Video Forgery Detection Using a Bayesian RJMCMC-Based Approach

Sami Bourouis
Taif University, KSA
LR-SITI, Tunis El Manar University, Tunis, Tunisia
Email: s.bourouis@tu.edu.sa

F. R. Al-Osaimi
Umm Al-Qura University, KSA
Email: frosaimi@uqu.edu.sa

Nizar Bouguila
Concordia University, Canada
Email: nizar.bouguila@concordia.ca

Hassen Sallay
Umm Al-Qura University, KSA
Email: hmsallay@uqu.edu.sa

Fahd Aldosari
Umm Al-Qura University, KSA
Email: fmdosari@uqu.edu.sa

Mohamed Al Mashrgy
Concordia University, Canada
Email: m_almash@encs.concordia.ca

***Abstract*—We propose a Bayesian approach to learn finite generalized inverted Dirichlet mixture models. The developed approach performs simultaneous parameters estimation, model complexity determination, and feature selection via a reversible jump Markov Chain Monte Carlo (RJMCMC) algorithm. A challenging application that concerns video forgery detection is deployed to validate our statistical framework and to show its merits.**

I. Introduction

Video forgery detection has been an active research topic in the last few years [1, 2, 3, 4]. It can be viewed as the problem of reliably distinguishing between tampered videos and untampered original ones. Indeed, with the advent of sophisticated multimedia editing softwares, that allow the manipulation of videos, the integrity of a given video content cannot be taken as granted. This strong interest is driven by a wide spectrum of promising applications in many areas such as forensics and security (e.g. a video can contain a hidden information), and journalism (e.g. using videos as critical evidence) to name a few. Unlike image forgery detection, the problem has received much attention just recently, and the development of new efficient methods is urgent. To solve such problem, one may think of extending techniques used for image tampering detection [5] to each frame of the treated video. Unfortunately, it will be difficult to detect some types of forgery because frames are evaluated independently and there is no consideration of the correlation between the frames. Similar to those for image forgery detection techniques, video forgery can be grouped as: inter-video based and intra-video based approaches. For intra-video forgery-objects are clipped or replaced with duplicates from the same video in order to mask other objects. While for inter-video forgery-objects from different videos are used to tamper other video. For example, frames from other video are superimposed on a specific frame in the current video. For example, Wang and Farid [1] studied the intra-video forgery in order to detect regions duplication. Their technique is based on the analysis of the correlation between original frames and duplicated ones to discover copy-paste forgery. Nevertheless, their approach cannot identify duplication and superimposition caused by adding regions or frames from other video sequences. The same authors proposed another work for inter-video forgery by checking the consistency of interlacing parameters [2]. The inconsistency-based method is used to convert an interlaced video to a non interlaced one. In their method, parameters are calculated through the expectation-maximization (EM) algorithm. They suggested that the motion between fields of a frame is closely related across fields in interlaced videos. Video noise-based correlation has also been studied to detect tampering. In fact, Hsu et al. [3] proposed a method that exploits noise residual as a feature characteristics extracted from the video and they used a block-level correlation technique. They model the distribution of correlation of temporal noise residue in a tampered video as a Gaussian mixture model (GMM). Consequently, a Bayesian classifier is used to find the optimal threshold value based on the estimated parameters. However, their approach greatly depends on the noise reduction technique and when the noise intensities of the original and tampered regions are different, it fails to reduce the noise accurately and can miss some forgeries because of the calculation error of noise residual. A different approach is presented in [6] in which the noise level function (NLF) is used to detect suspicious regions in static scene recorded from video. NLF is assumed linear and it is estimated in a least squares manner, and then pixels aare classified into authentic or tampered with respect to the distance from the estimated NLF. An extension of this work was also done in [4] to deal with nonlinear NLFs. In fact, a probabilistic model is applied with inconsistencies in noise to detect forged regions. The NLF controls the characteristics of the noise at each pixel. They calculated the probability of forgery and detect suspicious regions via MAP estimation. When a video is forged with frames or regions from a different video taken by different camera, the noise characteristics change. Their approach works for inter-video tampering and it was applied only for static video. In [7], a video forgery detection scheme was developed using Markov models of motion frame. Goodwin et al. [**?**] proposed also a technique to detect tampering by application different techniques such as

2161-5330/17 $31.00 © 2017 IEEE
DOI 10.1109/AICCSA.2017.43

the fusion process of the detected noise and the quantization residue features. Another work proposed in [8] is developed to detect the spatial and temporal copy-paste tampering in video. Their method is mainly based on the application of the Histogram of Oriented Gradients (HOG) feature matching technique and the use of video compression properties.

