# A GENERATIVE MODEL FOR SPATIAL COLOR IMAGE DATABASES CATEGORIZATION

*Nizar Bouguila and Walid ElGuebaly*

Concordia Institute for Information Systems Engineering
Concordia University
Montreal, Qc, Canada H3G 1T7
bouguila@ciise.concordia.ca, w_elgue@encs.concordia.ca

**ABSTRACT**

In this paper we analyze the problem of image databases categorization using a statistical generative model. Our model is based on the multinomial generalized Dirichlet distribution recently introduced to model discrete data. The model integrates also the spatial information with color histograms. We designed experiments to show the merits of our model.

***Index Terms*—** Generalized Dirichlet distribution, multinomial, image databases, prior distribution, mixture models.

## 1. INTRODUCTION

Images categorization is the process of grouping images into different classes. With the amount of digital information growing rapidly, the need for efficient automatic images categorization techniques has increased. Images categorization can be used for content-based images retrieval or browsing. A variety of techniques have been proposed to retrieve this digital content [1]. Although different, all these techniques agree on the fact that an efficient categorization scheme plays an important role. Color histograms are widely used as features vectors for images summarization and retrieval [2] and are used in different systems [1]. This can be explained by the fact that histograms provide a stable object recognition in the presence of occlusions and over views change [2]. However, histograms do not include any spatial information which is an important issue in human visual perception. Indeed, images with different appearance may have similar histograms which is a critical problem in large image databases. Different approaches have been proposed to integrate spatial information with color histograms [3, 4]. In this paper, we propose a generative model to take into account at the same time both the color and spatial information. Our model is based on the generalized Dirichlet distribution taken as a prior to the multinomial. Our generative model is presented in the next section. In section 3, we estimate the parameters of this model. Section 4, is devoted to experimental results.

The completion of this research was made possible thanks to the Natural Sciences and Engineering Research Council of Canada (NSERC), a NATEQ Nouveaux Chercheurs Grant, and a start-up grant from Concordia University.

## 2. THE GENERATIVE MODEL

Suppose that we have $N$ labeled images $\mathcal{I}_i$, $i = 1, \dots, N$ classified in $R$ classes and that the number of labeled images in each class $r$ is equal to $n_r$ ($\sum_{r=1}^{R} n_r = N$). By associating a distribution and a weight $p(r)$ to each class in the training set, we can suppose that each image $\mathcal{I}_i$ is generated by a mixture of $R$ distributions with parameters $\vec{\pi} = (\vec{\pi}_1, \dots, \vec{\pi}_R)$

$$p(\mathcal{I}_i|\vec{\pi}) = \sum_{r=1}^{R} p(r)p(\mathcal{I}_i|\vec{\pi}_r) \quad (1)$$

The problem now is the determination of $p(\mathcal{I}_i|\vec{\pi}_r)$. For this, let us introduce some notations. An $L \times K$ image $\mathcal{I}_i$ is considered to be a set of pixels $\{X_{i_{lk}}, l = 1, \dots, L; k = 1, \dots, K\}$, where $X_{i_{lk}}$ is the pixel in position $(l, k)$ of image $\mathcal{I}_i$. The colors in $\mathcal{I}_i$ are quantized into $C$ colors $c_1, \dots, c_C$. The distribution $p(\mathcal{I}_i|\vec{\pi}_r)$ can be described in terms of the features of the image. In our case, the features are the pixels. In order to introduce the spatial information, the probability of a pixel should be conditioned on its neighborhood. By taking the neighborhood consisting of the pixels at a distance $d \in D = \{d_1, \dots d_D\}$ measured using the $L_\infty$ norm, $p(\mathcal{I}_i|\vec{\pi}_r)$ will be given by

$$p(\mathcal{I}_i|\vec{\pi}_r) = \prod_{d=1}^{D}\prod_{l=1}^{L}\prod_{k=1}^{K} p(X_{i_{lk}}|\vec{\pi}_r; X_{i_{l'k'}}, d) \quad (2)$$

