SPATIO-TEMPORAL OBJECT RECOGNITION USING VARIATIONAL LEARNING OF AN INFINITE STATISTICAL MODEL

Wentao Fan

Concordia University
Electrical and Computer Engineering
Montreal, Canada
email: wenta_fa@encs.concordia.ca

Nizar Bouguila

Concordia University
Institute for Information Systems Engineering
Montreal, Canada
email:nizar.bouguila@concordia.ca

ABSTRACT

In this paper we present a sophisticated variational Bayes framework for learning infinite Beta-Liouville mixture models. A key feature of the proposed framework is that the appropriate mixture model complexity can be discovered automatically from the data to cluster as part of the inference procedure. Another important advantage is that the whole inference process itself is analytically tractable with closed-form solutions. Moreover, the problems of over-fitting and under-fitting are also prevented thanks to the nonparametric Bayesian nature of the proposed framework. The effectiveness of our statistical framework is investigated on two challenging motion recognition tasks including hand gesture and human activity recognition.

Index Terms— Clustering, mixture models, Dirichlet process, Beta-Liouville, variational Bayes, hand gesture, human activity.

1. INTRODUCTION

During the last decade, finite mixture models have drawn significant attention and have been applied in many fields such as machine learning, image processing and bioinformatics [1]. Conventionally, a finite mixture model, as a parametric approach, uses a fixed and finite number of components for learning the underlying structure of data. However, this may cause over-fitting or under-fitting of the data when there is a misfit between the complexity of the model (the number of components) and the amount of data that is available. Thus, a central problem in mixture modeling concerns model complexity (i.e. selecting the optimal number of components that best describes the data). Approaches based on frequentist learning (e.g. maximum likelihood) are generally prone to severe over-fitting and are fraught with difficulties for applying them in practice especially for high-dimensional data.
An alternative to parametric modeling and is the Bayesian nonparametric approach which allows the complexity of models to increase with data size [2]. One of the most popular Bayesian nonparametric models is the Dirichlet process (DP) [2] model which is often considered as an infinite mixture. Interest in infinite mixture models has grown considerably in recent years in domains such as image and signal processing, computer vision and machine learning [3]. In DP mixture modeling, the actual number of components used is not fixed and can be automatically inferred from the data set using a Bayesian posterior inference framework based, for instance, on Markov chain Monte Carlo (MCMC) techniques [4]. MCMC methods are well established and have been widely used for the learning of infinite mixture models, but are time-consuming and it is very difficult to assess their convergence. An efficient alternative to MCMC techniques is a deterministic approximation approach known as variational Bayes [5, 6]. Variational Bayes is based on analytical approximations to the posterior distribution and has been successfully applied for learning latent variable models in general and mixtures of distributions in particular [7, 8] due to its computational efficiency and convincing generalization power.
The purpose of this paper is to propose a Bayesian nonparametric approach for clustering based on Dirichlet processes mixtures. Rather than adopting Gaussian distributions as in many classic approaches, we use Beta-Liouville distributions to learn the underlying model of data. The motivation of employing Beta-Liouville distributions is due to its excellent modeling capabilities in the case of non-Gaussian data in general and proportional data (e.g. normalized histograms) in particular [9]. Our contributions are summarized as the following: first, we extend the finite Beta-Liouville mixture model into an infinite version using a stick-breaking construction [10]. Second, we develop a variational Bayes framework for learning the proposed model, such that the whole inference process is analytically tractable with closed-form solutions. Finally, we apply the proposed model on two challenging motion recognition tasks including hand gesture and human activity recognition.

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

The rest of this paper is organized as follows: Section 2 presents our infinite Beta-Liouville mixture model. In Section 3, we describe our variational Bayes framework for learning the proposed model. Section 4 presents the experimental results. Section 5 closes this paper with conclusions.

