SPATIAL COLOR IMAGE DATABASES SUMMARIZATION

Nizar Bouguila

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

ABSTRACT

This paper presents a finite discrete mixture model based on both the Dirichlet and the multinomial distributions to add spatial information to color histograms. The estimation of the parameters and the determination of the number of components in our model are based on the classification expectation-maximization approach and the integrated complete likelihood criterion, respectively. The developed model is applied with success for color images databases summarization.

Index Terms— Dirichlet distribution, multinomial, finite mixture models, Image databases.

1. INTRODUCTION

In recent years, there has been a tremendous increasing in the generation of digital images. As this content grows, the need for tools to summarize, filter and retrieve image databases becomes more accurate. A variety of techniques have been proposed to retrieve this content [1]. Although different, all these techniques agree on the fact that an efficient summarization scheme plays an important role. Summarizing an image database is very important because it simplifies the task of retrieval by restricting the search for similar images to a smaller domain of the database. Summarization is also very efficient for browsing. In this paper, we propose a summarization approach based on finite discrete mixture models. Color histograms are widely used as features vectors for images summarization [2]. However, histograms do not include any spatial information which is an important issue in human visual perception. One of the most successful approaches to integrate the spatial information with the color histograms is the color correlogram [3]. The color correlogram describes the spatial correlation of colors as a function of spatial distance. Let $\mathcal{I}$ be an $L \times C$ image composed of pixels $p(x, y)$. The colors in $\mathcal{I}$ are quantized into m colors $c_1, \ldots, c_m$. For a pixel p, let $\mathcal{I}(p)$ denotes its color. Let $\mathcal{I}_c = \{p | \mathcal{I}(p) = c\}$ and $D = \{d_1, \ldots, d_D\}$ a set of D fixed distances, which are measured using the L_∞ norm. The correlogram of image $\mathcal{I}$ is defined for color pair (c_i, c_j), $i, j = 0, \ldots, m$ and distance $d \in D$ as:

$$\gamma^d_{c_i,c_j}(\mathcal{I}) \equiv Pr\big[p_2 \in \mathcal{I}_{c_i} \,||p_1 - p_2| = d\big], p_1 \in \mathcal{I}_{c_i}, p_2 \in \mathcal{I} \quad (1)$$

Which gives the probability that given any pixel p_1 of color c_i, a pixel p_2 at a distance d from pixel p_1 is of color c_j. In order to compute the correlogram it suffices to compute the following count:

$$\Gamma^d_{c_i,c_j}(\mathcal{I}) \equiv Card\big\{(p_1, p_2) \in \mathcal{I}_{c_i} \times \mathcal{I}_{c_j} \,||p_1 - p_2| = d\big\} \quad (2)$$

where $Card\{\}$ refers to the number of elements of a set. Note that the size of the correlogram is $O(Dm^2)$, then a large D would result in expensive computation and large storage requirement [3]. In this paper, we propose to model the spatial color information using discrete finite mixture models by observing that for each color pair (c_i, c_j), we can associate a D-dimensional vector of counts described as follows:

$$\vec{f}_{c_i,c_j} = (f^{d_1}_{c_i,c_j}, \ldots, f^{d_D}_{c_i,c_j}) \quad (3)$$

where $f^d_{c_i,c_j}(\mathcal{I}) = Card\big\{(p_1, p_2) \in \mathcal{I}_{c_i} \times \mathcal{I}_{c_j} \,||p_1 - p_2| = d\big\}$. Then, the spatial color information is represented by m^2 D-dimensional vectors of counts which can be modeled by a discrete mixture.

2. THE FINITE DISCRETE MIXTURE MODEL

Finite mixtures can be viewed as a superimposition of a finite number of component densities [4]. The choice of the component model is very critical in mixture decomposition. The number of components required to model the mixture and the modeling capabilities are directly related to the component model used. For multivariate data, attention has focused on the use of multivariate Gaussian components. However, for count data, the Gaussian assumption turns out to be inadequate and the majority of the researchers consider the Multinomial distribution. If we suppose that $\vec{f}_{c_i,c_j}$ follows a Multinomial distribution with parameters $\vec{P} = (P_1, \ldots, P_D)$:

$$p(\vec{f}_{c_i,c_j} | \vec{P}) = \frac{(\sum_{d=1}^{D} f^d_{c_i,c_j})!}{f^{d_1}_{c_i,c_j}! f^{d_2}_{c_i,c_j}! \ldots f^{D}_{c_i,c_j}!} \prod_{d=1}^{D} P_d^{f^d_{c_i,c_j}} \quad (4)$$

