A Hierarchical Dirichlet Process Mixture of GID Distributions with Feature Selection for Spatio-Temporal Video Modeling and Segmentation

Wentao Fan*, Nizar Bouguila† and Xin Liu‡
*Department of Computer Science and Technology, Huaqiao University, Xiamen, China
Email: fwt@hqu.edu.cn
†Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada
Email: nizar.bouguila@concordia.ca
‡Department of Computer Science and Technology, Huaqiao University, Xiamen, China
Email: xliu@hqu.edu.cn

***Abstract*—In this paper, a hierarchical Dirichlet process (HDP) mixture model of generalized inverted Dirichlet (GID) distributions with an unsupervised feature selection scheme is developed. The proposed model is learned via a principled variational framework and then deployed for video modeling and segmentation. Experimental results show the merits of our developed statistical framework.**

***Keywords*-Mixture models, Dirichlet process, feature selection, variational learning, video segmentation.**

I. Introduction

Semantic video segmentation is an important step in many applications and necessitates the development of strong machine learning techniques [1], [2], [3]. The main goal is to automatically partition video sequences into spatiotemporal segments. Several approaches have been proposed in the past [4]. In this paper, we approach this problem by developing a framework based on HDP mixture model, which is a hierarchical nonparametric Bayesian framework that has shown promising performance in clustering of grouped data with sharing clusters [5], [6]. This model is particularly useful in many real-world problems where one cluster may be highly overlapped or even could be embedded into another cluster. A HDP mixture model is described as follows: Suppose that we have collected N observations that are organized into M groups, for each observation X_{ji} that is drawn independently from a mixture model and, we associate a factor θ_{ji}, where the index ji indicates the observation i within group j. In order to form a Bayesian approach, each factor θ_{ji} is distributed according to a prior G_j. Then, we have

$$\theta_{ji}|G_j \sim G_j, \qquad X_{ji}|\theta_{ji} \sim F(\theta_{ji}) \tag{1}$$

where $F(\theta_{ji})$ denotes the probability distribution of X_{ji} given θ_{ji}. The prior G_j is distributed according to the HDP model, which is built on the Dirichlet process (DP) [7] that contains a Bayesian hierarchy where the base measure of a DP is itself a drawn from a DP:

$$\begin{aligned} G_0 &\sim \mathrm{DP}(\gamma, H) \\ G_j &\sim \mathrm{DP}(\lambda, G_0), \qquad \text{for each } j \in \{1, \ldots, M\} \end{aligned} \tag{2}$$

where each group is associated with a group-level DP G_j, and this indexed set of DPs $\{G_j\}$ shares a common base (i.e. a global-level) distribution G_0. A crucial problem when using such models is the choice of a parent distribution and the selection of the relevant modeling features. In this paper we propose a principled approach to tackle the video segmentation problem by considering the GID that has been shown to provide a principled approach for simultaneous clustering and feature selection. The resulting model is learned within a variational framework that we have developed. The rest of this paper is organized as follows. Section II presents our model. Section III is devoted to the experimental results. The conclusion is given in Section III.

II. HDP Mixture of GID Distributions

In the global-level, the global measure G_0 follows the Dirichlet process $\mathrm{DP}(\gamma, H)$ and can be described using the stick-breaking representation [8], [9] as

$$\begin{aligned} \xi'_k &\sim \mathrm{Beta}(1, \gamma), \qquad \Lambda_k \sim H \\ \xi_k &= \xi'_k \prod_{s=1}^{k-1}(1-\xi'_s), \qquad G_0 = \sum_{k=1}^{\infty} \xi_k \delta_{\Lambda_k} \end{aligned} \tag{3}$$

where $\{\Lambda_k\}$ is a set of independent random variables drawn from H, δ_{Λ_k} is an atom centered at Λ_k. The stick-breaking weights ξ_k satisfy the constraint that $\sum_{k=1}^{\infty} \xi_k = 1$. Since G_0 is the base measure of G_j, the atoms Λ_k are therefore shared among all G_j and are only differ in weights. Next, we construct each group-level DP G_j:

$$\begin{aligned} \pi'_{jt} &\sim \mathrm{Beta}(1, \lambda), \qquad \Omega_{jt} \sim G_0 \\ \pi_{jt} &= \pi'_{jt} \prod_{s=1}^{t-1}(1-\pi'_{js}), \qquad G_j = \sum_{t=1}^{\infty} \pi_{jt} \delta_{\Omega_{jt}} \end{aligned} \tag{4}$$