2. MODEL SPECIFICATION

2.1. Finite Beta-Liouville Mixture Model

Assume that a D-dimensional vector $\vec{X} = (X_1, \ldots, X_D)$ is distributed according to a Beta-Liouville distribution, then its probability density function (pdf) is defined by [9]:

$$\mathrm{BL}(\vec{X}|\vec{\theta}) = \frac{\Gamma(\sum_{l=1}^{D}\alpha_l)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \prod_{l=1}^{D} \frac{X_l^{\alpha_l-1}}{\Gamma(\alpha_l)} \times \left(\sum_{l=1}^{D} X_l\right)^{\alpha-\sum_{l=1}^{D}\alpha_l} \left(1-\sum_{l=1}^{D} X_l\right)^{\beta-1} \quad (1)$$

where $\vec{\theta} = (\alpha_1, \ldots, \alpha_D, \alpha, \beta)$ are the parameters of the Beta-Liouville distribution. Now let us consider a set of N vectors $\mathcal{X} = \{\vec{X}_1, \ldots, \vec{X}_N\}$, where each vector $\vec{X}_i = (X_{i1}, \ldots, X_{iD})$ is assumed to be generated from a finite Beta-Liouville mixture model with M components as

$$p(\vec{X}_i|\vec{\pi}, \vec{\theta}) = \sum_{j=1}^{M} \pi_j \mathrm{BL}(\vec{X}_i|\theta_j) \quad (2)$$

where $\theta_j = (\alpha_{j1}, \ldots, \alpha_{jD}, \alpha_j, \beta_j)$ are the parameters of a Beta-Liouville distribution corresponding to component j. $\vec{\pi} = (\pi_1, \ldots, \pi_M)$ in Eq. (2) represents the vector of mixing coefficients which are subject to the following constraints: $0 \le \pi_j \le 1$ and $\sum_{j=1}^{M} \pi_j = 1$.

2.2. Infinite Beta-Liouville Mixture Model

In this subsection, we extend the finite Beta-Liouville mixture model to the infinite case by exploiting a Dirichlet process (DP) framework. In our work, we construct the DP using a stick-breaking representation, which is defined as follows [10]: a random distribution G is distributed according to a DP with a base distribution H and concentration parameter ψ (denoted as $G \sim \mathrm{DP}(\psi, H)$), if the following requirements are satisfied:

$$\lambda_j \sim \mathrm{Beta}(1, \psi), \qquad \Omega_j \sim H$$
$$\pi_j = \lambda_j \prod_{k=1}^{j-1}(1-\lambda_k), \qquad G = \sum_{j=1}^{\infty} \pi_j \delta_{\Omega_j} \quad (3)$$

where δ_{Ω_j} denotes the Dirac delta measure centered at Ω_j, and π_j is the mixing proportion in terms of mixture modeling terminology and is defined by recursively breaking a unit length stick into an infinite number of pieces.

Assuming now that we have observed a dataset $\mathcal{X}$ which is generated from a Beta-Liouville mixture model with a countably infinite number of components. Then, the infinite Beta-Liouville mixture model can be written as

$$p(\vec{X}_i|\vec{\pi}, \vec{\theta}) = \sum_{j=1}^{\infty} \pi_j \mathrm{BL}(\vec{X}_i|\theta_j) \quad (4)$$

Next, a binary latent variable $\vec{Z}_i = (Z_{i1}, Z_{i2}, \ldots)$ is placed over each vector $\vec{X}_i$, such that $Z_{ij} \in \{0, 1\}$ and $Z_{ij} = 1$ if $\vec{X}_i$ belongs to component j and 0, otherwise. The prior distribution of latent variables $\mathcal{Z} = (\vec{Z}_1, \ldots, \vec{Z}_N)$ is given by

$$p(\mathcal{Z}|\vec{\pi}) = \prod_{i=1}^{N}\prod_{j=1}^{\infty} \pi_j^{Z_{ij}} \quad (5)$$

Notice that, $\vec{\pi}$ is a function of $\vec{\lambda}$ according to the stick-breaking construction of DP as shown in (3). Then, we have

$$p(\mathcal{Z}|\vec{\lambda}) = \prod_{i=1}^{N}\prod_{j=1}^{\infty}\left[\lambda_j \prod_{s=1}^{j-1}(1-\lambda_s)\right]^{Z_{ij}} \quad (6)$$

