



Generalized Inverted Dirichlet Optimal Predictor for Image Inpainting

Omar Graja^(✉), Fatma Najjar^(✉), and Nizar Bouguila^(✉)

Concordia Institute for Information and Systems Engineering (CIISE),
Concordia University, Montreal, QC, Canada

omar.graja@mail.concordia.ca, f_najjar@encs.concordia.ca,
nizar.bouguila@concordia.ca

Abstract. Predicting a given pixel from surrounding neighbouring pixels is of great interest for several image processing tasks. Previous works focused on developing different Gaussian based models. Simultaneously, in real-world applications, the image texture and clutter are usually known to be non-Gaussian. In this paper, we develop a pixel prediction framework based on a finite generalized inverted Dirichlet (GID) mixture model that has proven its efficiency in several machine learning applications. We propose a GID optimal predictor, and we learn its parameters using a likelihood-based approach combined with the Newton-Raphson method. We demonstrate the efficiency of our proposed approach through a challenging application, namely image inpainting, and we compare the experimental results with related-work methods.

Keywords: Generalized inverted Dirichlet · Mixture models · Optimal predictor · Image pixel prediction · Image inpainting

1 Introduction

Pixel prediction has shown to be one of the most needed tool to perform several applications in image processing such as anomaly detection [18, 31], object detection [8, 17], edge detection [28], video compression [7, 24], semantic segmentation [27, 36], image restoration [11, 32] and keypoint prediction [37]. Meanwhile, pixel prediction is represented by approximating the predicted pixel using its neighbors. For that, it is usually represented by a linear or a non-linear combination of the neighboring pixels plus an error value [29, 38]. In this paper, inspired by the work proposed in [20], we take the optimal predictor of x_n as the conditional expectation $E(x_{ij} \mid \forall x_{i'j'} \in N_{i,j})$, where $x_{i'j'}$ are the neighbors of x_{ij} within the set of pixels $N_{i,j}$.

Exploiting the ease of analytical derivations, the authors in [40] derived optimal predictors for Gaussian distribution and mixture of Gaussians. However, the field of non-Gaussian distributions exhibited an exciting expansion in terms of mathematical theorems [23, 35]. Yet, a large number of researchers proved that Gaussian assumption is generally inappropriate and other alternative distributions are more effective in modeling data than Gaussian distribution by unveiling

more appropriate patterns and correlations among data features [4–6, 10]. In our recent work [19], we proved that finite inverted Dirichlet mixture model effectively represents positive vectors [3, 9, 39]. However, it suffers from significant drawbacks such as its minimal, strictly positive covariance structure. Therefore, to overcome this limitation, we consider applying the generalized inverted Dirichlet which belongs to Liouville family of distributions [22]. This distribution provides a more decent representation of the variability of the data [12]. Indeed, considering the fact that generalized inverted Dirichlet could be factorized into a set of inverted Beta distributions [13], gives more flexibility for modeling data in real-world applications.

In this work, we derive a novel optimal predictor based on generalized inverted Dirichlet distribution which results in a linear combination of the neighboring pixels. Meanwhile, we evaluate the proposed approach on image inpainting application. We choose a publicly available dataset namely Paris StreetView to validate our approach [16]. For the purpose of proving the efficiency of our proposed optimal predictor, we consider two types of pixel discarding. The first pixel removal is random, whereas, in the second experiment, we discard lines from the image. We use a 3rd order non-symmetrical half-plane casual (NSHP) neighborhood system to compute the missing pixel [11]. Finally, we perform two image comparison metrics to evaluate our proposed model and compare it to other similar optimal based predictors.

The rest of the paper is organized as follows: in Sect. 2, we describe our prediction model, and we derive the analytical expression of the GID optimal predictor. In Sect. 3, we consider the image inpainting application on Paris StreetView dataset with two different data masks to demonstrate the effectiveness of the proposed predictor, and we discuss the experimental results. Finally, a summary is provided in Sect. 4.

2 GID Prediction Model

The generalized inverted Dirichlet mixture model has shown high flexibility for modeling and clustering positive vectors. In this section, we start by reviewing the finite GID mixture model. Then, we introduce the parameters learning through the EM algorithm and, later, we extend this model to the prediction.

