19}$$

where the entries of the Fisher information matrix are defined by [45]:

$$G_{ii}(\beta) = \frac{1}{4}\Big(\frac{3d+6\beta}{d+2} - 1\Big),\tag{5.20}$$

$$G_{ij}(\beta) = \frac{1}{4}\Big(\frac{d+2\beta}{d+2} - 1\Big), i \neq j,\tag{5.21}$$

for $i, j = 1, \ldots, K$.

Afterwards, an iterative algorithm based on Newton–Raphson technique is applied to compute the maximum likelihood estimation of the shape parameter as the two previous estimator algorithms.

We summarize the EM-algorithm for the multivariate generalized Gaussian mixture model in the following algorithm:

5.3   Online Learning Algorithm

The deterministic framework presented in the above section was based on batch learning; the parameters are updated on the entire dataset at once. In this section, we introduce an online EM learning approach. We suppose the dataset was represented by M multivariate generalized Gaussian distributions with parameters Θ_N. Assume now at time $t + 1$, a new data $\mathbf{X}_{N+1}$ is inserted to the database, thus, we should

---

Algorithm 1 MGGMM learning algorithm

Require: $\mathcal{X}$, K
Ensure: Θ^*
  Initialization
  Apply the K-Means to obtain the parameters of each component.
  Then, apply the Method-of-Moment for each component j
  repeat
    for j:=1 **to** K **do**
      E-step: Compute posterior probabilities using Eq. (5.3)
      M-step:
      Update $\boldsymbol{\mu}_j$ using Eq. (5.7)
      Update β_j using Eq. (5.11)
      Update Σ_j using FP algorithm ((5.10), (5.9)), RA-FP algorithm (5.13) or FS algorithm (5.19).
    end for
  until Convergence of Likelihood
  Return the model's parameters Θ^*.

---

update the different mixture model parameters with the new input vector. For which reason, a stochastic approximation for obtaining the maximum likelihood of mixture parameters was considered. Namely, we have used the stochastic ascent gradient parameter updating proposed in [47] where the updating parameters are given by:

$$\Theta_{N+1}^{(t+1)} = \Theta_N^{(t)} + \delta_N \frac{\partial \log(p(\mathbf{X}_{N+1}|\Theta_N^{(t)}, p_j^{(t)}))}{\partial \Theta_N^{(t)}} \tag{5.22}$$

In terms of updating the mixing weight, the above updating equation does not ensure the constraints: $0 \leq p_j \leq 1$, $\sum_{j=1}^{M} p_j = 1$. For this aim, a logit parameterization was presented to overcome this problem.

$$\pi_j^{(t)} = \log \frac{p_j}{p_M}, \, j = 1, \ldots, M-1 \tag{5.23}$$

$$\pi_j^{(t+1)} = \pi_j^{(t)} + \delta_N \left(Z_{N+1,j}^{(t+1)} - p_j^{(t)} \right), \, j = 1, \ldots, M-1 \tag{5.24}$$

So that, for $j = 1 \ldots, M-1$, the updating mixing weight is given by:

$$p_j^{(t+1)} = \frac{\exp(\pi_j^{(t+1)})}{1 + \sum_{j=1}^{M-1} \exp\left(\pi_j^{(t+1)}\right)}, \tag{5.25}$$

$$p_M^{(t+1)} = \frac{1}{1 + \sum_{j=1}^{M-1} \exp\left(\pi_j^{(t+1)}\right)} \tag{5.26}$$

The updating mean parameter, shape parameter, and covariance matrix are as follows:

$$\boldsymbol{\mu}_j^{(t+1)} = \boldsymbol{\mu}_j^{(t)} + \delta_N * Z_{N+1,j}^{(t+1)} * \left[(\mathbf{X}_{N+1} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{X}_{N+1} - \boldsymbol{\mu}_j) \right]^{\beta_j - 1} \quad (5.27)$$

$$\beta_j^{(t+1)} = \beta_j^{(t)} + \delta_N * Z_{N+1,j}^{(t+1)} * \left[\left(\frac{1}{\beta_j} + \frac{d\psi(d/2\beta_j)}{2\beta_j^2} + \frac{d\log(2)}{2\beta_j^2} \right) \right.$$

$$- ((\mathbf{X}_{N+1} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{X}_{N+1} - \boldsymbol{\mu}_j)^{\beta_j})$$

$$\left. \log((\mathbf{X}_{N+1} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{X}_{N+1} - \boldsymbol{\mu}_j)) \right] \quad (5.28)$$

$$\Sigma_j^{(t+1)} = \Sigma_j^{(t)} + \delta_N * Z_{N+1,j}^{(t+1)} * \left[\left(-\frac{1}{2} tr(\Sigma_j^{-1} d\Sigma) \right) \right.$$

$$+ \frac{\beta_j}{2} (\mathbf{X}_{N+1} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} d\Sigma_j \Sigma_j^{-1}$$

$$\left. ((\mathbf{X}_{N+1} - \boldsymbol{\mu}_j)(\mathbf{X}_{N+1} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{X}_{N+1} - \boldsymbol{\mu}_j))^{(\beta_j - 1)} \right] \quad (5.29)$$

Thus, the complete online updating MGGMM algorithm was resumed as follows.