According to (3), the prior of $\vec{\lambda}$ is a specific Beta distribution:

$$p(\vec{\lambda}|\vec{\psi}) = \prod_{j=1}^{\infty} \mathrm{Beta}(1, \psi_j) = \prod_{j=1}^{\infty} \psi_j (1-\lambda_j)^{\psi_j - 1} \quad (7)$$

The next step is to place priors over parameters α_l, α and β in our Bayesian model. Since α_l, α and β are positive, Gamma distribution $\mathcal{G}(\cdot)$ is adopted to approximate conjugate priors for these parameters. Thus, we can obtain the following prior distributions for α_l, α and β, respectively: $p(\alpha_l) = \mathcal{G}(\alpha_l|u_l, v_l)$, $p(\alpha) = \mathcal{G}(\alpha|g, h)$, $p(\beta) = \mathcal{G}(\beta|s, t)$.

3. MODEL LEARNING VIA VARIATIONAL BAYES

In this section, we develop a variational Bayes framework for learning the infinite Beta-Liouville mixture model. To simplify notations, we define $\Theta = \{\mathcal{Z}, \vec{\lambda}, \vec{\theta}\}$. The main idea in variational learning is to find an approximation $Q(\Theta)$ for the posterior distribution $p(\Theta|\mathcal{X})$. In this work, we adopt factorial approximation [6] (or mean fields approximation [11]), which has been successfully applied in the past to complex models involving incomplete data, to factorize $Q(\Theta)$ into disjoint tractable factors. Furthermore, motivated by [12], we truncate the stick-breaking representation for the infinite Beta-Liouville mixture model at a value of M as: $\lambda_M = 1$, $\pi_j = 0$ when $j > M$, $\sum_{j=1}^{M} \pi_j = 1$. Here, the truncation level M is a variational parameter which can be freely initialized and will be optimized automatically during the learning process. In variational Bayes learning, the general expression for updating a variational factor is given by [6]:

$$Q_a(\Theta_a) = \frac{\exp\langle \ln p(\mathcal{X}, \Theta)\rangle_{\neq a}}{\int \exp\langle \ln p(\mathcal{X}, \Theta)\rangle_{\neq a} d\Theta} \quad (8)$$

where $\langle \cdot \rangle_{\neq a}$ denotes an expectation with respect to all the factor distributions except for a. Then, we can obtain the variational solution for each factor as

$$Q(\mathcal{Z}) = \prod_{i=1}^{N}\prod_{j=1}^{M} r_{ij}^{Z_{ij}}, \quad Q(\vec{\lambda}) = \prod_{j=1}^{M} \mathrm{Beta}(\lambda_j | c_j, d_j) \tag{9}$$

$$Q(\vec{\alpha}_l) = \prod_{j=1}^{M}\prod_{l=1}^{D} \mathcal{G}(\alpha_{jl}|u_{jl}^*, v_{jl}^*) \tag{10}$$

$$Q(\vec{\alpha}) = \prod_{j=1}^{M} \mathcal{G}(\alpha_j | g_j^*, h_j^*) \quad Q(\vec{\beta}) = \prod_{j=1}^{M} \mathcal{G}(\beta_j | s_j^*, t_j^*) \tag{11}$$

where we have

$$r_{ij} = \frac{\widetilde{r}_{ij}}{\sum_{j=1}^{M} \widetilde{r}_{ij}}, \quad c_j = 1+\sum_{i=1}^{N}\langle Z_{ij}\rangle, \quad d_j = \psi_j + \sum_{i=1}^{N}\sum_{k=j+1}^{N}\langle Z_{ik}\rangle \tag{12}$$