In this paper, we tackle video forgery detection using a Bayesian approach based on finite generalized inverted Dirichlet (GID) mixture model. The consideration of inverted generalized Dirichlet mixture is motivated by its flexibility and the excellent results previously obtained by deploying this model as shown in [9] where an MCMC algorithm has been developed using the fact that the generalized inverted Dirichlet belongs to the exponential family of distribution to develop its parameters priors. The application of MCMC algorithms to learning statistical models in general and mixture models in particular [10, 11, 12] has produced interesting results. A crucial problem when we consider finite mixture models is the selection of the number of components. A good approach that we shall develop in this paper is RJMCMC that generalizes the traditional MCMC to the case where the dimension of the unknown parameters in the model is also unknown. RJMCMC can be viewed as a Bayesian model averaging procedure that selects the model automatically by producing the posterior probability of the number of components, upon which one draws a conclusion on how many components are needed to model the data [13]. To the best of our knowledge RJMCMC inference of GID mixtures has not been done before. A challenging problem when using RJMCMC in real-life applications is the high dimensionality of the data. To circumvent this difficulty, we integrate a feature selection approach within our model.

The rest of this paper is organized as follows. As background, we review the generalized inverted Dirichlet mixture model and how to integrate feature selection within it in Section 2. Section 3 is dedicated to the development of our RJMCMC algorithm. Section 4 presents our experimental results. The conclusion is given in Section 5.

II. The Bayesian Model

Let $\mathcal{Y} = \{\vec{Y}_1, \ldots, \vec{Y}_2\}$ be a data set containing N objects where $\vec{Y}_i$ is a D-dimensional positive vector describing the i-th object that needs to be assigned into one of the M groups in the data set. A good generative model to represent such a data set is the GID mixture [9]:

$$p(\vec{Y}_i|\Theta) = \sum_{j=1}^{M} p_j p(\vec{Y}_i|\vec{\theta}_j) \quad (1)$$

where $\Theta = (\vec{\theta}_1, \vec{\theta}_2, \ldots, \vec{\theta}_M, p_1, p_2, \ldots, p_M)$, $(p_1, p_2, \ldots, p_M)$ is the vector of mixing weights which must be positive and sum to one, $p(\vec{Y}_i|\vec{\theta}_j)$ is the GID distribution with parameters $\vec{\theta}_j = (\alpha_{j1}, \beta_{j1}, \alpha_{j2}, \beta_{j2}, \ldots, \alpha_{jD}, \beta_{jD})$:

$$p(\vec{Y}_i|\vec{\theta}_j) = \prod_{d=1}^{D} \frac{\Gamma(\alpha_{jd} + \beta_{jd})}{\Gamma(\alpha_{jd})\Gamma(\beta_{jd})} \frac{Y_{id}^{\alpha_{jd}-1}}{(1 + \sum_{l=1}^{d} Y_{il})^{\gamma_{jd}}} \quad (2)$$

where $\vec{\theta}_j = (\alpha_{j1}, \beta_{j1}, \ldots, \alpha_{jD}, \beta_{jD})$, $\gamma_{jd} = \beta_{jd} + \alpha_{jd} - \beta_{jd+1}$ for $d = 1, \ldots, D$ with $\beta_{jD+1} = 0$.

The GID has an interesting property, previously shown in [14]. Indeed, if a vector $\vec{Y}_i$ has a generalized inverted Dirichlet distribution with parameters $(\alpha_1, \beta_1, \ldots, \alpha_D, \beta_D)$, then we can construct a vector $\vec{X}$ using the following geometric transformation $X_{i1} = Y_{i1}$ and $X_{il} = \frac{Y_{il}}{1+\sum_{k=1}^{l-1} Y_{ik}}$ for $l > 1$, such that each X_{id} has an inverted Beta distribution with parameters (α_d, β_d):

$$p_{IBeta}(X_{id}|\alpha_d, \beta_d) = \frac{\Gamma(\alpha_d + \beta_d)}{\Gamma(\alpha_d)\Gamma(\beta_d)} \frac{X_{id}^{\alpha_d - 1}}{(1 + X_{id})^{(\alpha_d + \beta_d)}} \quad (3)$$

This property means, as shown in [14], that the generalized inverted Dirichlet can be transformed to a multidimensional inverted Beta mixture model with conditionally independent features:

$$p(\vec{X}_i|\Theta) = \sum_{j=1}^{M} p_j \prod_{d=1}^{D} p_{IBeta}(X_{id}|\alpha_{jd}, \beta_{jd}) \quad (4)$$