2.1 Mixture of Generalized Inverted Dirichlet Distributions

Let $\mathcal{X} = (\mathbf{X}_1, \dots, \mathbf{X}_N)$ be a set of N d -dimensional positive vectors where each vector \mathbf{X}_i follows a mixture of K generalized inverted Dirichlet (GID) distributions characterized by parameters $\boldsymbol{\theta}_j = (\boldsymbol{\alpha}_j, \boldsymbol{\beta}_j)$ and mixing weight π_j of the j th component.

$$P(\mathbf{X}_i|\Theta) = \sum_{j=1}^K \pi_j P(\mathbf{X}_i|\boldsymbol{\theta}_j) \quad (1)$$

where $\Theta = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K, \pi_1, \dots, \pi_K)$ represents the GID mixture model parameters and $P(\mathbf{X}_i|\boldsymbol{\theta}_j)$ is the generalized inverted Dirichlet distribution, which has the following form [26]:

$$P(\mathbf{X}_i|\boldsymbol{\theta}_j) = \prod_{l=1}^d \frac{\Gamma(\alpha_{jl} + \beta_{jl})}{\Gamma(\alpha_{jl})\Gamma(\beta_{jl})} \frac{X_j^{\alpha_{jl}-1}}{(1 + \sum_{s=1}^l X_{is})^{\gamma_{jl}}} \tag{2}$$

where $\gamma_{jl} = \beta_{jl} + \alpha_{jl} - \beta_{j,l+1}$, for $l = 1, \dots, d$ ($\beta_{j,d+1} = 1$). It is to be noted that the GID is reduced to the inverted Dirichlet distribution when the parameter γ_{jl} is set to zero ($\gamma_{j1} = \dots = \gamma_{jd} = 0$).

The flexibility of the generalized inverted Dirichlet is by dint of the concept of ‘‘Force of mortality’’ of the population where we introduce, here, a doubly non-central Y independent-variables defined as

$$Y_{i1} = 1, \quad Y_{jl} = \frac{X_{il}}{T_{il-1}}, \quad l > 1 \tag{3}$$

where $T_{il} = 1 + X_{i1} + X_{i2} + \dots + X_{i,l-1}$, $l = 1, \dots, d$

The characteristic function underlying the $\mathcal{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_N)$ independent variables follows a product of 2-parameters inverted Beta distribution, where $\theta_l = (\alpha_l, \beta_l)$

$$P(\mathbf{Y}_i|\boldsymbol{\theta}) = \prod_{l=1}^d P_{IBeta}(Y_{il}|\theta_l) \tag{4}$$

In which the probability of inverted Beta is given by:

$$P_{IBeta}(Y_{il}|\theta_l) = \frac{\Gamma(\alpha_l + \beta_l)}{\Gamma(\alpha_l)\Gamma(\beta_l)} \frac{Y_{il}^{\alpha_l}}{(1 + Y_{il})^{\alpha_l + \beta_l}} \tag{5}$$

Many characteristics of the distribution are defined in [30]. We mention some interesting statistics for this distribution.

First, the mixed moments such as the n^{th} moment is given by:

$$E(Y^n) = \frac{\Gamma(\alpha + \beta)\Gamma(\beta - n)}{\Gamma(\alpha)\Gamma(\beta)} \tag{6}$$

where $\beta - n$ is positive.

Then, the covariance between two variable Y_1 and Y_2 is defined as:

$$COV(Y_1, Y_2) = \frac{(\alpha_1 - 1)(\alpha_2 - 1)}{(\alpha_1\alpha_2 - 1)(\beta_1 - 1)(\beta_2 - 1)} \tag{7}$$

The covariance between two features for inverted Beta is always positive, which means that both they tend to increase or decrease together.