Algorithm 2 Online MGGMM learning algorithm

Require: $\mathcal{X} = \{X_1, \ldots, X_N\}, \Theta_N^{(t)}, K$
Ensure: $\Theta_{N+1}^{(t+1)}$
  At $t + 1$, new data vector $\mathbf{X}_{N+1}$
  repeat
    for j:=1 **to** K **do**
      Compute posterior probability

$$z_{N+1j} = \frac{p_j p(\mathbf{X}_{N+1}|\Theta_j)}{\sum_{m=1}^K p_m p(\mathbf{X}_{N+1}|\Theta_m)} \quad (5.30)$$

      Affect $\mathbf{X}_{N+1}$ to a cluster using the Bayes rule: $\mathbf{X}_{N+1}$ is affected to cluster j_1 if $z_{N+1j_1} > z_{N+1j}, \forall j \neq j_1$
      Update the weights using Eqs. (5.25) and (5.26)
      Update $\boldsymbol{\mu}_j$ using Eq. (5.27)
      Update β_j using Eq. (5.28)
      Update Σ_j using Eq. (5.29).
    end for
  until Convergence of Likelihood
  Return the model's parameters Θ^{N+1}.

5.4   Experiments Results

5.4.1   Datasets

In our experiments, we are using three different datasets to evaluate the performance of the proposed online mixture model. For human action recognition, we use the well-known KTH dataset [42]. This human action dataset presents to date the most tremendous at handset of video sequences for human actions. It contains 2391 sequences categorized in six different human actions: walking, jogging, running, boxing, hand waving, and hand clapping. Each action is performed in four diverse scenarios: outdoors, outdoors with scale variations, outdoors with different clothes, and indoors. We present some examples of frames from video sequences in each category in Fig. 5.2. For recognizing facial expression, the set of data that we have used is the Cohn–Kanade dataset [22]. It contains 486 sequences where each sequence starts with a neutral expression and proceeds to a target expression of anger, surprise, joy, fear, sadness, or disgust. The sequences are collected from 97 university students ranging in age from 18 to 30 years. Sixty-five percent were female, 15% were African-American, and 3% were Asian or Latino with one to six emotions per subject. Sample images from this database with different facial expressions are shown in Fig. 5.3. In regard to the detecting pedestrian from infrared images, we make use of a challenging dataset of thermal imagery, namely the OSU thermal dataset. It is composed of 10 test collections with the total of 284 thermal images. Those images contain 984 pedestrians captured from Ohio State University campus using a Raytheon 300D thermal sensor core with 75 mm lens mounted on an 8-story building. We display an exemplary of different number of pedestrians in Fig. 5.4.
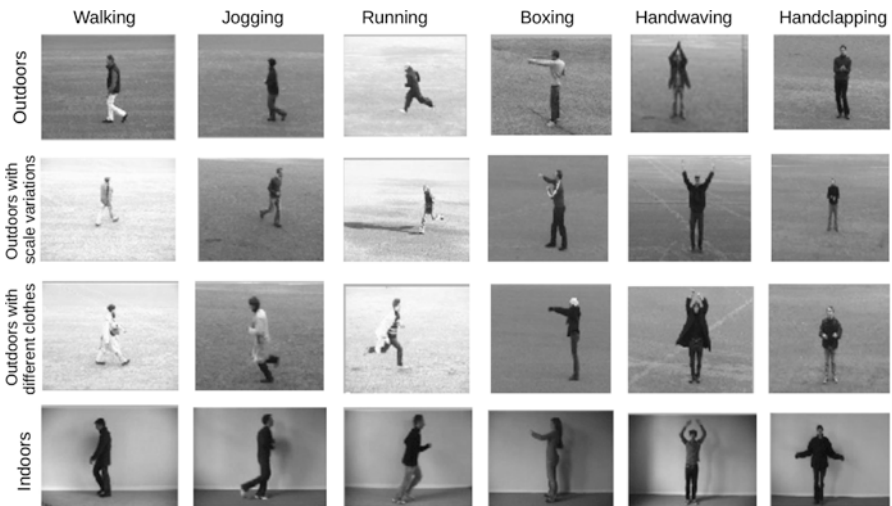
Fig. 5.2 Examples of frames from the KTH dataset of different human actions within different scenarios

Fig. 5.3 Sample face expression images from the Cohn–Kanade database

Fig. 5.4 Example of pedestrian images from the OSU-thermal database

5.4.2 Database Preprocessing Approach

The methodology that we have adopted for each application can be summarized as follows (Fig. 5.5). Basically, we have adopted the bag-of-words approach to represent our images and video sequences. In this model, each image or video of the dataset is depicted as a set of features. First, we extracted local spatio-temporal features from each video sequence from KTH database using space-time interest point detector [28] and SIFT3D descriptor [43]. From Cohn–Kanade videos dataset, we extracted dynamic textures features using LBP-TOP descriptor within 9×8 blocks [49]. Besides, for the infrared images, we used dense SIFT descriptors [30]
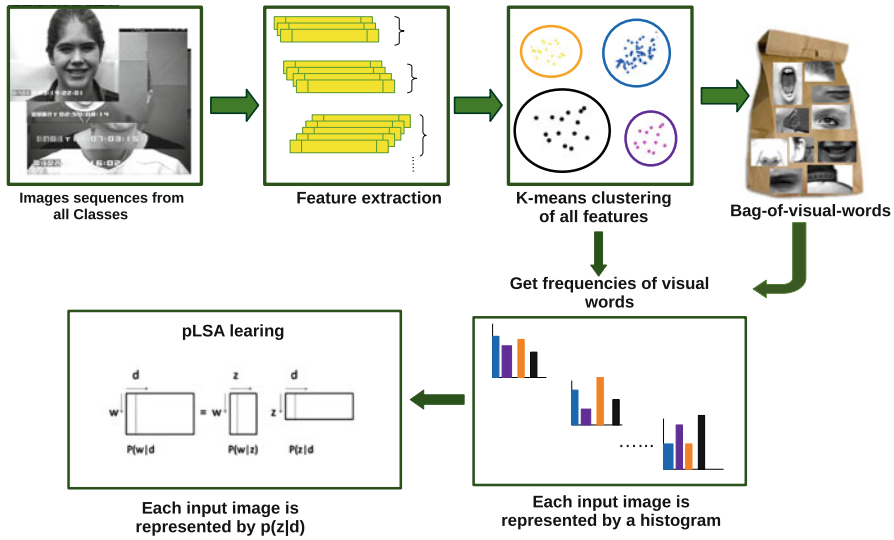
Fig. 5.5 Database preprocessing approach