The mean and the variance of the inverted Beta distribution are as following

$$\mu_{jd} = \frac{\alpha_{jd}}{\beta_{jd} - 1} \quad (5)$$

$$\sigma_{jd}^2 = \frac{\alpha_{jd}(\alpha_{jd} + \beta_{jd} - 1)}{(\beta_{jd} - 2)(\beta_{jd} - 1)^2} \quad (6)$$

An interesting unsupervised feature selection formulation has been previously applied successfully for the GID mixture in [14] and we shall adopt it here within Bayesian settings in our RJMCMC learning framework:

$$\begin{aligned} p(\vec{X}_i|\Theta^*) &= \sum_{j=1}^{M} p_j \prod_{d=1}^{D} \Big[\rho_d p_{IBeta}(X_{id}|\theta_{jd}) \\ &+ (1 - \rho_d) p_{IBeta}(X_{id}|\lambda_d) \Big] \end{aligned} \quad (7)$$

where $\Theta^* = \{\{p_j\}, \{\theta_{jd}\}, \{\rho_l\}, \{\lambda_l\}\}$ is the set of all our unsupervised feature selection model parameters that should be estimated, $\theta_{jd} = (\mu_{jd}, \sigma_{jd}^2)$, ρ_l represents the probability that feature X_{il} is relevant for clustering, and $p_{IBeta}(X_{il}|\lambda_l)$ is an inverted Beta distribution with parameter $\lambda_l = (\mu_{\lambda|l}, \sigma_{\lambda|l}^2)$, common to all clusters to generate irrelevant features.

III. RJMCMC-based learning

Unlike classic mixture models formulations, the number of components M is considered here as a parameter in the model for which a conditional distribution should be found. Thus, our unknowns are M, $\vec{\theta} = (\vec{\theta}_1, \ldots, \vec{\theta}_M)$, such that $\vec{\theta}_j = (\theta_{j1}, \ldots, \theta_{jD})$, and $\vec{P} = (p_1, \ldots, p_M)$ and are regarded as random variables drawn from some prior distributions that we have to specify. By imposing common conditional independencies [13], we get the following joint distribution

$$p(M, \vec{P}, Z, \vec{\theta}, \mathcal{Y}) = p(M)p(\vec{P}|M)P(Z|\vec{P}, M)p(\vec{\theta}|M)p(\mathcal{Y}|\vec{\theta}, Z) \quad (8)$$

where $Z = \{\vec{Z}_1, \ldots, \vec{Z}_N\}$, and each $\vec{Z}_i$ is a M-dimensional membership vector (known also as the unobserved or missing vector) that indicates to which component $\vec{Y_i}$ belongs, such that: Z_{ij} will be equal 1 if $\vec{Y_i}$ belongs to class j or 0, otherwise. It is noteworthy that the Z_i are supposed to be drawn independently from the following distribution

$$p(Z_i = j) = p_j \quad j = 1, \ldots, M. \tag{9}$$

For our model, we have chosen an inverted Beta and inverse Gamma distributions as priors for the mean μ_{jd} and the variance σ^2_{jd}, respectively. In addition, the typical prior choice for the mixing weight $\vec{P}$ is the Dirichlet distribution since it is defined under the constraint of $p_1, \ldots, p_M : \sum_{j=1}^{M} p_j < 1$. The RJMCMC methodology developed in [13] enables the simultaneous estimation of the posterior probabilities of several models under consideration and the parameters conditional on a specific model. It generalizes the classic Metropolis-Hastings (M-H) algorithm into the dimension varying situation. It is designed such that the sampler moves across different dimensions according to six types of moves:

1) Update the mixing parameters $\vec{P}$
2) Update the parameters $\vec{\mu_j}$ and $\vec{\sigma}^2_j$
3) Update the allocation Z
4) Update the hyperparameters.
5) Split one component into two, or combine two into one
6) The birth or death of an empty component