where $|(l, k) - (l', k')| = \max\{|l - l'|, |k - k'|\} = d$. Note that Eq. 2 will represent the classic image histogram, if we suppose that each pixel $X_{i_{lk}}$ is independent of its neighborhood, which is actually the standard naive Bayes assumption [5]. According to Eq. 2 the parameters of an individual mixture component are a multinomial distribution over the $C \times C$ possible color pairs and can be written as $\pi_{c_{t_1}, c_{t_2}, d|r}$, where $t_1, t_2 = 1, \dots, C$ and $\pi_{c_{t_1}, c_{t_2}, d|r} = p\Big(X_{i_{lk}} = c_{t_1}, X_{i_{l'k'}} = c_{t_2} \big| |(l, k) - (l', k')| = d\Big), l, l' = 1, \dots, L, k, k' = 1, \dots, K$, which is the probability that a pixel of color $c_{t_1}$ has at a distance $d$ a pixel of color $c_{t_2}$. Then, Eq. 2 could be written as a multinomial distribution

$$p(\mathcal{I}_i|\vec{\pi}_r) = \prod_{d=1}^{D}\prod_{c_{t_1}=1}^{C}\prod_{c_{t_2}=1}^{C}\pi_{c_{t_1},c_{t_2},d|r}^{f_{c_{t_1},c_{t_2},d}} \tag{3}$$

$f_{c_{t_1},c_{t_2},d} \equiv Card\{(X_{i_{lk}}, X_{i_{l'k'}}) = (c_{t_1}, c_{t_2}) \big| |(l,k)-(l',k')| = d\}$ and $Card\{\}$ refers to the number of elements of a set. Learning our model consists of estimating the parameters $\vec{\pi}_r$ using the $n_r$ labeled images in class $r$. Having all the parameters describing Eq. 1, we can assign a given test image $\mathcal{I}_t$ to a particular mixture component in Eq. 1 by using the Bayes' rule: $\mathcal{I}_t \longmapsto \arg\max_r p(r)p(\mathcal{I}_t|\vec{\pi}_r)$, where $p(r) = \frac{n_r}{N}$ [6].

## 3. PARAMETERS ESTIMATION

### 3.1. Prior Distributions

By associating a $C^2$-dimensional vector of frequencies $\vec{f}_{i,d} = (f_{c_1,c_1,d} = f_{i1}, \ldots, f_{c_1,c_C,d} = f_{ic_1\times c_C}, \ldots, f_{c_C,c_C,d} = f_{iC^2})$ to each image $\mathcal{I}_i$ for each distance $d$, the parameters can be estimated as the following:

$$\pi_{c_{t_1},c_{t_2},d|r} = \frac{f_{c_{t_1},c_{t_2},d}}{\sum_{c_{t_1}=1}^{C}\sum_{c_{t_2}=1}^{C} f_{c_{t_1},c_{t_2},d}} = \frac{f_{ic_{t_1}\times c_{t_2}}}{\sum_{c=1}^{C^2} f_{ic}} \tag{4}$$

which gives poor estimate [7]. A possible solution to overcome this problem is to assign a single Dirichlet or a Dirichlet mixture prior to the parameter vector $\vec{\pi}_r$ to moderate the extreme estimates given by Eq. 6 [7]. The Dirichlet distribution with $C^2$ positive parameters $\vec{\alpha} = (\alpha_1, \ldots, \alpha_{C^2})$ is defined by

$$p(\vec{\pi}_r|\vec{\alpha}) = \frac{\Gamma(\sum_{c=1}^{C^2}\alpha_c)}{\prod_{c=1}^{C^2}\Gamma(\alpha_c)}\prod_{c=1}^{C^2}\pi_c^{\alpha_c-1} \tag{5}$$

where $(\pi_{c_1,c_1,d|r} = \pi_1, \ldots, \pi_{c_1,c_C,d|r} = \pi_c, \ldots, \pi_{c_C,c_1,d|r} = \pi_{c_C\times c_1}, \ldots, \pi_{c_C,c_C,d|r} = \pi_{C^2})$. Using the Dirichlet as a prior, it is easy to show that we obtain the following estimate [7]:

$$\pi_{c_{t_1},c_{t_2},d|r} = \frac{f_{ic_{t_1}\times c_{t_2}} + \alpha_{c_{t_1}\times c_{t_2}}}{\sum_{c_{t_1}=1}^{C}\sum_{c_{t_2}=1}^{C}(f_{ic_{t_1}\times c_{t_2}} + \alpha_{c_{t_1}\times c_{t_2}})} \tag{6}$$