Finally, the variance of a variable Y is conveyed by:

$$VAR(Y) = \frac{(\alpha - 1)(\alpha + \beta - 2)}{(\beta - 1)^2(\beta - 2)} \tag{8}$$

2.2 Likelihood-Based Learning

Theoretically, a plethora of literature agrees on the effectiveness of the likelihood-based approach for estimating the mixture parameters. One of the well-known methodologies is the Expectation-Maximization technique [15], beginning with a tuned initialization for the set of parameters to the expectation step where the posterior is inferred (named often as “responsibilities”), then the iterations are proceeded to update the required variables until convergence. The heart of the matter comes with estimating the parameters based on the second derivative of the log-likelihood function with regards to each parameter. First, we introduce the log-likelihood as follows:

$$\log P(\mathcal{Y}|\Theta) = \sum_{i=1}^N \log \left[\sum_{j=1}^K \pi_j \prod_{l=1}^d P_{IBeta}(Y_{il}|\theta_{jl}) \right] \quad (9)$$

Initializing Process. As a first step, an unsupervised-method, namely “K-means,” is applied to obtain the initial K clusters. Consequently, for each cluster, the method-of-moments is implemented to get the initial α_j and β_j parameters of each component j . The mixing weight is set in the initial phase as the number of elements in each cluster divided by the total number of vectors. As mentioned earlier, with conditionally independent features, the GID is converted by the inverted Beta distribution factorization. Thus, given the moments of inverted Beta distribution [2], the initial α_{jl_0} and β_{jl_0} are deduced by

$$\alpha_{jl_0} = \frac{\mu_{jl}^2(1 + \mu_{jl}) + \mu_{jl}\sigma_{jl}^2}{\sigma_{jl}^2} \quad (10)$$

$$\beta_{jl_0} = \frac{\mu_{jl}(1 + \mu_{jl}) + 2\sigma_{jl}^2}{\sigma_{jl}^2} \quad (11)$$

where μ_{jl} is the mean and σ_{jl} is the standard-deviation, $j = 1, \dots, K$, $l = 1, \dots, D$.

Expecting the Responsibilities. The responsibilities or posterior probabilities play an essential role in the likelihood-based estimation technique. It affects the update of the parameters in the next following step using the current parameter value.

$$P(j|\mathbf{Y}_i) = \frac{\pi_j P(\mathbf{Y}_i|\theta_j)}{\sum_{m=1}^K \pi_m P(\mathbf{Y}_i|\theta_m)} \quad (12)$$

Maximizing and Upgrading the GID Parameters. At the beginning, we set the gradient of log-likelihood function with respect to the mixing weight parameter equals to zero:

$$\frac{\partial \log P(\mathcal{Y}, \Theta)}{\partial \pi_j} = 0 \quad (13)$$

Then, we obtain the update formula for π_j , for $j = 1, \dots, K$ as

$$\pi_j = \frac{1}{N} \sum_{i=1}^N P(j|\mathbf{Y}_i) \tag{14}$$

where $P(j|\mathbf{Y}_i)$ is the posterior computed in the E-step.

To learn the parameters α_j and β_j , the Fisher scoring algorithm [25] is used. Thus, we need to calculate the first and the second derivatives of the log-likelihood function based on the following update [1]:

$$\alpha_{jl}^{t+1} = \alpha_{jl}^t - \left(\frac{\partial^2}{\partial \alpha_{jl}^2} \log P(\mathcal{Y}, \Theta) \right)^{-1} \times \frac{\partial}{\partial \alpha_{jl}} \log P(\mathcal{Y}, \Theta) \tag{15}$$

$$\beta_{jl}^{t+1} = \beta_{jl}^t - \left(\frac{\partial^2}{\partial \beta_{jl}^2} \log P(\mathcal{Y}, \Theta) \right)^{-1} \times \frac{\partial}{\partial \beta_{jl}} \log P(\mathcal{Y}, \Theta) \tag{16}$$

The first derivatives of $\log P(\mathcal{Y}, \Theta)$ are given by:

$$\begin{aligned} \frac{\partial}{\partial \alpha_{jl}} \log P(\mathcal{Y}, \Theta) &= \sum_{i=1}^N P(j|\mathbf{Y}_i) \left(P_{IBeta}(Y_{il}|\theta_{jl}) [\Psi(\alpha_{jl} + \beta_{jl}) - \Psi(\alpha_{jl}) + \log Y_{il} \right. \\ &\quad \left. - \log(1 + Y_{il})] \right), \end{aligned} \tag{17}$$

$$\frac{\partial}{\partial \beta_{jl}} \log P(\mathcal{Y}, \Theta) = \sum_{i=1}^N P(j|\mathbf{Y}_i) \left(P_{IBeta}(Y_{il}|\theta_{jl}) [\Psi(\alpha_{jl} + \beta_{jl}) - \Psi(\alpha_{jl}) - \log(1 + Y_{il})] \right) \tag{18}$$