of 16×16 pixel patches computed over a grid with a spacing of 8 pixels. Second, we have quantized the extracted features into visual words using K-means algorithm [14], and each image and video is then represented as a frequency histogram over the visual words. Finally, we have applied a probabilistic latent semantic analysis (pLSA) to the obtained histograms in order to represent each image by a d-dimensional vector where d is the number of latent aspects [6].

5.4.3 Online Human Action Recognition

Recognizing human from videos is a widely studied problem in many applications, both offline and online. The interest of online processing is motivated by the promise of many applications. If we take the example of video surveillance systems for airports, online human recognition plays a key role in protecting against acts of terrorism and in providing real-time surveillance in various airport departments. Another application of online human action processing is in smart environments such as health care and assisted living geared to provide housing facilities for elderly population and people with disabilities. Accordingly, many recent researchers have concentrated on online human action recognition. An approach based on motion data and location information has been adapted in [50] to indoor human daily activity recognition. A combination between neural network

and HMM has been proposed to model motion data and location information. A more recent approach [46] based on semi-supervised learning was proposed to robustly recognize moves online from unsegmented data. In [24], an online activity recognition on smart phones using the built-in accelerometers was proposed to classify the basic movement of the user. This method was performed using the KNN classifier and evaluated by applying Naive Bayes classification method. To deal with the problem of continuous activities and personalized learning, an online multitask learning method for large-scale personalized activity recognition has been introduced in [44]. Using a dataset obtained from real-home settings, the authors in [36] have proposed an evolving fuzzy systems to recognize activities of the daily living (ADLs) from sensor streams.

5.4.4 Online Human Facial Expression Recognition

Recognizing human facial expression is an active research problem in the recent years. Various works have focused on online facial expression recognition that have been used for different applications such as smart environment, video surveillance systems, e-education, and many other interesting utilization. The interest in online human facial expression recognition is motivated by the promise of automatically categorizing the different types of human face expression used in computer interaction, medicine, e-learning, access control, monitoring, and marketing. For example, knowing the client's emotional state, computer can become a more effective interface to detect patient feeling about medical treatment. For instance, interpreting autism's expressions could help in developing a therapy system. In tutoring system, detecting the state of the learner may enhance the presentation style of e-learning program. Another interesting application is to detect drivers' state, helping the driver monitor their stress level and alert other cars. If we take the example of video surveillance systems for ATM, facial expression recognition plays a key role in protecting against acts of terrorism and theft as it doesn't dispense money when someone is scared. Many researches have focused on facial expression recognition but few of them were interested how could we understand the facial expression in an interactive way. A system built on elastic graph matching [21] was proposed to track and detect the face of a person in a live video sequence. In [7], a study on understanding how babies learn to recognize facial expressions is presented. They have used a cognitive system algebra combined with a neural network model to online recognize facial expressions. A method for collecting and analyzing facial responses over the web was introduced in [31]. The proposed framework was utilized to crowdsource over three million face videos in response to thousands of media clips ranging from advertisements to movie trailers to TV shows and political debates.

5.4.5   Online Pedestrian Detection in Infrared Images

Infrared (IR) thermography is an imaging method for visualizing radiation not observable by human eye. Analyzing thermal images has occurred growing interest both in research and in industry with a wide area of applications. In military surveillance uses, they have the need to mount infrared detection system on vehicles or towers for border surveillance. Fire-fighters use infrared imaging as a mechanism to find missing people in buildings on fire. As well, car-pedestrians accidents which occur at night acquire the use of far-infrared camera in order to discern the thermal energy. Online detection pedestrians from infrared images has not been much explored yet. In [27], a real-time online learning was proposed to track pedestrians using boosted random ferns and using Weber–Fechner's law to detect pedestrians according to the season and the weather. Another interesting work on online pedestrians detection based on particle filters and combination of a local intensity distribution (LID) with oriented center symmetric local binary patterns was introduced in [23]. The proposed algorithm was applied to various thermal videos to detect the most likely target position in the subsequent frame.

5.4.6   Results

In this section, we evaluate our proposed framework in different experiments and we compare recognition rates with methods from literature and offline methods. After preprocessing our databases, we used 15 subjects from each activity from KTH dataset to construct the visual vocabulary and the remaining 10 subjects to test. For facial expression and pedestrian detection experiments, we selected 70% of the data to the training and the remaining for the test.