$$\widetilde{r}_{ij} = \exp\Big[\widetilde{\mathcal{I}}_j + \widetilde{\mathcal{H}}_j + (\bar{\alpha}_j - \sum_{l=1}^{D}\bar{\alpha}_{jl})\ln(\sum_{l=1}^{D} X_{il}) + \sum_{k=1}^{j-1}\langle \ln(1-\lambda_k)\rangle + \sum_{l=1}^{D}(\bar{\alpha}_{jl}-1)\ln X_{il} + (\bar{\beta}_j - 1)\ln(1-\sum_{l=1}^{D} X_{il}) + \langle \ln \lambda_j\rangle\Big] \tag{13}$$

$$v_{jl}^* = v_{jl} - \sum_{i=1}^{N}\langle Z_{ij}\rangle\big[\ln X_{il} - \ln(\sum_{l=1}^{D} X_{il})\big] \tag{14}$$

$$t_j^* = t_j - \sum_{i=1}^{N}\langle Z_{ij}\rangle \ln(1-\sum_{l=1}^{D} X_{il}) \tag{15}$$

$$h_j^* = h_j - \sum_{i=1}^{N}\langle Z_{ij}\rangle \ln(\sum_{l=1}^{D} X_{il}) \tag{16}$$

$$u_{jl}^* = u_{jl} + \sum_{i=1}^{N}\langle Z_{ij}\rangle \bar{\alpha}_{jl}\Big[\Psi(\sum_{l=1}^{D}\bar{\alpha}_{jl}) + \Psi'(\sum_{l=1}^{D}\bar{\alpha}_{jl})\sum_{d\neq l}^{D}(\langle \ln \alpha_{jd}\rangle - \ln\bar{\alpha}_{jd})\bar{\alpha}_{jd} - \Psi(\bar{\alpha}_{jl})\Big] \tag{17}$$

$$g_j^* = g_j + \sum_{i=1}^{N}\langle Z_{ij}\rangle\big[\bar{\beta}_j\Psi'(\bar{\alpha}_j+\bar{\beta}_j)(\langle \ln\beta_j\rangle - \ln\bar{\beta}_j) - \Psi(\bar{\alpha}_j) + \Psi(\bar{\alpha}_j+\bar{\beta}_j)\big]\bar{\alpha}_j \tag{18}$$

$$s_j^* = s_j + \sum_{i=1}^{N}\langle Z_{ij}\rangle\big[\bar{\alpha}_j\Psi'(\bar{\alpha}_j+\bar{\beta}_j)(\langle \ln\alpha_j\rangle - \ln\bar{\alpha}_j) - \Psi(\bar{\beta}_j) + \Psi(\bar{\alpha}_j+\bar{\beta}_j)\big]\bar{\beta}_j \tag{19}$$

where $\Psi(\cdot)$ is the digamma function. $\widetilde{\mathcal{I}}_j$ and $\widetilde{\mathcal{H}}_j$ in (13) are the lower bounds of $\mathcal{I}_j = \Big\langle \ln \frac{\Gamma(\sum_{l=1}^{D}\alpha_{jl})}{\prod_{l=1}^{D}\Gamma(\alpha_{jl})}\Big\rangle$ and $\mathcal{H}_j = \Big\langle \ln\frac{\Gamma(\alpha_j+\beta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)}\Big\rangle$, respectively. Since these expectations are analytically intractable, we adopt the second-order Taylor series expansion to compute their lower bounds. The expected values in the above formulas are defined as

$$\bar{\alpha}_{jl} = \frac{u_{jl}^*}{v_{jl}^*}, \qquad \bar{\alpha}_j = \frac{g_j^*}{h_j^*}, \qquad \bar{\beta}_j = \frac{s_j^*}{t_j^*} \tag{20}$$

$$\langle Z_{ij}\rangle = r_{ij}, \qquad \big\langle \ln \alpha_{jl}\big\rangle = \Psi(u_{jl}^*) - \ln v_{jl}^* \tag{21}$$

$$\big\langle \ln \alpha_j\big\rangle = \Psi(g_j^*) - \ln h_j^*, \quad \big\langle \ln\beta_j\big\rangle = \Psi(s_j^*) - \ln t_j^* \tag{22}$$

$$\big\langle \ln\lambda_j\big\rangle = \Psi(c_j) - \Psi(c_j+d_j), \quad \big\langle \ln(1-\lambda_j)\big\rangle = \Psi(d_j) - \Psi(c_j+d_j) \tag{23}$$

Since the solutions to each variational factor are coupled together through the expected values of other factors, the optimization of the model can be solved in a way analogous to the EM algorithm. The variational Bayes algorithm for learning infinite Beta-Liouville mixture models is summarized in Algorithm 1.