The second derivative with respect to α_{jl} is given by:

$$\begin{aligned} \frac{\partial^2}{\partial^2 \alpha_{jl}} \log P(\mathcal{Y}, \Theta) &= \sum_{i=1}^N P(j|\mathbf{Y}_i) \left(\frac{\partial P_{IBeta}(Y_{il}|\theta_{jl})}{\partial \alpha_{jl}} [\Psi(\alpha_{jl} + \beta_{jl}) - \Psi(\alpha_{jl}) \right. \\ &\quad \left. - \log(1 + Y_{il})] + P_{IBeta}(Y_{il}|\theta_{jl}) [\Psi'(\alpha_{jl} + \beta_{jl}) - \Psi'(\alpha_{jl})] \right), \end{aligned} \tag{19}$$

The second derivative w.r.t β_{jl} is obtained through the same development.

2.3 GID Optimal Predictor

In this section, we present our novel non-linear optimal predictor method based on generalized inverted Dirichlet distribution. We consider the conditional expectation property to predict one random variable from the other neighboring variables.

We consider p data points $(X_i, X_{i+1}, \dots, X_{i+p-1})$, knowing their values, we predict the neighboring data point \hat{X}_{i+p} on the base of minimizing the *mean*

squared error (MSE). Therefore, we model the joint density of X_{i+p} and its neighbors using the generalized inverted Dirichlet. We take $i = 0$ and we derive the equations.

$$\mathbf{X} \sim GID(\boldsymbol{\theta}) \tag{20}$$

Considering generalized inverted Dirichlet properties [22], the conditional random variable Y_p follows an inverted Beta distribution:

$$Y_p = \frac{X_p}{T_{p-1}} | X_{p-1}, \dots, X_1, X_0 \sim IB(\alpha_p, \beta_p) \tag{21}$$

Consequently, the conditional probability density function of X_p is

$$X_p | X_{p-1}, \dots, X_0 \sim T_{p-1} IB(\alpha_p, \beta_p) \tag{22}$$

where $T_{p-1} = 1 + \sum_{k=1}^{p-1} X_k$.

Hence, the conditional expectation expression of X_p is expressed as follows.

$$E(X_p | X_{p-1}, X_1, \dots, X_0) = T_{p-1} \frac{\alpha_p}{\beta_p - 1} \tag{23}$$

In the case of mixture models, the optimal predictor expression can be derived directly by following steps defined in [19] (more details are in [40]):

$$\hat{X}_p = E(X_p | X_{p-1}, \dots, X_0) \tag{24}$$

$$\begin{aligned} \hat{X}_p &= \int X_p P(X_p | X_{p-1}, \dots, X_0) dX_p \\ &= \sum_{j=1}^K \pi'_j E_j(X_j | X_{j-1}, \dots, X_0) \end{aligned} \tag{25}$$

where

$$\pi'_j = \pi_j \frac{\int P_j(X_p, \dots, X_0) dX_p}{\int P(X_p, \dots, X_0) dX_p} \tag{26}$$

$$\pi'_j = \pi_j \frac{P_j(X_{p-1}, \dots, X_0)}{\sum_{j=1}^K \pi_j P_j(X_{p-1}, \dots, X_0)} \tag{27}$$

Finally, the GID optimal predictor is resumed in the following linear combination of X_p neighbors:

$$\hat{X}_p = \sum_{j=1}^K \pi'_j \left(1 + \sum_{k=1}^{p-1} X_k \right) \frac{\alpha_p}{\beta_p - 1} \tag{28}$$