where $\delta_{\Omega_{jt}}$ are group-level atoms centered at Ω_{jt}, $\{\pi_{jt}\}$ is a set of stick-breaking weights which satisfies $\sum_{t=1}^{\infty} \pi_{jt} = 1$. Since Ω_{jt} is distributed according to the base distribution G_0, it takes on the value Λ_k with probability ξ_k. Thus, it is straightforward to introduce a binary latent variable W_{jtk} as

an indicator variable, such that $W_{jtk} \in \{0,1\}$, $W_{jtk} = 1$ if ϖ_{jt} maps to the base-level atom Λ_k; otherwise, $W_{jtk} = 0$. Thus, we have $\Omega_{jt} = \Lambda_k^{W_{jtk}}$. The probability distribution of the indicator variable $\vec{W} = (W_{jt1}, W_{jt2}, \ldots)$ is

$$p(\vec{W}) = \prod_{j=1}^{M}\prod_{t=1}^{\infty}\prod_{k=1}^{\infty}\xi_k^{W_{jtk}} = \prod_{j=1}^{M}\prod_{t=1}^{\infty}\prod_{k=1}^{\infty}\Big[\xi'_k\prod_{s=1}^{k-1}(1-\xi'_s)\Big]^{W_{jtk}}$$

Using Eq. 3, the prior over $\vec{\xi}'$ is $p(\vec{\xi}') = \prod_{k=1}^{\infty}\mathrm{Beta}(1,\gamma_k) = \prod_{k=1}^{\infty}\gamma_k(1-\xi'_k)^{\gamma_k-1}$. Another binary latent variable $Z_{jit} \in \{0,1\}$ is introduced as an indicator variable, such that $Z_{jit} = 1$ if θ_{ji} is associated with component t and maps to the group-level atom Ω_{jt}; otherwise, $Z_{jit} = 0$. Thus, we have $\theta_{ji} = \Omega_{jt}^{Z_{jit}}$. Since Ω_{jt} maps to the global-level atom Λ_k as well, we can write $\theta_{ji} = \Omega_{jt}^{Z_{jit}} = \Lambda_k^{W_{jtk}Z_{jit}}$. The probability distribution of the indicator variable $\vec{Z} = (Z_{ji1}, Z_{ji2}, \ldots)$ is given by

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

According to the stick-breaking construction:

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

Given a D-dimensional random vector $\vec{Y} = (Y_1, \ldots, Y_D)$ which is distributed according to a GID with parameters $\vec{\alpha} = (\alpha_1, \ldots, \alpha_D)$ and $\vec{\beta} = (\beta_1, \ldots, \beta_D)$, then its pdf is [10], [11]

$$\mathrm{GID}(\vec{Y}|\vec{\alpha},\vec{\beta}) = \prod_{l=1}^{D}\frac{\Gamma(\alpha_l+\beta_l)}{\Gamma(\alpha_l)\Gamma(\beta_l)}\frac{Y_l^{\alpha_l-1}}{(1+\sum_{l=1}^{D}Y_l)^{\vartheta_l}} \quad (5)$$

where $\vartheta_l = \alpha_l + \beta_l - \beta_{l+1}$ for $l = 1, \ldots, D$, and $\beta_{l+1} = 0$. $\Gamma(\cdot)$ is the gamma function. Next, as discussed in [10], we can transform the data vector $\vec{Y}$ into another D-dimensional data point $\vec{X}$ with independent features, through the geometric transformation: $X_1 = Y_1$ and $X_l = Y_l/(1+\sum_{s=1}^{l-1}Y_s)$ for $l > 1$. Then, the estimation of a D-dimensional GID is transformed to D estimations of inverted Beta distributions [12], [13] $\mathrm{GID}(\vec{X}|\vec{\alpha},\vec{\beta}) = \prod_{l=1}^{D}\mathrm{IB}(X_l|\alpha_l,\beta_l)$, where $\mathrm{IB}(X_l|\alpha_l,\beta_l)$ is an inverted Beta with parameters $\{\alpha_l,\beta_l\}$:

$$\mathrm{IB}(X_l|\alpha_l,\beta_l) = \frac{\Gamma(\alpha_l+\beta_l)}{\Gamma(\alpha_l)\Gamma(\beta_l)}X_l^{\alpha_l-1}(1+X_l)^{-(\alpha_l+\beta_l)} \quad (6)$$