In spite of its flexibility and the fact that it is conjugate to the multinomial the Dirichlet has a very restrictive covariance matrix [8]. Another restriction of the Dirichlet distribution is that the variables with the same mean must have the same variance. All these disadvantages can be handled by using the generalized Dirichlet distribution. The generalized Dirichlet pdf in our case is defined by [9]

$$p(\vec{\pi}_r|\vec{\xi}) \quad = \quad \prod_{c=1}^{C^2-1}\frac{\Gamma(\alpha_c+\beta_c)}{\Gamma(\alpha_c)\Gamma(\beta_c)}\pi_c^{\alpha_c-1}(1-\sum_{l=1}^{c}\pi_c)^{\gamma_c}$$

where $\vec{\xi} = (\alpha_1, \beta_1, \ldots, \alpha_{C^2-1}, \beta_{C^2-1})$ ,$\alpha_c > 0$, $\beta_c > 0$, $\gamma_c = \beta_c - \alpha_{c+1} - \beta_{c+1}$ for $c = 1 \ldots C^2 - 2$ and $\gamma_{C^2-1} = \beta_{C^2-1} - 1$. The generalized Dirichlet distribution is reduced to a Dirichlet distribution with $(\alpha_1, \ldots, \alpha_{C^2-1}, \alpha_{C^2} = \beta_{C^2-1})$ when $\beta_c = \alpha_{c+1} + \beta_{c+1}$. Thus, the generalized Dirichlet includes the Dirichlet as a special case. Comparing to the Dirichlet, the generalized Dirichlet has $C^2 - 2$ extra parameters which is a very important advantage. Indeed, as the Dirichlet has $C^2$ parameters, when constructing a Dirichlet prior and if the mean probabilities of the variables have been fixed, it remains only one degree of freedom (by fixing the value of $\sum_{c=1}^{C^2}\alpha_c$) to adjust the distribution. For the generalized Dirichlet, however, it remains $C^2 - 1$ degrees of freedom which makes it more flexible for several applications [9]. We can show that the mean of the generalized Dirichlet distribution is [9, 10]

$$E(\pi_c) = \frac{\alpha_c}{\alpha_c+\beta_c}\prod_{k=1}^{c-1}\frac{\beta_k}{\alpha_k+\beta_k} \tag{7}$$

In addition to these properties, we can easily show that the generalized Dirichlet is conjugate to the multinomial distribution and using it as a prior gives us the following estimate

$$\begin{aligned}\pi_{c_{t_1},c_{t_2},d|r} \quad &= \quad \frac{\alpha_{c_{t_1}\times c_{t_2}} + f_{ic_{t_1}\times c_{t_2}}}{\alpha_{c_{t_1}\times c_{t_2}} + \beta_{c_{t_1}\times c_{t_2}} + n_{c_{t_1}\times c_{t_2}}} \\ &\times \prod_{l=1}^{c_{t_1}\times c_{t_2}-1}\frac{\beta_l + n_{l+1}}{\alpha_l+\beta_l+n_l}\end{aligned} \tag{8}$$

where $n_{c_{t_1}\times c_{t_2}} = f_{c_{t_1},c_{t_2},d} \ldots + f_{c_C,c_C,d}$. When $\beta_l = \alpha_{l+1} + \beta_{l+1}$, it is straightforward to verify that this equation is reduced to Eq. 6.

### 3.2. Estimation

By using the generalized Dirichlet as a prior, we suppose actually that the vectors $\mathcal{F}_r = \{\vec{f}_{i,d}, i = 1, \ldots, n_r\}$ in each training set follow a multinomial generalized Dirichlet as follows

$$p(\vec{f}_{i,d}|\vec{\xi}) = \prod_{c=1}^{C^2-1}\frac{\Gamma(\alpha_c+\beta_c)}{\Gamma(\alpha_c)\Gamma(\beta_c)}\prod_{c=1}^{C^2-1}\frac{\Gamma(\alpha_c^{'})\Gamma(\beta_c^{'})}{\Gamma(\alpha_c^{'}+\beta_c^{'})}$$

where $\alpha_c^{'} = \alpha_c + f_{ic}$ and $\beta_c^{'} = \beta_c + f_{ic+1} + \ldots + f_{iC^2}$ for $c = 1, \ldots, C^2 - 1$. The log-likelihood corresponding to this distribution is given by

$$L(\mathcal{F}|\vec{\xi}) = \sum_{i=1}^{n_r}\sum_{c=1}^{C^2-1}\log\left(\frac{\Gamma(\alpha_c+\beta_c)}{\Gamma(\alpha_c)\Gamma(\beta_c)}\frac{\Gamma(\alpha_c^{'})\Gamma(\beta_c^{'})}{\Gamma(\alpha_c^{'}+\beta_c^{'})}\right) \tag{9}$$

The estimation of the parameters is based on the maximization of this function. For this goal, we have used a Newton-Raphson method based on the computation of the first and second derivatives.

