Spatial Finite Non-Gaussian Mixtures for Color Image Segmentation

Ali Sefidpour

A Thesis

in

The Concordia Institute

for

Information Systems Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Applied Science (Quality Systems Engineering) at

Concordia University

Montréal, Québec, Canada

August 2011

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Ali Sefidpour

Entitled: Spatial Finite Non-Gaussian Mixtures for Color Image Segmentation

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Quality Systems Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. C. Assi Chair

Dr. J. Bentahar Examiner

Dr. A. Hamou-Lhadj Examiner

Dr. N. Bouguila Supervisor

Approved by _____
Chair of Department or Graduate Program Director

_____
Dean of Faculty

Date September 7, 2011

Abstract

Spatial Finite Non-Gaussian Mixtures for Color Image Segmentation

Ali Sefidpour

Finite mixture models are one of the most widely and commonly used probabilistic techniques for image segmentation. Although the most well known and commonly used distribution when considering mixture models is the Gaussian, it is certainly not the best approximation for image segmentation and other related image processing problems. It is well known, for instance, that the statistics of natural images are not Gaussian at all. In this thesis, we propose to use finite Dirichlet mixture model (DMM), finite generalized Dirichlet mixture model (GDMM) and finite Beta-Liouville mixture model (BLMM), which offer more flexibility in data modeling, for image segmentation. A maximum likelihood (ML) based algorithm is applied for estimating the resulted segmentation model's parameters. Spatial information is also employed for figuring out the number of regions in an image and two color spaces are investigated and compared. The experimental results show that the proposed segmentation framework yields good overall performance that is better than a comparable technique based on Gaussian mixture model.

Acknowledgements

It is a great pleasure to express my utmost gratitude to my supervisor, Professor Nizar Bouguila, who has supported me throughout my research career in Concordia University, and also for his intelligent guidance, encouragement and patience for allowing me to experience in my own way.

I would like to thank all my friendly fellow lab mates who made our lab a delightful and convenient place for research, in particular many thanks goes to Taoufik and Ali that I took benefit of their discussions during my research.

It is noteworthy to mention my gratitude to the faculty members and administrative staff of CIISE department, which were always ready to help with smiling faces.

Last but not the least, my deepest gratitude goes to my family for their endless love and support throughout my life, without their support, simply I could not do this.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Introduction and Related Works

Image segmentation is one of the essential image processing techniques receiving considerable attention in various applications. The importance of segmentation is obvious by the central role it plays in a number of applications that involve image and video processing, like content based image retrieval [1, 2], remote sensing [3], medical analysis [4, 5], intelligent vehicles [6], image compression [5], and so on. The success or failure of any of these applications is considerably dependent on the segmentation algorithm used before.

Although various approaches have been proposed and adopted in the past for image segmentation, the problem is still, however, challenging and open [7–9]. Finding the number of regions automatically, defining the smooth and meaningful regions for a given application and handling the direct effect of color space choice are some of the problems in image segmentation. Among the different image segmentation methods which have been employed as a solution for the mentioned problems, statistical models [10–13] have been widely used in the past. This can be justified by the simplicity of defining and describing image features by statistical models (mixture models and hidden markov models, for instance) and the great ability of these models for data classification.

In this thesis we focus on the adoption of a particular statistical modeling approach for image segmentation namely finite mixtures. The four main problems in this specific case are: (1) the

automatic determination of the number of clusters (i.e. number of regions in a given image), (2) the integration of the spatial information to achieve smooth segmentation results, (3) the accurate choice of the probability density functions that shall describe the image regions, and (4) the estimation of the resulted segmentation mixture model's parameters. In the majority of the works that have dealt with image segmentation, the often-cited difficulty is the automatic determination of the number of regions and a lot of approaches have been proposed with relative success. For instance, a bootstrapping approach has been applied in [13]. A variety of information criteria have been considered to select automatically the number of regions (see, [14, 15], for instance, for detailed discussions), also. The most successful approaches, however, have been based on the consideration of the spatial information as an implicit prior information about the expected number of regions. Indeed, since each region is composed by similar adjacent pixels which follow almost the same color or pattern, using spatial information will naturally lead us to find out smooth and accurate segments. For instance, an adaptive clustering algorithm based on K-means and eight-neighbor Gibbs random field model has been proposed in [16] and applied to pictures of industrial objects, buildings, optical characters, faces and aerial photographs. The authors in [17, 18] have used this information to improve their segmentation results on medical images while the authors in [10, 19] have employed spatial information for segmenting brain MR images. In another interesting paper [11] a segmentation statistical model, inspired from [10], based on the integration of spatial information and Gaussian mixtures has been developed to automatically determine the number of regions and then generate smooth regions. The main problem with the majority of the approaches that have tackled the combination of mixture models and spatial information is that the image regions have been supposed to follow Gaussian distributions. This assumption is unfortunately very restrictive and unrealistic in many real-world vision problems. Although the finite Gaussian mixture is a popular model in image analysis, it is not necessarily the best solution as it is thoroughly shown in several studies (see, for instance, [20–23]).