Assume that there is an observed data set $\mathcal{X}$ that contains N D-dimensional random vectors grouped into M groups, where each vector $\vec{X}_{ji} = (X_{ji1}, \ldots, X_{jiD})$ is drawn from our hierarchical model. Then, the likelihood of the model is

$$p(\mathcal{X}) = \prod_{j=1}^{M}\prod_{i=1}^{N}\prod_{t=1}^{\infty}\prod_{k=1}^{\infty}\Big[\prod_{l=1}^{D}\mathrm{IB}(X_{jil}|\alpha_{kl},\beta_{kl})\Big]^{Z_{jit}W_{jtk}}$$

We integrate a feature selection scheme [14], [15], [16], [17], [18], so that an irrelevant feature is defined as the one having a distribution independent from class labels:
$p(X_{jil}) = \mathrm{IB}(X_{jil}|\alpha_{kl},\beta_{kl})^{\phi_{jil}}\mathrm{IB}(X_{jil}|\alpha'_l,\beta'_l)^{1-\phi_{jil}}$,
where ϕ_{jil} is a binary latent variable indicates the feature relevance. The prior of $\vec{\phi}$ is given by $p(\vec{\phi}|\vec{\epsilon}) = \prod_{j=1}^{M}\prod_{i=1}^{N}\prod_{l=1}^{D}\epsilon_{l_1}^{\phi_{jil}}\epsilon_{l_2}^{1-\phi_{jil}}$, where $\vec{\epsilon} = (\vec{\epsilon}_1, \ldots, \vec{\epsilon}_D)$ denotes the features saliencies such that $\vec{\epsilon}_l = (\epsilon_{l_1}, \epsilon_{l_2})$ and $\epsilon_{l_1} + \epsilon_{l_2} = 1$. The prior of $\vec{\epsilon}$ is

$$p(\vec{\epsilon}) = \prod_{l=1}^{D}\mathrm{Dir}(\vec{\epsilon}_l|\vec{\zeta}) = \prod_{l=1}^{D}\frac{\Gamma(\zeta_1+\zeta_2)}{\Gamma(\zeta_1)\Gamma(\zeta_2)}\epsilon_{l_1}^{\zeta_1-1}\epsilon_{l_2}^{\zeta_2-1} \quad (7)$$

Then, the likelihood can be written as

$$p(\mathcal{X}|\vec{Z},\vec{W},\vec{\theta},\vec{\phi}) = \prod_{j=1}^{M}\prod_{i=1}^{N}\prod_{t=1}^{\infty}\prod_{k=1}^{\infty}\Big[\prod_{l=1}^{D}\mathrm{IB}(X_{jil}|\alpha_{kl},\beta_{kl})^{\phi_{jil}} \times\mathrm{IB}(X_{jil}|\alpha'_l,\beta'_l)^{(1-\phi_{jil})}\Big]^{Z_{jit}W_{jtk}} \quad (8)$$

where $\vec{\theta} = \{\vec{\alpha},\vec{\beta},\vec{\alpha}',\vec{\beta}'\}$. For parameters $\vec{\alpha}$, $\vec{\beta}$, $\vec{\alpha}'$ and $\vec{\beta}'$, we adopt Gamma distributions as their priors:

$$p(\vec{\alpha}) = \mathcal{G}(\vec{\alpha}|\vec{u},\vec{v}), \qquad p(\vec{\beta}) = \mathcal{G}(\vec{\beta}|\vec{g},\vec{h})$$
$$p(\vec{\alpha}') = \mathcal{G}(\vec{\alpha}'|\vec{u}',\vec{v}'), \qquad p(\vec{\beta}') = \mathcal{G}(\vec{\beta}'|\vec{g}',\vec{h}')$$

A truncation technique [19], is adopted at K and T as: $\xi'_K = 1$, $\sum_{k=1}^{K}\xi_k = 1$, $\xi_k = 0$ when $k > K$; $\pi'_{jT} = 1$, $\sum_{t=1}^{T}\pi_{jt} = 1$, $\pi_{jt} = 0$ when $t > T$. We also adopt factorial approximation [20]:

$$q(\Theta) = q(\vec{Z})q(\vec{W})q(\vec{\phi})q(\vec{\pi}')q(\vec{\xi}')q(\vec{\alpha})q(\vec{\beta})q(\vec{\alpha}')q(\vec{\beta}')q(\vec{\epsilon})$$