It is noteworthy that (1), (2), (3) and (4) are usual parameter update moves via the Gibbs sampling. On the other hand, (5) and (6) are trans-dimensional moves (involve changing the number of components by one), they constitute the reversible jump and are used for model selection via the M-H algorithm which has been extensively studied and adopted in the context of many applications from different disciplines [15, 16]. Each step $t = 1, \ldots, 6$ is called a move and a sweep is defined as a complete pass over the six moves. Assume that we are in state Δ_M, where $\Delta_M = (Z, P, M)$. The MCMC step representing move (5) takes the form of a Metropolis-Hastings step by proposing a move from a state Δ_M to $\hat{\Delta}_M$ with target probability distribution (posterior distribution) $p(\Delta_M|\mathcal{Y})$ and proposal distribution $q_t(\Delta_M, \hat{\Delta}_M)$ for the move t. When we are in the current state Δ_M, a given move t to destination $\hat{\Delta}_M$ is accepted with probability

$$p_t(\Delta_M, \hat{\Delta}_M) = \left(1, \frac{p(\hat{\Delta}_M|\mathcal{Y})q_t(\hat{\Delta}_M, \Delta_M)}{p(\Delta_M|\mathcal{Y})q_t(\Delta_M, \hat{\Delta}_M)}\right) \tag{10}$$

In the case of a move type where the dimension of the parameter does not change we use an ordinary ratio of densities. A move from a point Δ_M to $\hat{\Delta}_M$ in a higher dimensional space is done by drawing a vector of continuous random variables u, independent of Δ_M and the new state $\hat{\Delta}_M$ is determined by using an invertible deterministic function of Δ_M and u: f(Δ_M, u) [13]. On the other hand, the move from $\hat{\Delta}_M$ to Δ_M can be carried out using the inverse transformation. Hence, the move acceptance probability is given by

$$p_t(\Delta_M, \hat{\Delta}_M) = \min\left(1, \frac{p(\hat{\Delta}_M|\mathcal{Y})r_m(\hat{\Delta}_M)}{p(\Delta_M|\mathcal{Y})r_m(\Delta_M)q(u)}\left|\frac{\partial(\hat{\Delta}_M)}{\partial(\Delta_M, u)}\right.\right) \tag{11}$$

where $r_m(\Delta_M)$ is the probability of choosing move type m when we are in state Δ_M, and $q(u)$ is the density function of u. The last term $\frac{\partial(\hat{\Delta}_M)}{\partial(\Delta_M, u)}$ is the Jacobian function arising from the variable change from (Δ_M, u) to state $\hat{\Delta}_M$. All RJMCMC moves are discussed in the following.

The first four steps of RJMCMC are based on simple Gibbs sampling where the parameters are drawn from their known full conditional distributions. In move (5), we make a random choice between attempting to split or combine, with probabilities a_M and b_M where $b_M = 1 - a_M$, respectively. It is clear that, $a_{M_{max}} = 0$ and $b_1 = 0$, otherwise we choose $a_M = b_M = 0.5$ for $M = 1, \ldots, M_{max}$, where M_{max} is the maximum value allowed for M. The combine move is constructed by randomly choosing a pair of components (j_1, j_2), which must be adjacent; in other words they must meet the following constraint: $\vec{\mu}_{j_1} < \vec{\mu}_{j_2}$, where there is no other $\vec{\mu}_j$ in the interval $[\vec{\mu}_{j_1}, \vec{\mu}_{j_2}]$. Then, these two components can be merged and M is reduced by 1. We denote the new formed component by j^* which contains all the observations that were allocated to j_1 and j_2. Finally, we generate the parameter values for the new components $p_{j^*}, \vec{\mu}_{j^*}, \vec{\sigma}^2_{j^*}$ by preserving the zeroth, first, and second moments. For the split type move, a component j^* is chosen randomly and we split it into two components j_1 and j_2 with new parameters $p_{j_1}, \vec{\mu}_{j_1}$, and $\vec{\sigma}^2_{j_1}$ and $p_{j_2}, \vec{\mu}_{j_2}$, and $\vec{\sigma}^2_{j_2}$, respectively. Since there are 3 degrees of freedom in achieving this, we need to generate, from a Beta distribution, a three-dimensional random vector $u = [u_1, u_2, u_3]$ to define the new parameters [13].