where $P_D = 1 - \sum_{i=1}^{D-1} P_i$. Then, the frequencies will be used to set the probabilities, obtaining:

$$\hat{P}_d = \frac{f^d_{c_i,c_j}}{\sum_{l=1}^{D} f^l_{c_i,c_j}} \quad (5)$$

The frequencies alone are a poor estimate, however. An appropriate and efficient solution to address this issue is the introduction of a prior information into the construction of the statistical model. The prior information, for the Multinomial assumption, is given by the Dirichlet distribution to form *good* $\hat{P}_d$ estimates. The Dirichlet distribution with parameter vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_D)$ is defined by [5]:

$$p(P_1, \ldots, P_d) = \frac{\Gamma(\alpha_1 + \alpha_2 + \ldots + \alpha_D)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\ldots\Gamma(\alpha_D)} \prod_{d=1}^{D} P_d^{\alpha_d - 1} \quad (6)$$

The Dirichlet distribution depends on D parameters $\alpha_1, \ldots, \alpha_D$, which are all real and positive. This distribution is the multivariate extension of the 2-parameter Beta distribution. The mean of the Dirichlet distribution is given by:

$$E(P_d) = \frac{\alpha_d}{\sum_{l=1}^{D} \alpha_l} \quad (7)$$

The majority of the researchers assign a Dirichlet prior to the parameter vector of a multinomial distribution. This is due to the fact that the Dirichlet is a conjugate prior to the multinomial distribution, i.e the posterior is also Dirichlet. Indeed, we have:

$$p(\vec{f}_{c_i,c_j}, \vec{P}|\vec{\alpha}) = p(\vec{f}_{c_i,c_j}|\vec{P})p(\vec{P}|\vec{\alpha})$$
$$= \frac{\Gamma(\sum_{d=1}^{D} f^d_{c_i,c_j} + 1)}{\prod_{d=1}^{D} \Gamma(f^d_{c_i,c_j} + 1)} \frac{\Gamma(\sum_{d=1}^{D} \alpha_d)}{\prod_{d=1}^{D} \Gamma(\alpha_d)} \prod_{d=1}^{D} p_d^{f^d_{c_i,c_j} + \alpha_d - 1} \quad (8)$$

Integrating over $\vec{P}$, straightforward manipulations give us the marginal distribution of $\vec{X}_i$:

$$p(\vec{f}_{c_i,c_j}|\vec{\alpha}) = \int_{\vec{P}} p(\vec{f}_{c_i,c_j}, \vec{P}|\vec{\alpha}) d\vec{P}$$
$$= \frac{\Gamma(\sum_{d=1}^{D} f^d_{c_i,c_j} + 1)\Gamma(\sum_{d=1}^{D} \alpha_d)}{\Gamma(\sum_{d=1}^{D} f^d_{c_i,c_j} + \sum_{d=1}^{D} \alpha_d)} \prod_{d=1}^{D} \frac{\Gamma(f^d_{c_i,c_j} + \alpha_d)}{\Gamma(\alpha_d)\Gamma(f^d_{c_i,c_j} + 1)}$$

We call this density the multinomial Dirichlet distribution which is the multi-dimensional case of the widely studied Beta binomial distribution. Note that comparing to the multinomial, the multinomial Dirichlet has one extra degree of freedom, since its parameters are not constrained to sum to one, which makes it more practical [6]. Then, the posterior is given by:

$$p(\vec{P}|\vec{f}_{c_i,c_j}, \vec{\alpha}) = \frac{p(\vec{f}_{c_i,c_j}, \vec{P}|\vec{\alpha})}{p(\vec{f}_{c_i,c_j}|\vec{\alpha})}$$
$$= \frac{\Gamma(\sum_{d=1}^{D} f^d_{c_i,c_j} + \sum_{d=1}^{D} \alpha_l)}{\prod_{d=1}^{D} \Gamma(\alpha_d + f^d_{c_i,c_j})} \prod_{d=1}^{D} P_d^{\alpha_d + f^d_{c_i,c_j} - 1} \quad (9)$$

which is a Dirichlet with parameters $(\alpha_1 + f^{d_1}_{c_i,c_j}, \ldots, \alpha_D + f^{d_D}_{c_i,c_j})$. Using Eq. 9 and Eq. 7, we obtain:

$$\hat{P}_d = \frac{\alpha_l + f^d_{c_i,c_j}}{\sum_{l=1}^{D} \alpha_l + \sum_{l=1}^{D} f^l_{c_i,c_j}} \quad (10)$$