Then, we obtain the update equations by maximizing the lower bound $\mathcal{L}(q)$ with respect to each of the factors:

$$q(\vec{Z}) = \prod_{j=1}^{M}\prod_{i=1}^{N}\prod_{t=1}^{T}\rho_{jit}^{Z_{jit}}, \quad q(\vec{W}) = \prod_{j=1}^{M}\prod_{t=1}^{T}\prod_{k=1}^{K}\sigma_{jtk}^{W_{jtk}}$$

$$q(\vec{\phi}) = \prod_{j=1}^{M}\prod_{i=1}^{N}\prod_{l=1}^{D}\varphi_{jil}^{\phi_{jil}}(1-\varphi_{jil})^{1-\phi_{jil}}$$

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

$$q(\vec{\epsilon}) = \prod_{l=1}^{D}\mathrm{Dir}(\vec{\epsilon}_l|\vec{\zeta}^*), \quad q(\vec{\xi}') = \prod_{k=1}^{K}\mathrm{Beta}(\xi'_k|c_k,d_k)$$

$$q(\vec{\alpha}) = \prod_{k=1}^{K}\prod_{l=1}^{D}\mathcal{G}(\alpha_{kl}|\tilde{u}_{kl},\tilde{v}_{kl}), \quad q(\vec{\alpha}') = \prod_{l=1}^{D}\mathcal{G}(\alpha'_l|\tilde{u}'_l,\tilde{v}'_l)$$

$$q(\vec{\beta}) = \prod_{k=1}^{K}\prod_{l=1}^{D}\mathcal{G}(\beta_{kl}|\tilde{g}_{kl},\tilde{h}_{kl}), \quad q(\vec{\beta}') = \prod_{l=1}^{D}\mathcal{G}(\beta'_l|\tilde{g}'_l,\tilde{h}'_l)$$

where the associated hyperparameters are updated as

$$\rho_{jit} = \frac{\exp(\widetilde{\rho}_{jit})}{\sum_{s=1}^{T}\exp(\widetilde{\rho}_{jis})}, \qquad \sigma_{jtk} = \frac{\exp(\widetilde{\sigma}_{jtk})}{\sum_{s=1}^{K}\exp(\widetilde{\sigma}_{jts})}, \tag{9}$$

$$\widetilde{\rho}_{jit} = \sum_{k=1}^{K}\langle W_{jtk}\rangle \sum_{l=1}^{D}\langle\phi_{jil}\rangle[\langle\ln\frac{\Gamma(\alpha_{kl}+\beta_{kl})}{\Gamma(\alpha_{kl})\Gamma(\beta_{kl})}\rangle + (\bar{\alpha}_{kl}-1)\ln X_{jil} - (\bar{\alpha}_{kl}+\bar{\beta}_{kl})\ln(1+X_{jil})] + \langle\ln\pi'_{jt}\rangle + \sum_{s=1}^{t-1}\langle\ln(1-\pi'_{js})\rangle \tag{10}$$

$$\widetilde{\sigma}_{jtk} = \sum_{i=1}^{N}\langle Z_{jit}\rangle \sum_{l=1}^{D}\langle\phi_{jil}\rangle[\langle\ln\frac{\Gamma(\alpha_{kl}+\beta_{kl})}{\Gamma(\alpha_{kl})\Gamma(\beta_{kl})}\rangle + (\bar{\alpha}_{kl}-1)\ln X_{jil} - (\bar{\alpha}_{kl}+\bar{\beta}_{kl})\ln(1+X_{jil})] + \langle\ln\xi'_{k}\rangle + \sum_{s=1}^{k-1}\langle\ln(1-\xi'_{s})\rangle \tag{11}$$

$$\varphi_{jil} = \frac{\exp(\widetilde{\varphi}_{jil})}{\exp(\widetilde{\varphi}_{jil})+\exp(\widehat{\varphi}_{jil})} \tag{12}$$

$$\widetilde{\varphi}_{jil} = \langle\ln\epsilon_{l_1}\rangle + \sum_{t=1}^{T}\sum_{k=1}^{K}\langle Z_{jit}\rangle\langle W_{jtk}\rangle[\langle\ln\frac{\Gamma(\alpha_{kl}+\beta_{kl})}{\Gamma(\alpha_{kl})\Gamma(\beta_{kl})}\rangle + (\bar{\alpha}_{kl}-1)\ln X_{jil} - (\bar{\alpha}_{kl}+\bar{\beta}_{kl})\ln(1+X_{jil})] \tag{13}$$

$$\widehat{\varphi}_{jil} = (\bar{\alpha}'_{l}-1)\ln X_{jil} - (\bar{\alpha}'_{l}+\bar{\beta}'_{l})\ln(1+X_{jil}) + \langle\ln\epsilon_{l_2}\rangle + \langle\ln\frac{\Gamma(\alpha'_{l}+\beta'_{l})}{\Gamma(\alpha'_{l})\Gamma(\beta'_{l})}\rangle \tag{14}$$