We start by studying the impact of the visual vocabulary sizes on the recognition accuracy for our online methods onFP-MGGMM, onRA-FP-MGGMM, onFS-MGGMM, and the other approaches (GMM,GGMM), as depicted in Fig. 5.6a for KTH dataset, Fig. 5.7a for Cohn–Kanade and in Fig. 5.8a for OSU thermal dataset. According to these results, the maximum accuracy value is obtained with visual vocabulary sizes of 200 for KTH, 20 for Cohn–Kanade, and 50 for OSU thermal dataset. Moreover, we have studied the impact of the number of aspects on the recognition accuracy as shown in Figs. 5.6b, 5.7b, and 5.8b and we found that the optimal accuracy was obtained when the number of aspects was set to 20 for KTH and 6 for Cohn–Kanade and OSU thermal dataset.

We achieved the best performance with human action recognition, facial expression recognition, and infrared pedestrian detection in different proposed online learning multivariate generalized Gaussian methods. For instance, for human action recognition, the online Riemannian averaged fixed-point multivariate generalized Gaussian mixture achieves the best recognition rates (99.37%) as shown in Table 5.1. In recognizing facial expression, the online fixed-point MGGMM

(96.84%) outperforms the other related works, Gaussian-based models, and also the two proposed online mixture models as indicated in Table 5.2. With respect to infrared pedestrian detection, experiments results on OSU thermal dataset shown in Table 5.3 that the Fisher-scoring MGGMM provides the best performance (96.64%) as compared to other online Gaussian-based models. We notice from those tables (Tables 5.1–5.3) that our three proposed discriminative online learning methods reached superior performance where the accuracy increases approximately by 20% comparing to the basic Gaussian mixture model and the univariate generalized Gaussian mixture. We display also the confusion matrix for the proposed online learning methods for KTH database in Tables 5.4, 5.5, 5.6, for Cohn–Kanade in Tables 5.7, 5.8, 5.9, and for OSU-thermal database in Tables 5.10, 5.11, and 5.12.
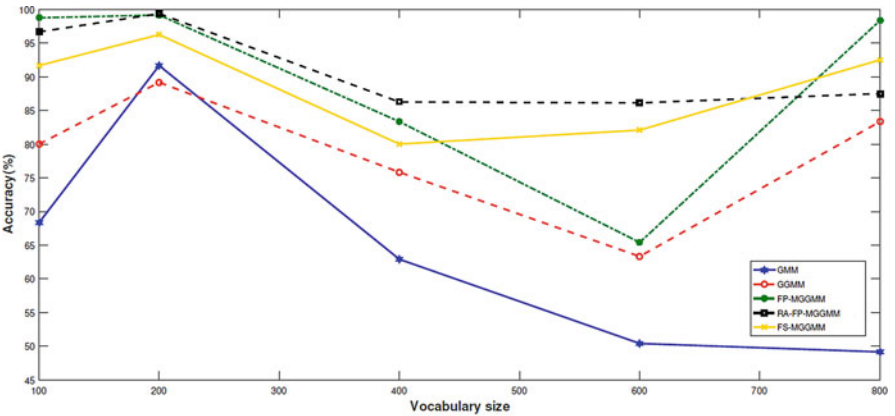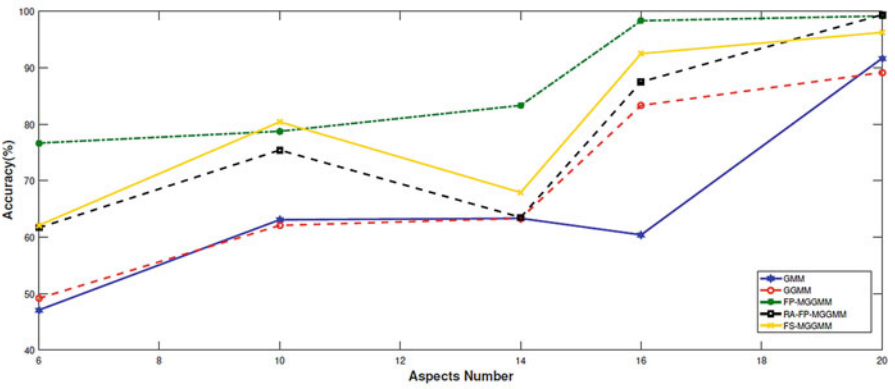
Fig. 5.6 (**a**) Recognition accuracy vs. vocabulary size for the KTH dataset; (**b**) recognition accuracy vs. the number of aspects for the KTH dataset
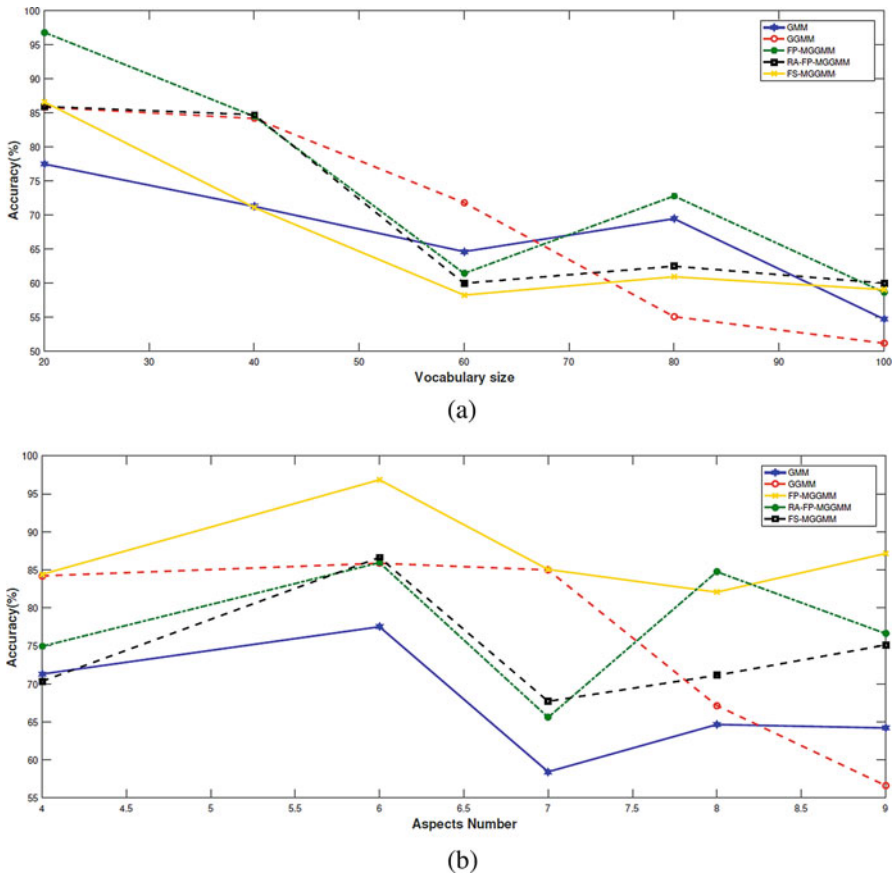
Fig. 5.7 (a) Recognition accuracy vs. vocabulary size for the Cohn–Kanade dataset. (b) Recognition accuracy vs. the number of aspects for the Cohn–Kanade dataset