$$\frac{\partial L(\mathcal{F}|\vec{\xi})}{\partial\alpha_c} = \sum_{i=1}^{n_r}\left(\Psi(\alpha_c+\beta_c) - \Psi(\alpha_c) + \Psi(\alpha_c^{'}) - \Psi(\alpha_c^{'}+\beta_c^{'})\right)$$

$$\frac{\partial L(\mathcal{F}|\vec{\xi})}{\partial \beta_c} = \sum_{i=1}^{n_r}\Big(\Psi(\alpha_c+\beta_c)-\Psi(\beta_c)+\Psi(\beta_c^{'})-\Psi(\alpha_c^{'}+\beta_c^{'})\Big)$$

By computing the second and mixed derivatives we obtain

$$\frac{\partial^2 L(\mathcal{F}|\vec{\xi})}{\partial \alpha_{c_1}\partial \alpha_{c_2}} = \begin{cases} \sum_{i=1}^{n_r}\Big(\Psi^{'}(\alpha_c+\beta_c)-\Psi^{'}(\alpha_c) & \\ \quad +\Psi^{'}(\alpha_c^{'})-\Psi^{'}(\alpha_c^{'}+\beta_c^{'})\Big) & \text{if } c_1=c_2=c \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

$$\frac{\partial^2 L(\mathcal{F}|\vec{\xi})}{\partial \beta_{c_1}\partial \beta_{c_2}} = \begin{cases} \sum_{i=1}^{n_r}\Big(\Psi^{'}(\alpha_c+\beta_c)-\Psi^{'}(\beta_c) & \\ \quad +\Psi^{'}(\beta_c^{'})-\Psi^{'}(\alpha_c^{'}+\beta_c^{'})\Big) & \text{if } c_1=c_2=c \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

$$\frac{\partial^2 L(\mathcal{F}|\vec{\xi})}{\partial \beta_{c_1}\partial \alpha_{c_2}} = \frac{\partial^2 L(\mathcal{F}|\vec{\xi})}{\partial \alpha_{c_1}\partial \beta_{c_2}} = \begin{cases} \sum_{i=1}^{n_r}\Big(\Psi^{'}(\alpha_c+\beta_c) & \\ \quad -\Psi^{'}(\alpha_c^{'}+\beta_c^{'})\Big) & \text{if } c_1=c_2=c \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where $\Psi$ and $\Psi^{'}$ are the digamma and trigamma functions. Then, the Hessian matrix has a block-diagonal structure

$$H(\vec{\xi}) = \text{block-diag}\big\{H_1(\alpha_1,\beta_1),\ldots,H_{C^2}(\alpha_{C^2},\beta_{C^2})\big\} \tag{13}$$

where

$$H_c(\alpha_c,\beta_c) = \begin{pmatrix} \frac{\partial^2 L(\mathcal{F}|\vec{\xi})}{\partial^2 \alpha_c} & \frac{\partial^2 L(\mathcal{F}|\vec{\xi})}{\partial \alpha_c \partial \beta_c} \\ \frac{\partial^2 L(\mathcal{F}|\vec{\xi})}{\partial \beta_c \partial \alpha_c} & \frac{\partial^2 L(\mathcal{F}|\vec{\xi})}{\partial^2 \beta_c} \end{pmatrix} \tag{14}$$

Given a set of initial estimates, Newton-Raphson method can now be used to estimate the parameters:

$$\vec{\xi}^{(t)} = \vec{\xi}^{(t-1)} - H(\vec{\xi}^{(t-1)})^{-1}\frac{\partial L(\mathcal{F}|\vec{\xi})}{\partial \vec{\xi}^{(t-1)}} \tag{15}$$

where

$$H(\vec{\xi})^{-1} = \text{block-diag}\big\{H_1(\alpha_1,\beta_1)^{-1},\ldots,H_v(\alpha_{C^2},\beta_{C^2})^{-1}\big\} \tag{16}$$