$$\zeta_1^* = \zeta_1 + \sum_{j=1}^{M}\sum_{i=1}^{N}\langle\phi_{jil}\rangle, \quad \zeta_2^* = \zeta_2 + \sum_{j=1}^{M}\sum_{i=1}^{N}\langle 1-\phi_{jil}\rangle \tag{15}$$

$$a_{jt} = 1 + \sum_{i=1}^{N}\langle Z_{jit}\rangle, \qquad b_{jt} = \lambda_{jt} + \sum_{i=1}^{N}\sum_{s=t+1}^{T}\langle Z_{jis}\rangle \tag{16}$$

$$c_k = 1 + \sum_{j=1}^{K}\sum_{t=1}^{T}\langle W_{jtk}\rangle, \; d_k = \gamma_k + \sum_{j=1}^{M}\sum_{t=1}^{T}\sum_{s=k+1}^{K}\langle W_{jts}\rangle \tag{17}$$

$$\tilde{u}_{kl} = u_{kl} + \sum_{j=1}^{M}\sum_{t=1}^{T}\langle W_{jtk}\rangle\sum_{i=1}^{N}\langle Z_{jit}\rangle\langle\phi_{jil}\rangle\bar{\alpha}_{kl}\big[\Psi(\bar{\alpha}_{kl}+\bar{\beta}_{kl}) - \Psi(\bar{\alpha}_{kl}) + \bar{\beta}_{kl}\Psi'(\bar{\alpha}_{kl}+\bar{\beta}_{kl})(\langle\ln\beta_{kl}\rangle - \ln\bar{\beta}_{kl})\big] \tag{18}$$

$$\tilde{v}_{kl} = v_{kl} - \sum_{j=1}^{M}\sum_{t=1}^{T}\langle W_{jtk}\rangle\sum_{i=1}^{N}\langle Z_{jit}\rangle\langle\phi_{jil}\rangle\ln\frac{X_{jil}}{1+X_{jil}} \tag{19}$$

$$\tilde{g}_{kl} = g_{kl} + \sum_{j=1}^{M}\sum_{t=1}^{T}\langle W_{jtk}\rangle\sum_{i=1}^{N}\langle Z_{jit}\rangle\langle\phi_{jil}\rangle\bar{\beta}_{kl}\big[\Psi(\bar{\alpha}_{kl}+\bar{\beta}_{kl}) - \Psi(\bar{\beta}_{kl}) + \bar{\alpha}_{kl}\Psi'(\bar{\alpha}_{kl}+\bar{\beta}_{kl})(\langle\ln\alpha_{kl}\rangle - \ln\bar{\alpha}_{kl})\big] \tag{20}$$

$$\tilde{h}_{kl} = h_{kl} - \sum_{j=1}^{M}\sum_{t=1}^{T}\langle W_{jtk}\rangle\sum_{i=1}^{N}\langle Z_{jit}\rangle\langle\phi_{jil}\rangle\ln\frac{1}{1+X_{jil}} \tag{21}$$

$$\tilde{u}'_{l} = u'_{l} + \sum_{j=1}^{M}\sum_{i=1}^{N}\langle 1-\phi_{jil}\rangle\bar{\alpha}'_{l}\big[\Psi(\bar{\alpha}'_{l}+\bar{\beta}'_{l}) - \Psi(\bar{\alpha}'_{l}) + \bar{\beta}'_{l}\Psi'(\bar{\alpha}'_{l}+\bar{\beta}'_{l})(\langle\ln\beta'_{l}\rangle - \ln\bar{\beta}'_{l})\big] \tag{22}$$

$$\tilde{v}'_{l} = v'_{l} - \sum_{j=1}^{M}\sum_{i=1}^{N}\langle 1-\phi_{jil}\rangle\ln\frac{X_{jil}}{1+X_{jil}} \tag{23}$$

$$\tilde{g}'_{l} = g'_{l} + \sum_{j=1}^{M}\sum_{i=1}^{N}\langle 1-\phi_{jil}\rangle\bar{\beta}'_{l}\big[\Psi(\bar{\alpha}'_{l}+\bar{\beta}'_{l}) - \Psi(\bar{\beta}'_{l}) + \bar{\alpha}'_{l}\Psi'(\bar{\alpha}'_{l}+\bar{\beta}'_{l})(\langle\ln\alpha'_{l}\rangle - \ln\bar{\alpha}'_{l})\big] \tag{24}$$