3 Experimental Results

Image inpainting is the process of restoring deteriorating, damaged or missing parts of an image to produce a complete picture. It is an active area of image processing research [32–34] where machine learning has exciting results comparable to artists’ results. Mainly, in this process, we will be completing a missing pixel by an approximated value that depends on its neighborhood. In our work, we use the 3rd order non-symmetrical half-plane casual neighborhood system [19,21]. We apply the model on a publicly available dataset; Paris StreetView [16]. Then, we compare it with the widely used mixture of Gaussian predictor, generalized Dirichlet mixture predictor and inverted Dirichlet mixture predictor. We are not trying to restore the ground-truth image, our goal is to get an output image that is close enough or similar to the ground-truth. Therefore, we use the structural similarity index (SSIM) [11] to gauge the differences between the predicted images and the original ones. We also perform signal to noise ratio (PSNR) [14] to evaluate the performance of the models.

We reduce the size of the original images to 256×256 to minimize the complexity of computing the model’s parameters. We train the model on 70% of the database and we test on the rest. We apply two types of masks. The first one is randomly distributed as shown in Fig. 1a. And, for the second one, we discard lines of the images, as in Fig. 1b. Finally, we compute the SSIM and PSNR of each test image with its corresponding ground-truth, and we average all over the test set.

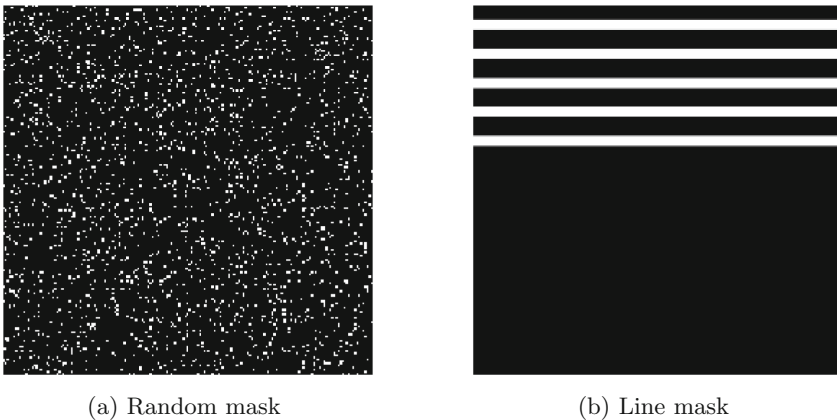


Fig. 1. Types of image mask

As we mentioned earlier, we discard around 15% of the pixels randomly. Figure 2 reveals that the difference between models’ prediction is undetectable visually. Moreover, Table 1 shows that the difference between the models is not significant. There is a slight advantage for the use of GIDM model compared



Fig. 2. Models' performance on random masked images. 1st column is for the ground truth images, 2nd column is for the masked images, 3rd column is for the GM prediction, 4th column is for the DM prediction, 5th column is for the IDM prediction, 6th column is for the GIDM prediction

to the others. Thus, we conclude that this approach of models' evaluation is not appropriate. For that, we decide to remove slightly thick lines of pixels and re-evaluate the models.

To evaluate the models' performance, we used TensorFlow to calculate the PSNR and Skimage python package for the SSIM metric.

After discarding lines from the images, we are able to generate back again the missing pixels, and Fig. 3 demonstrates that GIDM is the most efficient model among all the others. This is also clear in Table 2, where we can notice in the chosen images that GIDM is the most accurate re-generator of discarded pixels. Therefore, our work has shown that image data is better represented by

Table 1. Models' evaluation for the randomly masked images.

	PSNR	SSIM
GM	21.832	0.853
DM	25.963	0.878
IDM	25.672	0.875
GIDM	26.126	0.887



Fig. 3. Models' performance on line masked images. 1st column is for the ground truth images, 2nd column is for the masked images, 3rd column is for the GM prediction, 4th column is for the DM prediction, 5th column is for the IDM prediction, 6th column is for the GIDM prediction

Table 2. Models' evaluation for the line masked images.

	PSNR	SSIM
GM	20.366	0.833
DM	25.851	0.856
IDM	27.673	0.868
GIDM	29.398	0.891

generalized inverted Dirichlet. It is noteworthy to mention that these models' performance is hugely dependent on the size of the masks, the hyper-parameters, the type and order of the neighbouring system.