## 4. EXPERIMENTAL RESULTS: SPATIAL COLOR IMAGE DATABASES SUMMARIZATION

For our experiments, we used a database containing 45100 images. This database contains 10 homogeneous classes (see Figure 1). We divided the database on two sets. A data set containing 22550 images used for training. The remaining images were used for testing. The repartition of the different classes in the training and test sets is given in table 1. We considered the RGB space with color quantization into 512 colors ($8 \times 8 \times 8$) and the set of distances $D = \{1,3,5,7,9,11\}$. Besides, we have considered only probabilities of pixels having same colors in order to reduce zero frequencies, which is a common approach and used, for instance, in the case of the autocorrelogram proposed by Huang et al. [4]. The accuracy classification produced by our classifier was measured by counting the number of misclassified images, yielding a confusion matrix. In this confusion matrix, the cell $(i,j)$ represents the number of images from category $i$ which are classified as category $j$. The number of images misclassified when we used generalized Dirichlet as a prior, was 2189, which represents an accuracy of 90.29 percent (See Table 2). Table 3 represents the confusion matrix when we used a Dirichlet (3711 misclassified images which represents an accuracy of 83.54 percent). Table 4 shows the confusion matrix when we use only the frequencies. In this case, the accuracy was 80.35 percent (4430 misclassified images). We have also tested the

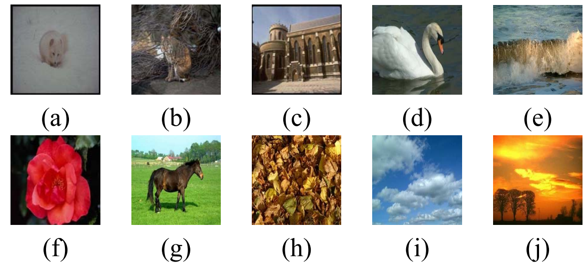


**Fig. 1**. Sample images from each group. (a) Class1, (b) Class2, (c) Class3, (d) Class4, (e) Class5, (f) Class6, (g) Class7, (h) Class8, (i) Class9, (j) Class10,

**Table 1**. Repartition of the different classes in the training and test sets.

| class | Training set | Testing set |
|---|---|---|
| Class1 | 2250 | 2250 |
| Class2 | 2500 | 2500 |
| Class3 | 3000 | 3000 |
| Class4 | 1900 | 1900 |
| Class5 | 2000 | 2000 |
| Class6 | 2100 | 2100 |
| Class7 | 2250 | 2250 |
| Class8 | 2200 | 2200 |
| Class9 | 2050 | 2050 |
| Class10 | 2300 | 2300 |

| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 |
|---|---|---|---|---|---|---|---|---|---|---|
| C1 | 1961 | 89 | 21 | 45 | 28 | 22 | 14 | 27 | 20 | 23 |
| C2 | 32 | 2168 | 21 | 89 | 17 | 33 | 45 | 37 | 22 | 36 |
| C3 | 18 | 71 | 2632 | 23 | 37 | 22 | 28 | 31 | 84 | 54 |
| C4 | 51 | 35 | 14 | 1700 | 26 | 12 | 15 | 18 | 19 | 10 |
| C5 | 29 | 17 | 21 | 52 | 1775 | 13 | 14 | 18 | 37 | 24 |
| C6 | 13 | 25 | 7 | 10 | 10 | 1925 | 31 | 49 | 5 | 25 |
| C7 | 17 | 55 | 16 | 19 | 20 | 23 | 2033 | 27 | 6 | 34 |
| C8 | 23 | 36 | 4 | 8 | 12 | 25 | 17 | 2047 | 8 | 20 |
| C9 | 23 | 7 | 3 | 15 | 25 | 2 | 4 | 3 | 1949 | 19 |
| C10 | 9 | 21 | 9 | 5 | 13 | 12 | 10 | 32 | 18 | 2171 |

**Table 2**. Confusion matrix for image classification using multinomial generalized Dirichlet.

| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 |
|---|---|---|---|---|---|---|---|---|---|---|
| C1 | 1739 | 123 | 43 | 81 | 49 | 41 | 29 | 53 | 41 | 51 |
| C2 | 49 | 1918 | 49 | 145 | 47 | 51 | 86 | 62 | 39 | 54 |
| C3 | 37 | 118 | 2347 | 48 | 67 | 37 | 56 | 60 | 135 | 95 |
| C4 | 89 | 67 | 26 | 1575 | 37 | 17 | 21 | 23 | 28 | 17 |
| C5 | 48 | 29 | 36 | 80 | 1674 | 18 | 19 | 21 | 46 | 29 |
| C6 | 17 | 39 | 13 | 17 | 19 | 1822 | 48 | 66 | 16 | 43 |
| C7 | 30 | 79 | 23 | 31 | 29 | 36 | 1927 | 45 | 8 | 42 |
| C8 | 41 | 57 | 7 | 13 | 23 | 46 | 29 | 1932 | 15 | 37 |
| C9 | 44 | 11 | 7 | 27 | 46 | 5 | 9 | 15 | 1859 | 27 |
| C10 | 20 | 43 | 21 | 12 | 25 | 25 | 21 | 59 | 28 | 2046 |