We can think that the hyperparameters α_d are hidden quantities added in order to represent our confidence about the estimates and to moderate the extreme estimates given by Eq. 5. As we can note, as the number of observations increases the estimates converge to Eq. 5. But, if the quantities α_d grow, our estimates tend to be further off from the estimates based just on the observed frequencies and given by Eq. 5. A multinomial Dirichlet mixture with M components is defined as:

$$p(\vec{f}_{c_i,c_j}|\Theta) = \sum_{k=1}^{M} p(\vec{f}_{c_i,c_j}|\vec{\alpha}_k)p(k) \quad (11)$$

where $p(k)$ ($0 < p(k) \leq 1$ and $\sum_{k=1}^{M} p(k) = 1$) are the mixing proportions and $p(\vec{f}_{c_i,c_j}|\vec{\alpha}_k)$ is the multinomial Dirichlet. The symbol Θ refers to the entire set of parameters to be estimated: $\Theta = (\vec{\alpha}_1, \ldots, \vec{\alpha}_M, p(1), \ldots, p(M))$ where $\vec{\alpha}_k = (\alpha_{k1}, \ldots, \alpha_{kD})$ is the parameter vector for the k^{th} population.

3. THE MULTINOMIAL DIRICHLET MIXTURE ESTIMATION AND SELECTION

3.1. Maximum Likelihood Estimation

Given the set of the m^2 independent vectors $\mathcal{F} = \{\vec{f}_{c_i,c_j}, i, j = 1 \ldots, m\}$, the log-likelihood corresponding to a M-component is:

$$L(\Theta, \mathcal{F}) = log \prod_{i,j}^{m} p(\vec{f}_{c_i,c_j}|\Theta) = \sum_{i,j}^{m} log \sum_{k=1}^{M} p(\vec{f}_{c_i,c_j}|\vec{\alpha}_k)p(k) \quad (12)$$

It is well known that the maximum likelihood (ML) estimate:

$$\hat{\Theta}_{ML} = arg \max_{\Theta}\{L(\Theta, \mathcal{F})\} \quad (13)$$

The ML estimates of the mixture parameters can be obtained using EM and related techniques [4]. The EM algorithm is a general approach to maximum likelihood in the presence of incomplete data. In EM, the "complete" data are considered to be $Y_{c_i,c_j} = \{\vec{f}_{c_i,c_j}, \vec{Z}_{c_i,c_j}\}$, where $\vec{Z}_{c_i,c_j} = (Z^1_{c_i,c_j}, \ldots, Z^M_{c_i,c_j})$, with:

$$Z^k_{c_i,c_j} = \begin{cases} 1 & \text{if } \vec{f}_{c_i,c_j} \text{ belongs to class } k \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

constituting the "missing" data. In this paper, we do not consider this approach, but the classification maximum likelihood approach (CML) [7]. The classification log-likelihood function is given by:

$$CL(\Theta, \mathcal{Z}, \mathcal{F}) = \sum_{k=1}^{M} \sum_{\vec{f}_{c_i,c_j} \in \mathcal{P}_k} log\big(p(\vec{f}_{c_i,c_j}|\vec{\alpha}_k)p(k)\big)$$

where $\mathcal{Z} = \{\vec{Z}_{c_i,c_j}, i,j = 1,\ldots,c\}$ and $\mathcal{P} = \{\mathcal{P}_1,\ldots,\mathcal{P}_k\}$ is a partition of $\mathcal{F}$ associated to $\mathcal{Z}$: $\mathcal{P}_k = \{\vec{f}_{c_i,c_j}/Z^k_{c_i,c_j} = 1\}$. In the classification approach of model-based clustering, the maximization of the classification log-likelihood function is based on the classification EM algorithm (CEM) [7]. Having an initial partition $\mathcal{P}^0$, the iteration t of CEM is composed of the following steps:

1. **E-step:** Compute $\hat{Z}^{k(t)}_{c_i,c_j}$:

$$\hat{Z}^{k(t)}_{c_i,c_j} = \frac{p(\vec{f}_{c_i,c_j}|\vec{\alpha}_k^{(t-1)})p(k)^{(t-1)}}{\sum_{k=1}^{M} p(\vec{f}_{c_i,c_j}|\vec{\alpha}_k^{(t-1)})p(k)^{(t-1)}} \quad (15)$$

2. **C-Step:** Assignment of the vectors $\vec{f}_{c_i,c_j}$ to clusters.