4 Conclusion

In this paper, we have developed a new optimal predictor based on finite generalized inverted Dirichlet mixtures. The GID demonstrates its efficiency in representing positive vectors due to its statistical characteristics through its covariance structure.

We learnt the model parameters using MLE approach with Newton Raphson method, and we considered the NSHP neighbouring system to compute the predicted pixel. We evaluated the GID optimal predictor on image inpainting and we compared the proposed model to other similar related works. The experimental results demonstrate its capability that offers reliable prediction and modeling potential.

References

1. Al Mashrgy, M., Bdiri, T., Bouguila, N.: Robust simultaneous positive data clustering and unsupervised feature selection using generalized inverted dirichlet mixture models. *Knowl. Based Syst.* **59**, 182–195 (2014)
2. Mashrgy, M.A., Bouguila, N.: A fully Bayesian framework for positive data clustering. In: Laaloui, Y., Bouguila, N. (eds.) *Artificial Intelligence Applications in Information and Communication Technologies*. SCI, vol. 607, pp. 147–164. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19833-0_7
3. Bdiri, T., Bouguila, N.: An infinite mixture of inverted Dirichlet distributions. In: Lu, B.-L., Zhang, L., Kwok, J. (eds.) *ICONIP 2011*. LNCS, vol. 7063, pp. 71–78. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24958-7_9
4. Bdiri, T., Bouguila, N.: Learning inverted dirichlet mixtures for positive data clustering. In: Kuznetsov, S.O., Ślezak, D., Hepting, D.H., Mirkin, B.G. (eds.) *RSFD-GrC 2011*. LNCS (LNAI), vol. 6743, pp. 265–272. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21881-1_42
5. Bdiri, T., Bouguila, N.: Positive vectors clustering using inverted dirichlet finite mixture models. *Expert Syst. Appl.* **39**(2), 1869–1882 (2012)
6. Bdiri, T., Bouguila, N.: Bayesian learning of inverted dirichlet mixtures for SVM kernels generation. *Neural Comput. Appl.* **23**(5), 1443–1458 (2013)
7. Bdiri, T., Bouguila, N., Ziou, D.: Visual scenes categorization using a flexible hierarchical mixture model supporting users ontology. In: *25th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2013*, Herndon, VA, USA, 4–6 November 2013, pp. 262–267. IEEE Computer Society (2013)
8. Bdiri, T., Bouguila, N., Ziou, D.: Object clustering and recognition using multi-finite mixtures for semantic classes and hierarchy modeling. *Expert Syst. Appl.* **41**(4), 1218–1235 (2014)
9. 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)
10. Bdiri, T., Bouguila, N., Ziou, D.: Variational Bayesian inference for infinite generalized inverted dirichlet mixtures with feature selection and its application to clustering. *Appl. Intell.* **44**(3), 507–525 (2016). <https://doi.org/10.1007/s10489-015-0714-6>
11. Bouguila, N.: Non-gaussian mixture image models prediction. In: *2008 15th IEEE International Conference on Image Processing*, pp. 2580–2583. IEEE (2008)
12. Bourouis, S., Al Mashrgy, M., Bouguila, N.: Bayesian learning of finite generalized inverted dirichlet mixtures: application to object classification and forgery detection. *Expert Syst. Appl.* **41**(5), 2329–2336 (2014)
13. Boutemedjet, S., Bouguila, N., Ziou, D.: A hybrid feature extraction selection approach for high-dimensional non-Gaussian data clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(8), 1429–1443 (2008)