$$\tilde{h}'_{l} = h'_{l} - \sum_{j=1}^{M}\sum_{i=1}^{N}\langle 1-\phi_{jil}\rangle\ln\frac{1}{1+X_{jil}} \tag{25}$$

where $\Psi(\cdot)$ is the digamma and expected values are:

$$\bar{\alpha}_{kl} = \frac{\tilde{u}_{kl}}{\tilde{v}_{kl}}, \quad \bar{\beta}_{kl} = \frac{\tilde{g}_{kl}}{\tilde{h}_{kl}}, \quad \bar{\alpha}'_{l} = \frac{\tilde{u}'_{l}}{\tilde{v}'_{l}}, \quad \bar{\beta}'_{l} = \frac{\tilde{g}'_{l}}{\tilde{h}'_{l}}$$
$$\langle Z_{jit}\rangle = \rho_{jit}, \quad \langle W_{jtk}\rangle = \sigma_{jtk}, \quad \langle\phi_{jil}\rangle = \varphi_{jil}$$
$$\langle\ln\alpha_{kl}\rangle = \Psi(\tilde{u}_{kl}) - \ln\tilde{v}_{kl}, \quad \langle\ln\beta_{kl}\rangle = \Psi(\tilde{g}_{kl}) - \ln\tilde{h}_{kl}$$
$$\langle\ln\alpha'_{l}\rangle = \Psi(\tilde{u}'_{l}) - \ln\tilde{v}'_{l}, \quad \langle\ln\beta'_{l}\rangle = \Psi(\tilde{g}'_{l}) - \ln\tilde{h}'_{l}$$
$$\langle\ln\epsilon_{l_1}\rangle = \Psi(\zeta_1^*) - \Psi(\zeta_1^*+\zeta_2^*), \; \langle\ln\epsilon_{l_2}\rangle = \Psi(\zeta_2^*) - \Psi(\zeta_1^*+\zeta_2^*)$$
$$\langle\ln\pi'_{jt}\rangle = \Psi(a_{jt}) - \Psi(a_{jt}+b_{jt}), \; \langle\ln\xi'_{k}\rangle = \Psi(c_k) - \Psi(c_k+d_k)$$

III. Experimental Results

We initialize K and T, in our model (referred to as *HDP-GID*), to 0.85 and 0.64, respectively. The hyperparameters of the feature saliency ζ_1 and ζ_2 are both initialized to 0.5. The hyperparameters of the stick-breaking weights λ_{jt} and γ_k are initialized to 0.25, and we set $(u_{kl}, v_{kl}, g_{kl}, h_{kl}, u'_l, v'_l, g'_l, h'_l) = (0.1, 0.05, 0.1, 0.05, 0.1, 0.05, 0.1, 0.05)$.

A. Video Segmentation Methodology

We adopted the idea of "frame saliency" as described in [4], so that only a subset of frames with high relevancy are used for model training. This is motivated by the observation obtained in [4] that only uses the most relevant frames may significantly reduce redundancy and improve the quality of video modeling. In our first step, for each pixel, we construct a feature vector that contains its three-dimensional color descriptor in the $L^*a^*b^*$ color space, the spatial information (i.e., the (x, y) position of the pixel) and the time feature (r) that indicates the number of frames in a video shot. Then, the obtained feature vectors are modeled using the proposed *HDP-GID* in which each frame $\mathcal{F}_j$ is considered as a "group" and is therefore associated with a Dirichlet process mixture model G_j. Therefore, the density function of $\vec{X}_{ji}$ (i.e. the ith pixel of jth frame) can be described as $p(\vec{X}_{ji}) = \sum_{k=1}^{\infty}\xi_k[\epsilon_j \mathrm{IB}(\vec{X}_{ji}|\vec{\alpha}_k, \vec{\beta}_k) + (1-\epsilon_j)\mathrm{IB}(\vec{X}_{ji}|\vec{\alpha}', \vec{\beta}')]$, where the frame saliency $\epsilon_j = p(\phi_j = 1)$ of frame j represents

the probability that frame j is highly relevant. By including the idea of frame saliency into the proposed HDP mixture model, we then have the video segmentation approach in which each mixture component [21], [22], [23] in a frame represents a segment and the components are shared among all frames.