For the new generated components, the adjacency condition defined in the combine move must be checked to make sure that the split/combine is reversible or not. If this condition is rejected, split/combine move is not reversible, the split move is rejected. Otherwise, the split move is accepted and we reallocate the j^* into the new components j_1 and j_2 using Eq. 9. According to Eq. 11, the acceptance probability R for the split and combine moves types can be calculated using the following

$$R = \frac{P(Z, P, M+1, \varepsilon, \zeta, \varpi, \vartheta|\mathcal{Y})b_{M+1}}{p(Z, P, M, \varepsilon, \zeta, \varpi, \vartheta|\mathcal{Y})a_M P_{alloc} q(u)}\left|\frac{\partial\hat{\Delta}_M}{\partial(\Delta_M, u)}\right| \tag{12}$$

where the acceptance probability for the split is $min(1, R)$, and for the combine move is $min(1, R^{-1})$. P_{alloc} is the probability of making this particular allocation to components j_1 and j_2, $\left|\frac{\partial\hat{\Delta}_M}{\partial(\Delta_M, u)}\right|$ is the Jacobian of the transformation from the state $(p_{j^*}, \vec{\mu}_{j^*}, \vec{\sigma}^2_{j^*}, u_1, u_2, u_3)$ to state $(p_{j_1}, \vec{\mu}_{j_1}, \vec{\sigma}^2_{j_1}, p_{j_2}, \vec{\mu}_{j_2}, \vec{\sigma}^2_{j_2})$. In Death/Birth move, we first make a random choice between birth and death with the same a_m and b_M as above. If the birth move is chosen, the values of the parameters of the new components $(\vec{\mu}_{j^*}, \vec{\sigma}^2_{j^*})$ are drawn from the associated prior distributions. The weight of the new component, p_{j*}, is generated from the marginal distribution of p_{j*} derived from the distribution of $\vec{P} = (p_1, \ldots, p_M, p_{j*})$. The vector $\vec{P}$ follows a Dirichlet with parameters $(\delta_1, \ldots, \delta_M, \delta_{j*})$, thus the marginal of p_{j*} is a

Beta distribution with parameters $(\delta_{j*}, \sum_{j=1}^{M} \delta_j)$. Note that in order to keep the mixture constraint $\sum_{j=1}^{M} p_j + p_{j*} = 1$, the previous weights $p_j, j = 1, \ldots, M$ have to be rescaled and then all multiplied by $(1 - p_{j*})$.

IV. Experimental Results

Among video forgery detection the successful approaches, the authors in [3] proposed to exploit noise residual as a feature characteristics extracted from the video and to use block-level correlation technique. The distribution of correlation of temporal noise residue in a tampered video, in forged and normal regions, was supposed to be a Gaussian mixture model which parameters are learned using the expectation maximization algorithm. Consequently, a Bayesian classifier is used to find the optimal threshold value based on the estimated parameters. The goal of this application is to investigate the detection results when a generalized inverted Dirichlet mixture is considered instead to motivate further this particular choice. In our experiments, we consider five video sequences (see Figure 1) forged using the inpainting approach proposed in [17]. We have evaluated the forgery detection performance using precision and recall rates that we have calculated from available ground truth:

$$Precision = \frac{N_{hit}}{N_{hit} + N_{false}} \tag{13}$$

$$Recall = \frac{N_{hit}}{N_{hit} + N_{miss}} \tag{14}$$

where N_{hit} represents the number of correct detections, N_{false} denotes the number of false positives, and N_{miss} denotes the number of misses. Tables I and II display the classification results using the BGIDM and BGM models, respectively. The results show that our mixture model provides better detection results as compared to the Gaussian model which can be explained by its flexibility and ability to represent different forms and shapes.

TABLE I
Performance evaluation of the forgery detection for five test sequences using the Bayesian generalized inverted Dirichlet mixture (BGIDM).

Method	Recall	Precision
Sequence 1	65.23	93.45
Sequence 2	66.93	94.20
Sequence 3	64.87	92.96
Sequence 4	66.68	94.22
Sequence 5	65.77	93.08

TABLE II
Performance evaluation of the forgery detection for five test sequences using the Bayesian Gaussian mixture (BGM).

Method	Precision	Recall
Sequence 1	61.88	90.01
Sequence 2	62.07	90.15
Sequence 3	60.98	89.68
Sequence 4	62.37	90.82
Sequence 5	61.90	91.03

V. Conclusion

We have proposed a novel statistical framework based on Bayesian learning of finite generalized inverted Dirichlet mixture model. The proposed framework allows simultaneous learning and feature selection. Experimental results has concerned a challenging application namely video forgery detection and have shown the merits of the proposed approach. Future works could be devoted to the extension of the proposed model to the infinite case by using a similar approach as the one proposed in [18, 19]. Potential future works include the validation of the proposed Bayesian model via other applications such as concept modeling [20], image databases summarization [21], and video segmentation [22].

Acknowledgment

The authors would like to thank Umm al-Qura University, Kingdom of Saudi Arabia, for their funding support under grant number 15-COM-3-1-0007.