We compare the performance of the offline learning and online learning proposed in Sect. 5.2.2 on all the three datasets. Tables 5.13, 5.14, and 5.15 illustrate the accuracy and the running time for each proposed model and for the Gaussian mixture model and the generalized Gaussian mixture model. According to those tables, we notice that online learning has improved the quality of the clusters and decreased the time of running compared to the offline learning.
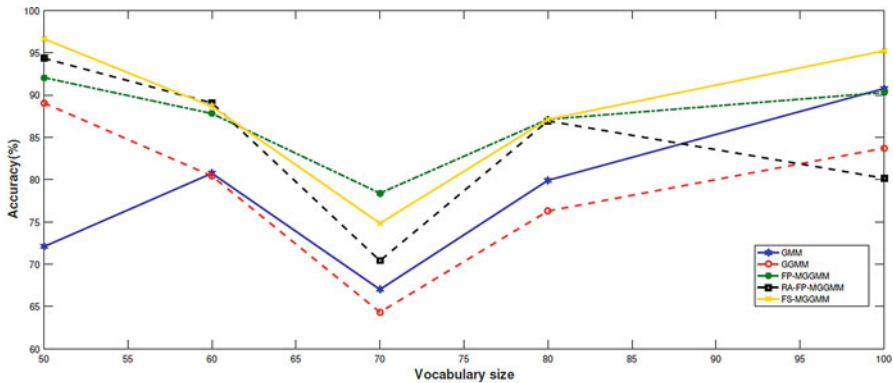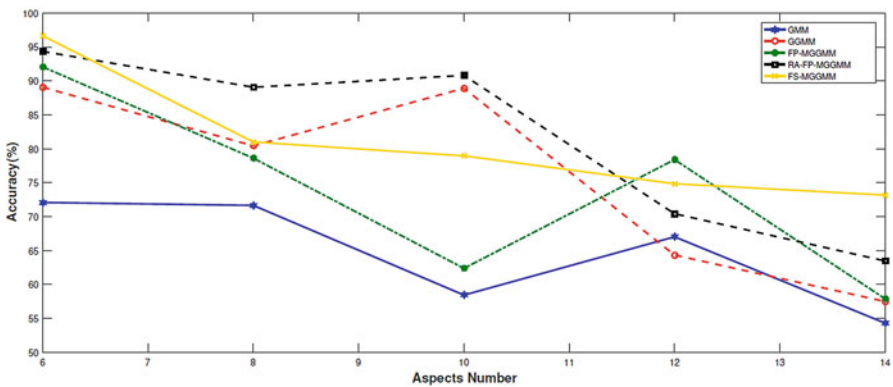
Fig. 5.8 (**a**) Recognition accuracy vs. vocabulary size for the OSU-thermal dataset. (**b**) Recognition accuracy vs. the number of aspects for the OSU-thermal dataset

Table 5.1 The average recognition rates using different online algorithms for KTH dataset

Approach	Recognition rates
sig min-Hash [20]	91.2
Baseline [20]	44.3
OHAC [37]	82.2
onGMM	91.66
onGGMM	89.16
onFP-MGGMM	99.16
onRA-FP-MGGMM	99.37
onFS-MGGMM	96.25

Table 5.2 The average recognition rates using different online algorithms for Cohn–Kanade dataset

Approach	Recognition rates
onBM [17]	84.57
onDM	79.69
onGDM	83.08
onGMM	77.50
onGGMM	85.85
onFP-MGGMM	96.84
onRA-FP-MGGMM	85.94
onFS-MGGMM	86.62

Table 5.3 The average recognition rates using different algorithms for OSU-thermal dataset

Approach	Recognition rates
onGMM	90.76
onGGMM	89.06
onFP-MGGMM	92.06
onRA-FP-MGGMM	94.34
onFS-MGGMM	96.64

Table 5.4 Confusion matrix for KTH dataset using onFP-MGGMM

	C1	C2	C3	C4	C5	C6
C1	97.5%	0%	2.5%	0%	0%	0%
C2	0%	100%	0%	0%	0%	0%
C3	0%	0%	100%	0%	0%	0%
C4	0%	0%	0%	100%	0%	0%
C5	0%	0%	0%	0%	100%	0%
C6	0%	0%	0%	0%	2.5%	97.5%