Moreover, the influence of the chosen color spaces have not been investigated within the proposed segmentation statistical models, despite the fact that this specific choice has been shown to have a

noticeable effect on the final results in several other image processing and computer vision applications [24, 25].

Hence, in this thesis we propose the integration of spatial information into finite Dirichlet mixture model, finite generalize Dirichlet mixture model and finite Beta-Liouville mixture model for the segmentation of color images. The choice of these mixture models is motivated by their flexibility (i.e. they can be symmetric or asymmetric) in modeling data [21, 23, 26]. As in [10, 11], the spatial information is used as indirect prior knowledge, encoded via contextual constraints of neighboring pixels, for estimating the number of clusters. Finally, two appropriate color spaces are investigated within the proposed segmentation framework. Experiments show that incorporating the spatial information into generalized Dirichlet mixture models and Beta-Liouville mixture models and choosing appropriate color spaces achieves accurate image segmentation results.

1.2 Contributions

The contributions of this thesis are as follows:

☞ **Integration of spatial information into different non Gaussian finite mixture models:** Our approach suggests the integration of spatial information into three different finite mixture models (Dirichlet mixture model, generalized Dirichlet mixture model and Beta-Liouville mixture model) to produce smooth and more meaningful regions in color image segmentation while offering more flexibility and ease of use for data modeling in comparison to well known and commonly used finite Gaussian mixture models.

☞ **Automatic determination of the number of regions by using the proposed approaches:** The proposed approach also made it possible to find the number of segmentation regions automatically, this would be an extensive boost for realtime applications of color image segmentation which needs no human interference.

☞ **Revealing the effect of color space on color image segmentation:** We compare the effect of two different color spaces on the color image segmentation process and we show that different color spaces can lead to different segmentation results.

1.3 Thesis Overview

The organization of this thesis is as follows:

- ❑ Chapter 1 introduces the challenging problems in color image segmentations, the major role of finite mixture models in clustering the data and the beneficial integration of spatial information to determine the accurate number of regions and to produce the homogeneous regions.

- ❑ Chapter 2 presents three powerful finite mixture models (Dirichlet mixture model, generalize Dirichlet mixture model and Beta-Liouville mixture model), gives the details of integration of spatial information into these models and follows by estimation of the required parameters for these models.

- ❑ Chapter 3 is devoted to the presentation of our experimental results and quantitative evaluation of proposed models as compared to the Gaussian mixture model.

- ❑ Chapter 4 summarizes the various approaches and concludes the thesis while proposing the room for future works.

CHAPTER 2

Segmentation Models

2.1 Introduction

In previous chapter we presented the challenging problems related to color image segmentation and we also pointed out to finite mixture models as a particular statistical technique for color image segmentation. We continue in this chapter by introducing finite mixture models and the way we integrate spatial information into these models. Three flexible and powerful distributions will be employed to present our new models which will be learned using well known maximum likelihood estimation (MLE) within an expectation maximization (EM) optimization framework. The whole segmentation process will be summarized at the end by presenting the steps of segmentation algorithm.

2.2 Finite Dirichlet Mixture Model

Let $\mathcal{X}$ be an image represented by a set of pixels $\mathcal{X} = \{\vec{X}_1, \ldots, \vec{X}_N\}$ where each pixel is denoted by a random vector $\vec{X}_n = (X_{n1}, \ldots, X_{nD})$ [1] and N is the number of pixels. Now if the random vector $\vec{X}$ follows a Dirichlet distribution with parameters $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{D+1})$, the joint density

[1] The dimensionality D will depend on the number of features used to describe a given pixel. For instance, if we use only the color information in an RGB space, then $D = 3$.