Algorithm 1

1: Choose the initial truncation level M.
2: Initialize the values for hyperparameters $\psi_j, u_{jl}, v_{jl}, g_j, h_j, s_j$ and t_j.
3: Initialize the values of r_{ij} by K-Means algorithm.
4: **repeat**
5: *The variational E-step*:
6: Estimate the expected values in (20)∼(23), use the current distributions over the model parameters.
7: *The variational M-step*:
8: Update the variational solutions for each factor using (9), (10) and (11) with the current values of the moments.
9: **until** Convergence criterion is reached.
10: Compute the expected value of λ_j as $\langle\lambda_j\rangle = c_j/(c_j+d_j)$ and substitute it into (3) to obtain the estimated values of the mixing coefficients π_j.
11: Detect the optimal number of components M by eliminating the components with small mixing coefficients close to 0.

4. EXPERIMENTAL RESULTS

We test the effectiveness of the proposed variational infinite Beta-Liouville mixture model (*InBLM*) on a challenging task namely spatio-temporal object (or motion) recognition. In our work, motion recognition was performed in two different domains including hand gesture and human activity recognition. We adopted a recently proposed interest point detector which has shown promising performance in motion recognition [13]. It is known as the NNMF interest point detector since it is based on non-negative matrix factorisation (NNMF). Compared with other popular spatiotemporal interest point detectors which use local information only, the NNMF interest point detector exploits global information from each video input and has demonstrated better results in motion recognition according to [13]. In all of our experiments, we initialize the truncation level M and the hyperparameter ψ_j to 20 and 0.1, respectively. In order to provide broad non-informative prior distributions, the initial values of hyperparameters u_{jl}, g_j, s_j of the Gamma priors are set to 1, and v_{jl}, h_j, t_j are set to 0.01. In order to show the merits of our approach, we have compared our approach with four other well-defined mixture-modeling approaches: the finite

Beta-Liouville mixture model (*FiBLM*), the infinite generalized Dirichlet mixture model (*InGDM*), the infinite Dirichlet mixture model (*InDM*) and the infinite Gaussian mixture model (*InGM*) [12]. To have a fair comparison, all of these approaches are learned by variational Bayes.

4.1. Experimental Design

The methodology that we have adopted for motion recognition can be summarized as follows. First, we detect spatiotemporal interest points and their locations from each video using the NNMF interest point detector. In our case, the spatial and temporal scales are set to 2 and 4, respectively. Next, K-Means algorithm is applied to quantize the obtained interest point features to form a visual vocabulary and each cluster center is treated as a visual word. Applying the paradigm of *bag-of-words*, a histogram representing the frequency of each visual word is calculated for each video. Then, we apply the probabilistic latent semantic analysis (pLSA) model [14] to reduce the dimensionality of the resulting histograms which allows the description of each video as a vector of proportions (50-dimensional). Finally, we employ the proposed *InBLM* as a classifier to recognize motions by assigning the testing video to the category which has the highest posterior probability according to Bayes' decision rule. We run the algorithm 20 times to investigate its performance. The results of recognizing hand gestures and human activities by our approach and other comparable approaches will be described in the following subsections.