3. **M-step:** Update the parameter estimates for each component j using the partition $\mathcal{P}_j$ according to:

$$p(k)^{(t)} = \frac{n_k^{(t)}}{m^2} \quad (16)$$

$$\hat{\vec{\alpha}}_k = argmax_{\vec{\alpha}_k} \sum_{\vec{f}_{c_i,c_j} \in \mathcal{P}_k} log\big(p(\vec{f}_{c_i,c_j}|\vec{\alpha}_k)p(k)\big) \quad (17)$$

The quantity $\hat{Z}^k_{c_i,c_j}$ is the conditional probability that observation $\vec{f}_{c_i,c_j}$ belongs to class k (the *posterior* probability) and n_k is the number of vectors affected to cluster k. When maximizing Eq. 17, we do not obtain a closed-form solution for the $\vec{\alpha}_k$ parameters.
Let $CL(\vec{\alpha}_k, \mathcal{P}_k) = \sum_{\vec{f}_{c_i,c_j} \in \mathcal{P}_k} log\big(p(\vec{f}_{c_i,c_j}|\vec{\alpha}_k)p(k)\big)$. Then, the partial derivative of $CL(\vec{\alpha}_k, \mathcal{P}_k)$ with respect to α_{kd} is:

$$\frac{\partial CL(\vec{\alpha}_k, \mathcal{P}_k)}{\partial \alpha_{kd}} = n_k\left(\Psi(\sum_{d=1}^{D}\alpha_{kd}) - \Psi(\alpha_{kd})\right) \quad (18)$$
$$+ \sum_{\vec{f}_{c_i,c_j} \in \mathcal{P}_k}\left(\Psi(\alpha_{kd} + f^d_{c_i,c_j}) - \Psi(\sum_{d=1}^{D}(\alpha_{kd} + f^d_{c_i,c_j}))\right)$$

and the estimation of α_{kd} is based on the following iteration scheme:

$$\alpha_{kd}^{(t)} = \frac{\alpha_{kd}^{(t-1)}\left(\sum_{\vec{f}_{c_i,c_j} \in \mathcal{P}_k}\Psi(\alpha_{kd} + f^d_{c_i,c_j}) - n_k\Psi(\alpha_{kd})\right)}{\sum_{\vec{f}_{c_i,c_j} \in \mathcal{P}_k}\Psi(\sum_{d=1}^{D}(\alpha_{kd} + f^d_{c_i,c_j})) - n_k\Psi(\sum_{d=1}^{D}\alpha_{kd})} \quad (19)$$

3.2. Complete Algorithm of Estimation and Selection

In order to determine the number of clusters, we have used the integrated complete likelihood (ICL) criterion proposed in [7]:

$$ICL(M) = CL(\Theta, \mathcal{Z}, \mathcal{X}) - \frac{N_p}{2}\log N \quad (20)$$

where $N_p = M(d+2)$ is the number of free parameters in the mixture model. Having the ICL criterion and the initialization algorithm presented in [5] in hand, the complete algorithm for estimation and selection is as the following:
Algorithm
For each candidate value of M:

1. Apply the INITIALIZATION Algorithm [5].

2. E-Step: Compute the *posterior* probabilities $\hat{Z}_{ij}^{(t)}$ using Eq. 15

3. C-Step: Assignment of the vectors $\vec{X}_i$ to clusters. The classification of an observation $\vec{X}_i$ is taken to be $\{k/Z^*_{ik} = max_j Z^*_{ij}\}$, which is the Bayes rule.