14. 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). <https://doi.org/10.1007/s11042-018-5808-9>
15. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. Ser. B (Methodol.)* **39**(1), 1–22 (1977)
16. Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A.A.: What makes Paris look like Paris? *ACM Trans. Graph. (SIGGRAPH)* **31**(4), 101:1–101:9 (2012)
17. Elguebaly, T., Bouguila, N.: Finite asymmetric generalized Gaussian mixture models learning for infrared object detection. *Comput. Vis. Image Underst.* **117**(12), 1659–1671 (2013)
18. Fan, W., Bouguila, N.: Topic novelty detection using infinite variational inverted dirichlet mixture models. In: Li, T., et al. (eds.) 14th IEEE International Conference on Machine Learning and Applications, ICMLA 2015, Miami, FL, USA, 9–11 December 2015, pp. 70–75. IEEE (2015)
19. Graja, O., Bouguila, N.: Finite inverted dirichlet mixture optimal pixel predictor. In: 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp. 1–5. IEEE (2019)
20. Grimmett, G., Grimmett, G.R., Stirzaker, D., et al.: *Probability and Random Processes*. Oxford University Press, New York (2001)
21. Guidara, R., Hosseini, S., Deville, Y.: Maximum likelihood blind image separation using nonsymmetrical half-plane Markov random fields. *IEEE Trans. Image Process.* **18**(11), 2435–2450 (2009)
22. Gupta, A., Song, D.: Generalized liouville distribution. *Comput. Math. Appl.* **32**(2), 103–109 (1996)
23. Kotz, S., Balakrishnan, N., Johnson, N.L.: *Continuous multivariate distributions, Volume 1: Models and applications*. Wiley, Hoboken (2004)
24. Li, Y., Sayood, K.: Lossless video sequence compression using adaptive prediction. *IEEE Trans. Image Process.* **16**(4), 997–1007 (2007)
25. Lindley, D., Rao, C.R.: *Advanced statistical methods in biometric research*. (1953)
26. Lingappaiah, G.: On the generalised inverted dirichlet distribution. *Demonstratio Mathematica* **9**(3), 423–433 (1976)
27. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440 (2015)
28. Maini, R., Aggarwal, H.: Study and comparison of various image edge detection techniques. *Int. J. Image Process. (IJIP)* **3**(1), 1–11 (2009)
29. Makhoul, J.: Linear prediction: a tutorial review. *Proc. IEEE* **63**(4), 561–580 (1975)
30. Mathai, A., Moschopoulos, P.: A multivariate inverted beta model. *Statistica* **57**, 189–198 (1997)
31. Matteoli, S., Diani, M., Corsini, G.: A tutorial overview of anomaly detection in hyperspectral images. *IEEE Aerosp. Electron. Syst. Mag.* **25**(7), 5–28 (2010)
32. Mosleh, A., Bouguila, N., Hamza, A.B.: Bandlet-based sparsity regularization in video inpainting. *J. Vis. Commun. Image Represent.* **25**(5), 855–863 (2014)
33. Mosleh, A., Bouguila, N., Hamza, A.B.: A video completion method based on bandlet transform. In: *Proceedings of the 2011 IEEE International Conference on Multimedia and Expo, ICME 2011, Barcelona, Catalonia, Spain, 11–15 July 2011*, pp. 1–6. IEEE Computer Society (2011)
34. Mosleh, A., Bouguila, N., Hamza, A.B.: Automatic inpainting scheme for video text detection and removal. *IEEE Trans. Image Process.* **22**(11), 4460–4472 (2013)

35. Ng, K.W., Tian, G.L., Tang, M.L.: *Dirichlet and Related Distributions: Theory, Methods and Applications*, vol. 888. Wiley, Hoboken (2011)
36. Sefidpour, A., Bouguila, N.: Spatial color image segmentation based on finite non-Gaussian mixture models. *Expert Syst. Appl.* **39**(10), 8993–9001 (2012)
37. Simon, T., Joo, H., Matthews, I., Sheikh, Y.: Hand keypoint detection in single images using multiview bootstrapping. In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1145–1153 (2017)
38. Therrien, C.W., Quatieri, T.F., Dudgeon, D.E.: Statistical model-based algorithms for image analysis. *Proc. IEEE* **74**(4), 532–551 (1986)
39. Tirdad, P., Bouguila, N., Ziou, D.: Variational learning of finite inverted dirichlet mixture models and applications. In: Laalaoui, Y., Bouguila, N. (eds.) *Artificial Intelligence Applications in Information and Communication Technologies*. SCI, vol. 607, pp. 119–145. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19833-0_6
40. Zhang, J., Ma, D.: Nonlinear prediction for gaussian mixture image models. *IEEE Trans. Image Process.* **13**(6), 836–847 (2004)