function is given by [21]

$$p(\vec{X}|\vec{\alpha}) = \frac{\Gamma\left(|\vec{\alpha}|\right)}{\prod_{d=1}^{D+1}\Gamma\left(\alpha_d\right)}\prod_{d=1}^{D+1}X_d^{\alpha_d-1} \tag{1}$$

where $\sum_{d=1}^{D}X_d < 1$, $|\vec{X}| = \sum_{d=1}^{D}X_d (0 < X_d < 1,\ d = 1,\ldots,D)$, $X_{D+1} = 1 - |\vec{X}|$, $|\vec{\alpha}| = \sum_{d=1}^{D+1}\alpha_d\,(0 < \alpha_d, d = 1,\ldots,D+1)$. The mean, variance and covariance of Dirichlet distribution are given by [21]:

$$E(X_d) = \frac{\alpha_d}{|\vec{\alpha}|} \tag{2}$$

$$Var(X_d) = \frac{\alpha_d(|\vec{\alpha}| - \alpha_d)}{|\vec{\alpha}|^2(|\vec{\alpha}| + 1)} \tag{3}$$

$$Cov(X_i, X_j) = \frac{-\alpha_i\alpha_j}{|\vec{\alpha}|^2(|\vec{\alpha}| + 1)} \tag{4}$$

Generally, an image is composed of different regions. Thus, it is appropriate to describe it by a Dirichlet Mixture Model with M clusters

$$p(\vec{X}|\theta) = \sum_{j=1}^{M}P_j p(\vec{X}|\vec{\theta}_j) \tag{5}$$

where P_j $(0 < P_j < 1$ and $\sum_{j=1}^{M}P_j = 1)$ are the mixing proportions, $p(\vec{X}|\vec{\theta}_j)$ is the Dirichlet distribution, $\vec{\theta}_j = (\alpha_1,\ldots,\alpha_{D+1})$, and $\theta = (P_1,\ldots,P_M,\vec{\theta}_1,\ldots,\vec{\theta}_M)$ is the set of all mixture parameters.

2.2.1 Integration of Spatial Information into Mixture Models

In the following we adopt the segmentation approach, based on Gaussian mixture models, proposed in [11] for the introduction of the spatial information into finite Dirichlet mixtures. This approach can be explained as follows. For each pixel $\vec{X}_n \in \mathcal{X}$ (we don't consider the boundary pixels which number is negligible as compared to the whole image pixels), there is an immediate neighbor $\vec{\hat{X}}_n \in \mathcal{X}$ which is supposed to have arisen from the same cluster of $\vec{X}_n$, we call it the peer of $\vec{X}_n$.

Since it is supposed that the peers stay in the same clusters, this spatial information can be used as indirect information for estimating the number of clusters. In this scenario, if a larger value is assigned to M, there would be a conflict with the indirect information, provided by the pixels spatial repartition, of M, which means that a true cluster is wrongly divided into two sub-clusters. These two sub-clusters have then to be merged to form a new cluster which related parameters have to be estimated again. In this case, one of the clusters' mixing probabilities will drop suddenly and approaches zero, that can be neglected easily, so the number of clusters will gradually decrease to reach the true number of clusters (i.e. image regions).

For more clarification, let us investigate an example. Figure 2.1(a) demonstrates a sample of size 100 generated by a one-component bivariate Dirichlet mixture model with parameters $E(X_1) = 0.5$, $E(X_2) = 0.25$, $Var(X_1) = 0.05$ and $Var(X_2) = 0.0375$. The samples are shown by dots, and the ellipse is used to show the related mean and standard deviations of bivariate Dirichlet model. Then in figure 2.1(b), it is assumed that the sample dataset is wrongly partitioned into two different clusters (one is shown by dots and other is shown by stars). The parameters of new two component Dirichlet mixture model is estimated by well-known Expectation Maximization method [27]. The new parameters for each component are shown by using new ellipses while the data points belonging to each cluster are shown with different characters. Now in figure 2.1(c), since the dataset is split wrongly, most of the data points left separated from their prospective peers, thus the strong connection between the data points and their related peers in different clusters is made and shown by the lines. In figure 2.1(d), the connected peers are added to the first cluster and made a new cluster (the connected stars are labeled as dots), this sudden decrease in the cluster size shows that the two clusters are very likely the result of wrongly division and should be considered as one cluster.