Figure 1. Sample frames from each video sequence: First row: Video 1; Second row: Video 2

B. Results

Our experiments were conducted using two video sequences collected by [24]. Here, video 1 contains 323 frames, and video 2 has 250 frames in total. Each video has a resolution of 450 $\times$ 350. Sample frames from each video can be viewed in Fig. 1. For comparison, we also applied the HDP mixture of Gaussians with feature selection (referred to as *HDP-GMM*), and the one proposed in [4] that is based on the finite Gaussian mixture model and minimum description length (MDL) criterion (referred to as *GMM-MDL*). The performance of our approach was reported in terms of the objective criteria that is used in [4] from three aspects: **1) Spatial uniformity**: It measures the color homogeneity of video segments which includes the texture (color) variance of segments ($text\_var$) [25] and the spatial color contrast along segments boundaries ($color\_con$) [26]; **2) Temporal stability:** It evaluates the color and spatial homogeneity of the segments for consecutive time instants. Here, it was measured by the inter-frame difference of segment size and elongation ($size\_diff$ and $elong\_diff$) [25], and the $\mathcal{X}^2$ metric [26]; **3) Motion uniformity**: it measures the segments' motion smoothness which contains the summation of motion vector variance in x and y directions ($motion\_var$) [25] in this work. The average segmentation results from 20 runs are shown in Tables I and II for each tested video sequence . As shown in those tables, our approach (*HDP-GID*) was able to provide the best performance among all tested approaches in terms of smaller $text\_var$, $\mathcal{X}^2$, $motion\_var$ and larger $color\_con$. This fact verifies the advantages of using the GID mixture model over the Gaussian one as well as the merits of using the HDP mixture model over the conventional mixture modeling. The

Table I
AVERAGE NUMERICAL EVALUATION OF SEGMENTATION PERFORMANCE FOR VIDEO 1.

Mesurements	*HDP-GID*	*HDP-GMM*	*GMM-MDL*
$text\_var$	541.28	596.37	661.06
$color\_con$	1.51	1.49	1.45
$size\_diff$	54.23	61.35	68.49
$elong\_diff$	0.57	0.59	0.68
$\mathcal{X}^2$	0.21	0.25	0.31
$motion\_var$	289.52	346.19	413.53

Table II
AVERAGE NUMERICAL EVALUATION OF SEGMENTATION PERFORMANCE FOR VIDEO 2.

Mesurements	*HDP-GID*	*HDP-GMM*	*GMM-MDL*
$text\_var$	395.73	427.64	463.17
$color\_con$	1.39	1.34	1.33
$size\_diff$	39.12	47.22	53.58
$elong\_diff$	0.37	0.42	0.46
$\mathcal{X}^2$	0.13	0.17	0.18
$motion\_var$	302.41	351.07	389.92

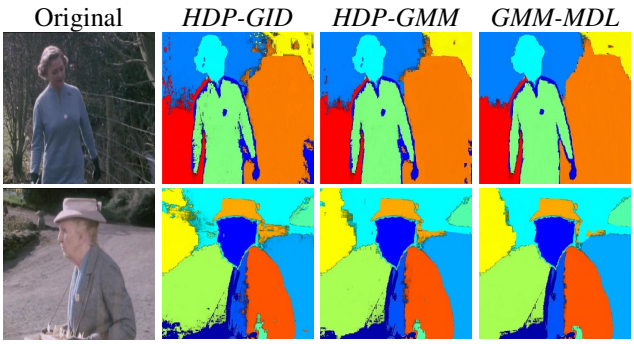

Figure 2. Examples of video segmentation results. First row: Video 1, Frame number 21; Second row: Video 2, Frame number 220.

example of video segmentation for each video sequence can be viewed in Fig. 2. According to this figure, it is clear that *HDP-GID* obtained better quality of object segmentation.

IV. CONCLUSION

A HDP mixture of GID distributions is developed to tackle the challenging task of video segmentation. The model is based on a feature selection approach and is learned within a principled variational framework. The experiments have shown promising results. Future works could be devoted to the extension of the proposed framework to online settings to improve further its generalization and flexibility.

ACKNOWLEDGMENT

This research was funded by the National Natural Science Foundation of China (61502183,61673185) and NSERC.

REFERENCES

[1] M. S. Allili, N. Bouguila, and D. Ziou, "A robust video foreground segmentation by using generalized gaussian mixture modeling," in *Fourth Canadian Conference on Computer and Robot Vision (CRV 2007), 28-30 May 2007, Montreal, Quebec, Canada*, pp. 503–509, 2007.