Table 5.5 Confusion matrix for KTH dataset using onRA-FP-MGGMM

	C1	C2	C3	C4	C5	C6
C1	92.5%	5%	0%	0%	2.5%	0%
C2	0%	100%	0%	0%	0%	0%
C3	0%	0%	100%	0%	0%	0%
C4	0%	0%	0%	100%	0%	0%
C5	0%	0%	0%	0%	100%	0%
C6	0%	0%	0%	0%	0%	100%

Table 5.6 Confusion matrix for KTH dataset using onFS-MGGMM

	C1	C2	C3	C4	C5	C6
C1	90%	2.5%	0%	0%	7.5%	0%
C2	0%	92.5%	2.5%	0%	0%	5%
C3	0%	0%	100%	0%	0%	0%
C4	0%	2.5%	0%	97.5%	0%	0%
C5	0%	0%	0%	2.5%	97.5%	0%
C6	0%	0%	0%	0%	0%	100%

Table 5.7 Confusion matrix for Cohn–Kanade dataset using onFP-MGGMM

	C1	C2	C3	C4	C5	C6
C1	100%	0%	0%	0%	0%	0%
C2	0%	91.66%	8.34%	0%	0%	0%
C3	0%	0%	93.75%	0%	6.25%	0%
C4	0%	0%	0%	100%	0%	0%
C5	0%	0%	0%	0%	100%	0%
C6	4.35%	0%	0%	0%	0%	95.65%

Table 5.8 Confusion matrix for Cohn–Kanade dataset using onRA-FP-MGGMM

	C1	C2	C3	C4	C5	C6
C1	86.66%	0%	6.6%	0%	6.74%	0%
C2	0%	91.66%	0%	8.34%	0%	0%
C3	6.25%	0%	87.5%	0%	6.25%	0%
C4	0%	14.05%	0%	80.95%	0%	5%
C5	0%	0%	0%	19.05%	80.95%	0%
C6	0%	0%	0%	17.4%	0%	82.60%

Table 5.9 Confusion matrix for Cohn–Kanade dataset using onFS-MGGMM

	C1	C2	C3	C4	C5	C6
C1	100%	0%	0%	0%	0%	0%
C2	8.34%	91.66%	0%	0%	0%	0%
C3	6.25%	0%	87.5%	0%	0%	6.25%
C4	3.4%	0%	11.13%	66.66%	18.81%	0%
C5	0%	9.53%	0%	0%	90.47%	0%
C6	8.7%	0%	0%	8.7%	0%	82.60%

Table 5.10 Confusion matrix for OSU-thermal dataset using onFP-MGGMM

	C1	C2	C3	C4	C5	C6	C7
C1	94.73%	0%	5.27%	0%	0%	0%	0%
C2	0%	86.66%	0%	13.34%	0%	0%	0%
C3	9.1%	0%	90.90%	0%	0%	0%	0%
C4	0%	7.69%	0%	84.61%	0%	7.7%	0%
C5	11.38%	0%	11.7%	0%	76.92%	0%	0%
C6	0%	36.37%	0%	0%	0%	63.63%	0%
C7	0%	0%	0%	16.67%	0%	0%	83.33%

Table 5.11 Confusion matrix for OSU-thermal dataset using onRA-FP-MGGMM

	C1	C2	C3	C4	C5	C6	C7
C1	94.73%	0%	5.27%	0%	0%	0%	0%
C2	13.3%	80%	0%	6.7%	0%	0%	0%
C3	0%	0%	90.90%	0%	9.1%	0%	0%
C4	0%	7.7%	0%	92.30%	0%	0%	0%
C5	0%	0%	23.08%	0%	76.92%	0%	0%
C6	0%	36.37%	0%	0%	0%	63.63%	0%
C7	0%	0%	0%	8.34%	0%	0%	91.66%

Table 5.12 Confusion matrix for OSU-thermal dataset using onFS-MGGMM

	C1	C2	C3	C4	C5	C6	C7
C1	92.10%	0%	0%	7.9%	0%	0%	0%
C2	0%	76.66%	0%	11.6%	0%	11.74%	0%
C3	0%	0%	90.90%	0%	0%	9.1%	0%
C4	11.54%	0%	0%	88.46%	0%	0%	0%
C5	0%	19.24%	0%	0%	80.76%	0%	0%
C6	0%	0%	0%	0%	13.64%	86.36%	0%
C7	0%	16.67%	0%	0%	0%	0%	83.33%

Table 5.13 Performance of human recognition using offline and online algorithms for KTH dataset

	Offline		Online	
	Accuracy (%)	Run time (s)	Accuracy (%)	Run time (s)
GMM	87.14	27.78	91.66	8.63
GGMM	88.33	38.36	89.16	3.39
FP-MGGMM	91.97	19.18	99.16	10.29
RA-FP-MGGMM	94.16	22.71	99.37	10.44
FS-MGGMM	94.58	15.87	96.25	10.03

Table 5.14 Performance of facial expression recognition using offline and online algorithms for Cohn–Kanade dataset

	Offline		Online	
	Accuracy (%)	Run time (s)	Accuracy (%)	Run time (s)
GMM	76.46	14.74	77.50	2.31
GGMM	85.85	9.55	85.85	1.36
FP-MGGMM	88.98	7.1	96.84	2.35
RA-FP-MGGMM	91.61	8.76	85.94	2.61
FS-MGGMM	93.87	3.8	86.62	2.24