4. M-Step:
   (a) Update the $p(j)^{(t)}$ using Eq. 16.
   (b) Update the $\vec{\alpha}_j^{(t)}$ using Eq. 19

5. Calculate the associated criterion ICL(M) using Eq. 20.

6. Select the optimal model M^* such that:

$$M^* = arg\ max_M ICL(M)$$

4. EXPERIMENTAL RESULTS

In order to validate our model, we use it for color images summarization. Our summarization approach is based on a classifier. The inputs to the classifier are images from different image database classes. These images are separated into unknown or test set of images, whose class is unknown, and the training set of images, whose class is known. The training set is necessary to adapt the classifier to each possible class before the unknown set is submitted to the classifier. All the input images are passed through the $\vec{\Gamma}_{c_i,c_j}$ computation stage, and then through the mixture's parameters estimation and selection stage, in which the spatial color information is modeled as a Multinomial Dirichlet mixture. After this stage, each class in the database is represented by a Multinimial Dirichlet mixture. Finally, the classification stage uses mixtures estimated from the unknown images to determine in which class they will be assigned. The estimated mixture is compared to the training mixtures by using the Battacharya distance between densities, $D(I,\omega_j) = \int_{\vec{C}}\sqrt{p(\vec{C}|I)p(\vec{C}|\omega_j)d\vec{C}}$, where

I and ω_j index, respectively, a Multinimial Dirichlet mixture for an image to be classified $p(\vec{C}|I)$ and a Multinimial Dirichlet mixture obtained from training and which represents a class $p(\vec{C}|\omega_j)$. Classification is performed using this rule: image represented by the Multinimial Dirichlet mixture I is assigned to class ω_{j_1} if $D(I, \omega_j) > D(I, \omega_{j_1})$ $\forall j \neq j_1$.
For our experiment, we used a database containing 12850 images. This database contains 10 classes (see Figure 1). We divided the database on two sets. A data set containing 6425 images used for training. The remaining images were used for testing. We considered the RGB space with color quantization into 64 color and the set of distances $D = \{1, 3, 5, 7, 9, 11\}$. The accuracy classification produced by our classifier was

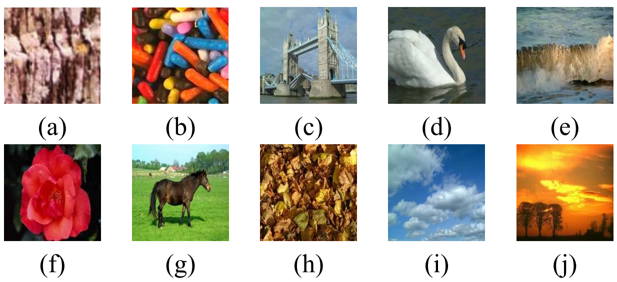

Fig. 1. Sample images from each group.

measured by counting the number of misclassified images, yielding a confusion matrix. The number of images misclassified when we used multinomial Dirichlet mixtures, was 637, which represents an accuracy of 90.08 percent. In this case of the multinomial mixture, the accuracy was 80.98 percent (1222 misclassified images). Note that the accuracy improvement when using multinomial Dirichlet mixture is statistically significant as shown by a Student's t-test. Figure 2 shows the accuracy of the classification, when using the two mixtures, as a function of the number of images in the training set. We can see clearly that the accuracy increase as we add images in the training set.

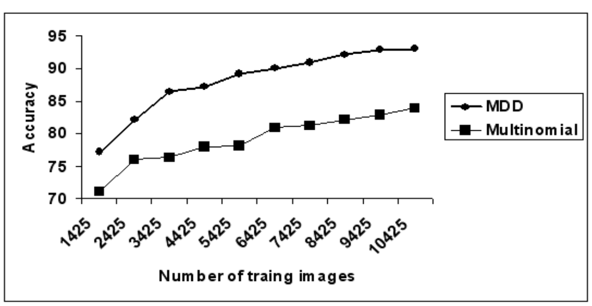

Fig. 2. Accuracy, using spatial color information, as a function of the number of images in the training set

5. SUMMARY AND CONCLUSIONS

Multinomial Dirichlet mixture models are an effective approach to model the spatial color information. The effectiveness of the our model was shown experimentally through an application which involves image databases summarization. Future work can be devoted to the use of this model for objects modeling and recognition. Another interesting extension of this model could be the use of the generalized Dirichlet mixture [8] as a prior to the multinomial.

Acknowledgments

The completion of this research was made possible thanks to a start-up grant from Concordia University.

6. REFERENCES

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Trans. PAMI*, 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] 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.

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

[5] 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.

[6] R. E. Madsen, D. Kauchak and C. Elkan, "Modeling Word Buristness Using the Dirichlet Distribution," in *Proc. of ICML2005*, 2005, pp. 545–552.

[7] C. Biernacki, G. Celeux and G. Govaert, "Assessing a Mixture Model for Clustering with the Integrated Complete Likelihood," *IEEE Trans. on PAMI*, vol. 22, no. 7, pp. 719–725, 2000.

[8] 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.