[2] M. S. Allili, N. Bouguila, and D. Ziou, "Finite generalized gaussian mixture modeling and applications to image and video foreground segmentation," in *Fourth Canadian Conference on Computer and Robot Vision (CRV 2007), 28-30 May 2007, Montreal, Quebec, Canada*, pp. 183–190, 2007.

[3] M. S. Allili, D. Ziou, N. Bouguila, and S. Boutemedjet, "Image and video segmentation by combining unsupervised generalized gaussian mixture modeling and feature selection," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 20, no. 10, pp. 1373–1377, 2010.

[4] X. Song and G. Fan, "Selecting salient frames for spatiotemporal video modeling and segmentation," *IEEE Transactions on Image Processing*, vol. 16, pp. 3035–3046, Dec 2007.

[5] 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.

[6] Y. W. Teh and M. I. Jordan, "Hierarchical Bayesian Nonparametric Models with Applications," in *Bayesian Nonparametrics: Principles and Practice* (N. Hjort, C. Holmes, P. Müller, and S. Walker, eds.), Cambridge University Press, 2010.

[7] T. S. Ferguson, "Bayesian Density Estimation by Mixtures of Normal Distributions," *Recent Advances in Statistics*, vol. 24, pp. 287–302, 1983.

[8] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica*, vol. 4, pp. 639–650, 1994.

[9] C. Wang, J. W. Paisley, and D. M. Blei, "Online variational inference for the hierarchical Dirichlet process," *Journal of Machine Learning Research - Proceedings Track*, vol. 15, pp. 752–760, 2011.

[10] M. A. Mashrgy, T. Bdiri, and N. Bouguila, "Robust simultaneous positive data clustering and unsupervised feature selection using generalized inverted dirichlet mixture models," *Knowledge-Based Systems*, vol. 59, pp. 182 – 195, 2014.

[11] T. Bdiri, N. Bouguila, and D. Ziou, "Variational bayesian inference for infinite generalized inverted dirichlet mixtures with feature selection and its application to clustering," *Applied Intelligence*, vol. 44, no. 3, pp. 507–525, 2016.

[12] T. Bdiri, N. Bouguila, and D. Ziou, "A statistical framework for online learning using adjustable model selection criteria," *Engineering Applications of Artificial Intelligence*, vol. 49, pp. 19–42, 2016.

[13] T. Bdiri and N. Bouguila, "Positive vectors clustering using inverted dirichlet finite mixture models," *Expert Syst. Appl.*, vol. 39, no. 2, pp. 1869–1882, 2012.

[14] M. H. C. Law, M. A. T. Figueiredo, and A. K. Jain, "Simultaneous feature selection and clustering using mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1154–1166, 2004.

[15] N. Bouguila, "A model-based approach for discrete data clustering and feature weighting using MAP and stochastic complexity," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 12, pp. 1649–1664, 2009.

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

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

[18] W. Fan, N. Bouguila, and D. Ziou, "Unsupervised hybrid feature extraction selection for high-dimensional non-gaussian data clustering with variational inference," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 7, pp. 1670–1685, 2013.

[19] D. M. Blei and M. I. Jordan, "Variational inference for Dirichlet process mixtures," *Bayesian Analysis*, vol. 1, pp. 121–144, 2005.

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

[21] N. Bouguila, "Spatial color image databases summarization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2007, Honolulu, Hawaii, USA, April 15-20, 2007*, pp. 953–956, IEEE, 2007.

[22] N. Bouguila and D. Ziou, "Improving content based image retrieval systems using finite multinomial dirichlet mixture," in *Proc. of the IEEE Workshop on Machine Learning for Signal Processing*, pp. 23–32, 2004.

[23] N. Bouguila and D. Ziou, "Dirichlet-based probability model applied to human skin detection [image skin detection]," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2004, Montreal, Quebec, Canada, May 17-21, 2004*, pp. 521–524, IEEE, 2004.

[24] T. Brox and J. Malik, "Object segmentation by long term analysis of point trajectories," in *Computer Vision: ECCV 2010* (K. Daniilidis, P. Maragos, and N. Paragios, eds.), vol. 6315 of *Lecture Notes in Computer Science*, pp. 282–295, Springer Berlin Heidelberg, 2010.

[25] P. Correia and F. Pereira, "Objective evaluation of video segmentation quality," *IEEE Transactions on Image Processing*, vol. 12, pp. 186–200, Feb 2003.

[26] C. Erdem, B. Sankur, and A. Tekalp, "Performance measures for video object segmentation and tracking," *IEEE Transactions on Image Processing*, vol. 13, pp. 937–951, July 2004.