**Table 3**. Confusion matrix for image classification using multinomial Dirichlet.

| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 |
|---|---|---|---|---|---|---|---|---|---|---|
| C1 | 1670 | 131 | 49 | 86 | 55 | 44 | 36 | 61 | 53 | 65 |
| C2 | 53 | 1858 | 54 | 169 | 51 | 58 | 86 | 67 | 47 | 57 |
| C3 | 43 | 132 | 2247 | 53 | 76 | 45 | 67 | 72 | 153 | 112 |
| C4 | 98 | 72 | 29 | 1515 | 45 | 23 | 25 | 29 | 37 | 27 |
| C5 | 51 | 31 | 39 | 89 | 1624 | 21 | 29 | 25 | 57 | 34 |
| C6 | 26 | 44 | 23 | 25 | 24 | 1742 | 61 | 78 | 23 | 54 |
| C7 | 37 | 91 | 36 | 42 | 41 | 49 | 1819 | 63 | 16 | 56 |
| C8 | 44 | 61 | 12 | 17 | 28 | 52 | 35 | 1877 | 23 | 51 |
| C9 | 56 | 14 | 11 | 31 | 46 | 13 | 14 | 18 | 1812 | 35 |
| C10 | 27 | 51 | 27 | 17 | 36 | 37 | 29 | 71 | 49 | 1956 |

**Table 4**. Confusion matrix for image classification using multinomials.

representation of the image colors in the HSV space and we did not remark much changes in the results. From the results, we can conclude that the multinomial Dirichlet and the multinomial Generalized Dirichlet perform better than the multinomial. This can be explained by the sparseness problem; i.e the zero frequency problem [11]. Indeed, many frequencies are actually zero or have very small probabilities. Then, when the multinomial is used for modeling, prediction and classification, a large number of observations will be judged to be impossible based on the training data. The introduction of the Dirichlet and the generalized Dirichlet as priors can be viewed as smoothing technique to deal with this problem.

## 5. CONCLUSION

We have proposed, discussed and evaluated a generative model to improve the color histogram by the spatial information. This model is based on the introduction of the generalized Dirichlet as a prior to multinomial distributions. The recently proposed multinomial Dirichlet has turned out to be a special case. We have also addressed the problem of parameters estimation. The proposed model is powerful and flexible enough to be adapted to a broad variety of applications where discrete data plays an important role such as information retrieval and filtering, natural language processing and bioinformatics.

## 6. REFERENCES


[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Guptaand and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.

[2] M. Swain and D. Ballard, "Color Indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[3] N. Bouguila, "Spatial Color Image Databases Summarization," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2007), Volume1*, Honolulu, HI, USA, 2007, pp. 953–956.

[4] J. Huang, S. R. Kumar, M. Mitra, W. Zhu and R. Zabih, "Spatial Color Indexing and Applications," *International Journal of Computer Vision*, vol. 35, no. 3, pp. 245–268, 1999.

[5] K. Nigam, A. McCallum, S. Thrun and T. Mitchell, "Text Classification from Labeled and Unlabeled Documents using EM," *Machine Learning*, vol. 39, no. 2, pp. 103–134, 2000.

[6] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, Wiley, New York, 2001.

[7] N. Bouguila and D. Ziou, "Unsupervised Learning of a Finite Discrete Mixture: Applications to Texture Modeling and Image Databases Summarization," *Journal of Visual Communications and Image Representation*, vol. 18, no. 4, pp. 295–309, 2006.

[8] N. Bouguila, D. Ziou and J. Vaillancourt, "Unsupervised Learning of a Finite Mixture Model Based on the Dirichlet Distribution and its Application," *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1533–1543, 2004.

[9] N. Bouguila and D. Ziou, "A Hybrid SEM Algorithm for High-Dimensional Unsupervised Learning Using a Finite Generalized Dirichlet Mixture," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2657–2668, 2006.

[10] N. Bouguila and D. Ziou, "High-Dimensional Unsupervised Selection and Estimation of a Finite Generalized Dirichlet Mixture Model Based on Minimum Message Length," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1716–1731, 2007.

[11] S. Katz, "Estimation of Probabilities From Sparse Data For The Language Model Component of a Speech Recognizer," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 3, pp. 400–401, 1987.