4.2. Hand Gesture Recognition

For hand gesture recognition, we adopted a publicly available database namely the Cambridge-Gesture database [1] [15]. It consists of 900 image sequences of 9 hand gesture classes, which are defined by 3 primitive hand shapes ("Flat", "Spread" and "V-shape") and 3 primitive motions (Leftward, rightward and contract). Each class contains 100 image sequences with a size of 320×240 pixels. Sample frames from this database can be viewed in Fig. 1. In our experiment, half of the data was randomly chosen for constructing the visual vocabulary and the other half was used for testing. Figure 2 shows the confusion matrix for the Cambridge-Gesture database using the *InBLM*. Table 1 illustrates the average recognition accuracy obtained by each tested approach. According to this table, it is obvious that the proposed *InBLM* gave the best performance in terms of the highest recognition accuracy rate (89.44%). The fact that *InBLM* outperformed *FiBLM* confirms the advantage of using infinite mixture model over finite mixture model. Moreover, we may notice that *InGM* provided the worst performance among all infinite mixture models. This result proves that

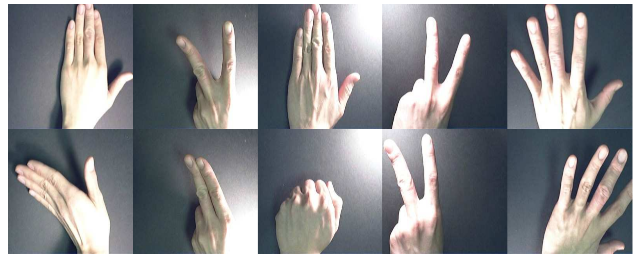

Fig. 1. Sample frames from the Cambridge-Gesture database.

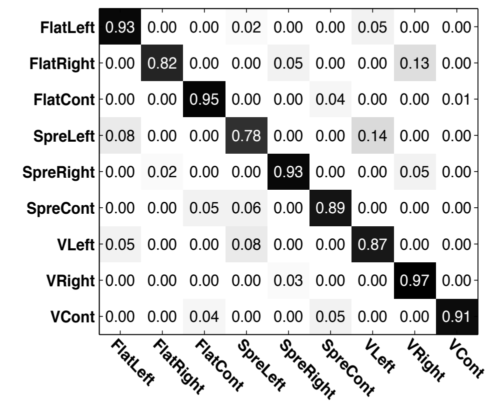

Fig. 2. Confusion matrix obtained by *InBLM* for the Cambridge-Gesture database.

Gaussian mixture model is not a good choice for handling proportional data.

Table 1. The average recognition rate (%) using different approaches for hand gestures and human activities.

Method	Hand Gesture	Human Activity
InBLM	89.44 ± 0.75	85.90 ± 0.85
FiBLM	85.98 ± 0.82	83.15 ± 1.12
InGDM	86.37 ± 0.78	83.09 ± 0.95
InDM	84.29 ± 0.91	80.30 ± 1.02
InGM	82.18 ± 1.04	77.56 ± 1.19

4.3. Human Activity Recognition

In this subsection, we have applied our approach on the Weizmann human action database [16] for recognizing human activities. It contains of 90 video sequences at a resolution of 180×144 pixels. Ten different types of human actions are performed by nine subjects. Some examples of frames from each action class are displayed in Fig. 3. In our case, we use a leave-one-out setup to test the performance of our approach. Specifically, we construct our visual vocabulary from the video sequences of eight subjects and perform testing on

[1]http://www.iis.ee.ic.ac.uk/icvl/ges_db.htm

the sequences of the remaining subject. The confusion matrix for the Weizmann human action database using the *InBLM* is demonstrated in Fig. 4. The overall accuracy is around 85.90%. As we can see, most errors are generated from similar action categorizes, such as "run" with "walk", "jump" with "skip", "skip" with "jump" and "run". The average recognition rate by each testing approach is shown in Table 1. It is clear that *InBLM* outperforms other approaches by providing the highest recognition accuracy rate.

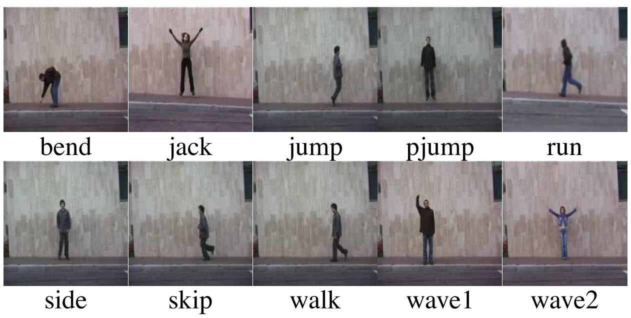

Fig. 3. Sample frames from the Weizmann database.