References

[1] W. Wang and H. Farid, "Exposing digital forgeries in video by detecting duplication," in *Proc. of the 9th Workshop on Multimedia & Security*. New York, NY, USA: ACM, 2007, pp. 35–42.

[2] ——, "Exposing digital forgeries in interlaced and deinterlaced video," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 438–449, 2007.

[3] C.-C. Hsu, T.-Y. Hung, C.-W. Lin, and C.-T. Hsu, "Video forgery detection using correlation of noise residue," in *2008 IEEE 10th Workshop on Multimedia Signal Processing*, 2008, pp. 170–174.

[4] M. Kobayashi, T. Okabe, and Y. Sato, "Detecting forgery from static-scene video based on inconsistency in noise level functions," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 883–892, 2010.

[5] T. Chihaoui, S. Bourouis, and K. Hamrouni, "Copy-move image forgery detection based on sift descriptors and svd-matching," in *2014 1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, March 2014, pp. 125–129.

[6] M. Kobayashi, T. Okabe, and Y. Sato, "Detecting video forgeries based on noise characteristics," in *Advances in Image and Video Technology, Third Pacific Rim Symposium, PSIVT 2009, Tokyo, Japan, January 13-16, 2009. Proceedings*, 2009, pp. 306–317.

[7] K. Kancherla and S. Mukkamala, "Novel blind video forgery detection using markov models on motion residue," in *Intelligent Information and Database Systems - 4th Asian Conference, ACIIDS 2012, Kaohsiung, Taiwan, March 19-21, 2012, Proceedings, Part III*, 2012, pp. 308–315.

[8] A. V. Subramanyam and S. Emmanuel, "Video forgery detection using HOG features and compression properties," in *14th IEEE International Workshop on Multimedia Signal Processing, MMSP 2012, Banff, AB, Canada, September 17-19, 2012*, 2012, pp. 89–94.

Fig. 1. Original and inpainted frames used to evaluate the detection approach.

[9] S. Bourouis, M. A. Mashrgy, and N. Bouguila, "Bayesian learning of finite generalized inverted dirichlet mixtures: Application to object classification and forgery detection," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2329–2336, 2014.

[10] C. R. B. Cabral, H. Bolfarine, and J. R. G. Pereira, "Bayesian density estimation using skew student-t-normal mixtures," *Computational Statistics & Data Analysis*, vol. 52, no. 12, pp. 5075–5090, 2008.

[11] N. Bouguila, D. Ziou, and R. I. Hammoud, "On bayesian analysis of a finite generalized dirichlet mixture via a metropolis-within-gibbs sampling," *Pattern Analysis and Applications*, vol. 12, no. 2, pp. 151–166, 2009.

[12] N. Bouguila, D. Ziou, and E. Monga, "Practical bayesian estimation of a finite beta mixture through gibbs sampling and its applications," *Statistics and Computing*, vol. 16, no. 2, pp. 215–225, 2006. [Online]. Available: http://dx.doi.org/10.1007/s11222-006-8451-7

[13] S. Richardson and P. J. Green, "On bayesian analysis of mixtures with an unknown number of components (with discussion)," *Journal of the Royal Statistical Society: Series B*, vol. 59, no. 4, pp. 731–792, 1997.

[14] 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.

[15] J. S. Liu, F. Liang, and W. H. Wong, "The multiple-try method and local optimization in Metropolis sampling," *Journal of the American Statistical Association*, vol. 95, no. 449, pp. 121–134, 2000.

[16] Z. Kato, "Segmentation of color images via reversible jump mcmc sampling," *Image and Vision Computing*, vol. 26, no. 3, pp. 361–371, 2008.

[17] A. Mosleh, N. Bouguila, and A. B. Hamza, "Video completion using bandlet transform," *IEEE Transactions on Multimedia*, vol. 14, no. 6, pp. 1591–1601, 2012.

[18] N. Bouguila and D. Ziou, "A dirichlet process mixture of dirichlet distributions for classification and prediction," in *2008 IEEE Workshop on Machine Learning for Signal Processing*, 2008, pp. 297–302.

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

[20] ——, "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*. IEEE, 2004, pp. 521–524.

[21] N. Bouguila, "Spatial color image databases summarization," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, 2007, pp. 953–956.

[22] M. S. Allili, N. Bouguila, and D. Ziou, "Finite generalized gaussian mixture modeling and applications to image and video foreground segmentation," in *Proc. of the Fourth Canadian Conference on Computer and Robot Vision (CRV)*, 2007, pp. 183–190.