Table 5.15 The average recognition rates using offline and online algorithms for OSU-thermal dataset

	Offline		Online	
	Accuracy (%)	Run time (s)	Accuracy (%)	Run time (s)
GMM	79.61	15.17	90.76	1.78
GGMM	84.26	8.56	89.06	1.05
FP-MGGMM	89.32	6.65	92.06	1.7
RA-FP-MGGMM	92.51	6.75	94.34	2.1
FS-MGGMM	89.81	6.21	96.64	2.02

5.5 Conclusion

In this paper, we have proposed an online learning based on deterministic framework that is able to estimate the multivariate parameters of the generalized Gaussian mixture model. To this aim, we were motivated by developing a robust maximum likelihood approach based on recent techniques, namely fixed-point, Riemannian-averaged fixed-point, and Fisher scoring and a stochastic gradient descent algorithm. We applied our algorithms on extensive experiments including challenging applications namely recognizing human action, facial expression and detecting pedestrian in infrared images. Comparisons revealed that our online methods achieve better recognition rates with respect to other offline methods, proposed methods in literature, and other Gaussian-based models. In spite of the promising results achieved, further enhancement could be done using online variational learning.

References

1. Allili, M.S., Bouguila, N., Ziou, D.: Online video foreground segmentation using general gaussian mixture modeling. In: 2007 IEEE International Conference on Signal Processing and Communications, pp. 959–962. IEEE, New York (2007)
2. Amayri, O., Bouguila, N.: Online spam filtering using support vector machines. In: 2009 IEEE Symposium on Computers and Communications, pp. 337–340. IEEE, New York (2009)
3. Amayri, O., Bouguila, N.: Online news topic detection and tracking via localized feature selection. In: The 2013 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE, New York (2013)
4. Bdiri, T., Bouguila, N., Ziou, D.: A statistical framework for online learning using adjustable model selection criteria. Eng. Appl. Artif. Intell. **49**, 19–42 (2016)
5. Beringer, J., Hüllermeier, E.: Online clustering of parallel data streams. Data Knowl. Eng. **58**(2), 180–204 (2006)
6. Bosch, A., Zisserman, A., Muñoz, X.: Scene classification via pLSA. In: Computer Vision–ECCV 2006, pp. 517–530 (2006)
7. Boucenna, S., Gaussier, P., Andry, P., Hafemeister, L.: Imitation as a communication tool for online facial expression learning and recognition. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5323–5328. IEEE, New York (2010)

8. Bouguila, N., Ziou, D.: Online clustering via finite mixtures of Dirichlet and minimum message length. Eng. Appl. Artif. Intell. **19**(4), 371–379 (2006)

9. Boukouvalas, Z., Fu, G.S., Adalı, T.: An efficient multivariate generalized gaussian distribution estimator: application to IVA. In: 2015 49th Annual Conference on Information Sciences and Systems (CISS), pp. 1–4. IEEE, New York (2015)

10. Boukouvalas, Z., Said, S., Bombrun, L., Berthoumieu, Y., Adalı, T.: A new Riemannian averaged fixed-point algorithm for MGGD parameter estimation. IEEE Signal Process. Lett. **22**(12), 2314–2318 (2015)

11. Channoufi, I., Bourouis, S., Bouguila, N., Hamrouni, K.: Color image segmentation with bounded generalized gaussian mixture model and feature selection. In: 4th International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2018, Sousse, March 21–24, 2018, pp. 1–6 (2018)

12. Channoufi, I., Bourouis, S., Bouguila, N., Hamrouni, K.: Image and video denoising by combining unsupervised bounded generalized gaussian mixture modeling and spatial information. Multimedia Tools Appl. **77**(19), 25591–25606 (2018)

13. Channoufi, I., Bourouis, S., Bouguila, N., Hamrouni, K.: Spatially constrained mixture model with feature selection for image and video segmentation. In: Image and Signal Processing - 8th International Conference, ICISP 2018, Proceedings, Cherbourg, July 2–4, 2018, pp. 36–44 (2018)

14. Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, ECCV, Prague, vol. 1, pp. 1–2 (2004)

15. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B (Methodol.) **39**(1), 1–38 (1977)

16. Elguebaly, T., Bouguila, N.: Semantic scene classification with generalized gaussian mixture models. In: International Conference Image Analysis and Recognition, pp. 159–166. Springer, New York (2015)

17. Fan, W., Bouguila, N.: Online facial expression recognition based on finite Beta-Liouville mixture models. In: 2013 International Conference on Computer and Robot Vision, pp. 37–44. IEEE, New York (2013)

18. Fan, W., Bouguila, N.: Online variational learning of generalized Dirichlet mixture models with feature selection. Neurocomputing **126**, 166–179 (2014)

19. Fan, W., Sallay, H., Bouguila, N., Bourouis, S.: A hierarchical Dirichlet process mixture of generalized Dirichlet distributions for feature selection. Comput. Electr. Eng. **43**, 48–65 (2015)

20. Gilbert, A., Bowden, R.: Image and video mining through online learning. Comput. Vis. Image Underst. **158**, 72–84 (2017)

21. Hong, H., Neven, H., Von der Malsburg, C.: Online facial expression recognition based on personalized galleries. In: Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 354–359. IEEE, New York (1998)

22. Kanade, T., Tian, Y., Cohn, J.F.: Comprehensive database for facial expression analysis. In: Conference Paper, Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), p. 46. IEEE, New York (2000)