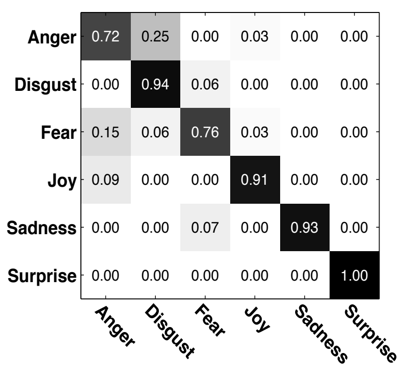

Fig. 4. Confusion matrix obtained by *InBLM* for the Weizmann human action database.

5. CONCLUSION

In this paper we have proposed a statistical framework that builds on recent developments in infinite mixture models and variational inference. More specifically, we have developed an infinite Beta-Liouville mixture which can provide a practical solution to the challenging problem of model selection. It is learnt using a variational Bayes framework in which the whole inference process is analytically tractable with closed-form solutions. The considered variational inference is computationally efficient and offers a deterministic effective alternative to fully Bayesian inference by maximizing a lower bound on the marginal likelihood. The merits of the proposed approach has been tested on two challenging motion recognition tasks namely hand gesture and human activity recognition.

6. REFERENCES

[1] G.J. McLachlan and D. Peel, *Finite Mixture Models*, New York: Wiley, 2000.

[2] T. S. Ferguson, "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, vol. 1, no. 2, pp. 209–230, 1973.

[3] N. Bouguila and D. Ziou, "A Dirichlet Process Mixture of Dirichlet Distributions for Classification and Prediction," in *Proc. of the IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, 2008, pp. 297–302.

[4] N. Bouguila, "Infinite Liouville Mixture Models with Application to Text and Texture Categorization," *Pattern Recognition Letters*, vol. 33, no. 2, pp. 103–110, 2012.

[5] H. Attias, "A Variational Bayes Framework for Graphical Models," in *Proc. of Advances in Neural Information Processing Systems (NIPS)*, 1999, pp. 209–215.

[6] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[7] A. Corduneanu and C. M. Bishop, "Variational Bayesian Model Selection for Mixture Distributions," in *Proc. of the International Conference on Artificial Intelligence and Statistics (AISTAT)*, 2001, pp. 27–34.

[8] W. Fan, N. Bouguila, and D. Ziou, "Variational Learning for Finite Dirichlet Mixture Models and Applications," *IEEE Trans. on Neural Netw. Learning Syst.*, vol. 23, no. 5, pp. 762–774, 2012.

[9] N. Bouguila, "Hybrid Generative/Discriminative Approaches for Proportional Data Modeling and Classification," *IEEE Trans. on Knowledge and Data Engineering*, vol. 24, no. 12, pp. 2184–2202, 2012.

[10] J. Sethuraman, "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, vol. 4, pp. 639–650, 1994.

[11] B. Wang and D.M. Titterington, "Lack of consistency of mean field and variational bayes approximations for state space models," *Neural Processing Letters*, vol. 20, pp. 151–170, 2004.

[12] D. M. Blei and M. I. Jordan, "Variational Inference for Dirichlet Process Mixtures," *Bayesian Analysis*, vol. 1, pp. 121–144, 2005.

[13] S. Wong and R. Cipolla, "Extracting Spatiotemporal Interest Points using Global Information," in *Proc. of IEEE International Conference on Computer Vision, (ICCV)*, 2007, pp. 1–8.

[14] T. Hofmann, "Unsupervised Learning by Probabilistic Latent Semantic Analysis," *Machine Learning*, vol. 42, no. 1/2, pp. 177–196, 2001.

[15] T. Kim, S. Wong, and R. Cipolla, "Tensor Canonical Correlation Analysis for Action Classification," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8.

[16] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2005, pp. 1395–1402.