23. Ko, B.C., Kwak, J.Y., Nam, J.Y.: Human tracking in thermal images using adaptive particle filters with online random forest learning. Opt. Eng. **52**(11), 113105 (2013)

24. Kose, M., Incel, O.D., Ersoy, C.: Online human activity recognition on smart phones. In: Workshop on Mobile Sensing: From Smartphones and Wearables to Big Data, vol. 16, pp. 11–15 (2012)

25. Kotz, S.: Multivariate distributions at a cross-road. In: Statistical Distributions in Scientific Work, vol. 1, pp. 247–270 (1975)

26. Kumar, K.N., Rao, K.S., Srinivas, Y., Satyanarayana, C.: Studies on texture segmentation using D-dimensional generalized gaussian distribution integrated with hierarchical clustering. Int. J. Image Graph. Signal Process. **8**(3), 45–54 (2016)

27. Kwak, J.Y., Ko, B.C., Nam, J.Y.: Pedestrian tracking using online boosted random ferns learning in far-infrared imagery for safe driving at night. IEEE Trans. Intell. Transp. Syst. **18**(1), 69–81 (2017)
28. Laptev, I.: On space-time interest points. Int. J. Comput. Vis. **64**(2–3), 107–123 (2005)
29. Lee, D.S.: Effective gaussian mixture learning for video background subtraction. IEEE Trans. Patt. Anal. Mach. Intell. **27**(5), 827–832 (2005)
30. Lowe, D.G., et al.: Object recognition from local scale-invariant features. In: ICCV, vol. 99, pp. 1150–1157 (1999)
31. McDuff, D., Kaliouby, R., Picard, R.: Crowdsourcing facial responses to online videos: extended abstract. In: Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII) (Xi'an), pp. 512–518 (2015)
32. Naiar, F., Bourouis, S., Bouguila, N., Belghith, S.: A fixed-point estimation algorithm for learning the multivariate GGMM: application to human action recognition. In: 2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE), pp. 1–4. IEEE, New York (2018)
33. Najar, F., Bourouis, S., Bouguila, N., Belghith, S.: A comparison between different gaussian-based mixture models. In: 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), pp. 704–708. IEEE, New York (2017)
34. Najar, F., Bourouis, S., Zaguia, A., Bouguila, N., Belghith, S.: Unsupervised human action categorization using a Riemannian averaged fixed-point learning of multivariate GGMM. In: International Conference Image Analysis and Recognition, pp. 408–415. Springer, New York (2018)
35. Najar, F., Bourouis, S., Bouguila, N., Belghith, S.: Unsupervised learning of finite full covariance multivariate generalized gaussian mixture models for human activity recognition. Multimed. Tools Appl. **1**, 1–23 (2019)
36. Ordóñez, F.J., Iglesias, J.A., De Toledo, P., Ledezma, A., Sanchis, A.: Online activity recognition using evolving classifiers. Expert Syst. Appl. **40**(4), 1248–1255 (2013)
37. Panzner, M., Beyer, O., Cimiano, P.: Human activity classification with online growing neural gas. In: Workshop on New Challenges in Neural Computation (NC2) (2013)
38. Pascal, F., Bombrun, L., Tourneret, J.Y., Berthoumieu, Y.: Parameter estimation for multivariate generalized gaussian distributions. IEEE Trans. Signal Process. **61**(23), 5960–5971 (2013)
39. Perkins, S., Theiler, J.: Online feature selection using grafting. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03), pp. 592–599 (2003)
40. Sailaja, V., Srinivasa Rao, K., Reddy, K.: Text independent speaker identification with finite multivariate generalized gaussian mixture model and hierarchical clustering algorithm. Int. J. Comput. Appl. **11**(11), 0975–8887 (2010)
41. Sato, M.A., Ishii, S.: On-line EM algorithm for the normalized gaussian network. Neural Comput. **12**(2), 407–432 (2000)
42. Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: a local SVM approach. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, vol. 3, pp. 32–36. IEEE, New York (2004)
43. Scovanner, P., Ali, S., Shah, M.: A 3-dimensional sift descriptor and its application to action recognition. In: Proceedings of the 15th ACM International Conference on Multimedia, pp. 357–360. ACM, New York (2007)
44. Sun, X., Kashima, H., Ueda, N.: Large-scale personalized human activity recognition using online multitask learning. IEEE Trans. Knowl. Data Eng. **25**(11), 2551–2563 (2013)
45. Verdoolaege, G., Scheunders, P.: On the geometry of multivariate generalized gaussian models. J. Math. Imag. Vis. **43**(3), 180–193 (2012)
46. Vieira, T., Faugeroux, R., Martínez, D., Lewiner, T.: Online human moves recognition through discriminative key poses and speed-aware action graphs. Mach. Vis. Appl. **28**(1–2), 185–200 (2017)
47. Yao, J.F.: On recursive estimation in incomplete data models. Statistics **34**(1), 27–51 (2000)

48. Zhang, J., Ghahramani, Z., Yang, Y.: A probabilistic model for online document clustering with application to novelty detection. In: Advances in Neural Information Processing Systems, pp. 1617–1624 (2005)
49. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Patt. Anal. Mach. Intell. **29**(6), 915–928 (2007)
50. Zhu, C., Sheng, W.: Motion-and location-based online human daily activity recognition. Pervasive Mob. Comput. **7**(2), 256–269 (2011)
51. Zivkovic, Z., van der Heijden, F.: Recursive unsupervised learning of finite mixture models. IEEE Trans. Pattern Anal. Mach. Intell. **26**(5), 651–656 (2004)