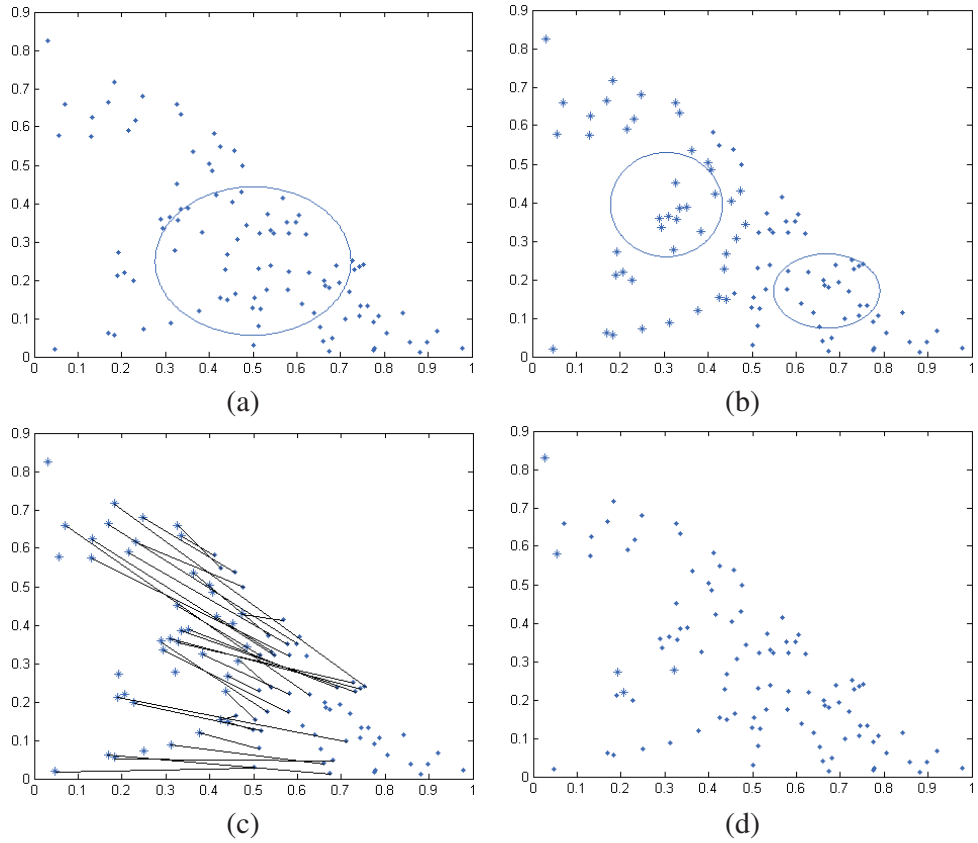


Figure 2.1: (a) Simulated data of a bivariate Dirichlet mixture. (b) Estimated results for wrongly supposed two clusters. (c) Strong relation between points and their prospective peers. (d) The combination of clusters.

2.2.2 Parameter Estimation

Now that the model is ready, the related parameters estimation would be the next step. Mixtures models parameters estimation has been studied as an interesting topic recently and many approaches have been developed [28]. Maximum likelihood (ML) approach is one of the most popular estimation methods. The main idea behind ML is to find the parameters which maximize

the joint probability density function of the available data (or the data likelihood). This can be performed through the expectation maximization (EM) algorithm [27] which is the most widely used technique in the case of missing data. The missing data in our case is the knowledge of the pixels classes. For convenience, we usually deal with the data log-likelihood, instead of the likelihood. Let $\mathcal{X}$ and the set of peers $\widehat{\mathcal{X}} = \{\vec{\widehat{X}}_1, \ldots, \vec{\widehat{X}}_N\}$ be our observed data. The set of group indicators for all pixels $\mathcal{Z} = \{\vec{Z}_1, \ldots, \vec{Z}_N\}$ will form the unobserved data, where $\vec{Z}_n = (z_{n,1}, \ldots, z_{n,M})$ denotes the missing group indicator and $z_{n,j}$ is equal to one if $\vec{X}_n$ and $\vec{\widehat{X}}_n$ belong to the same cluster j, or zero, otherwise. The complete data likelihood is given by

$$p(\mathcal{X}, \widehat{\mathcal{X}}, \mathcal{Z}|\Theta) = \prod_{n=1}^{N} \prod_{j=1}^{M} \left[P_j p(\vec{X}_n|\vec{\theta}_j) P_j p(\vec{\widehat{X}}_n|\vec{\theta}_j)\right]^{z_{n,j}} \tag{6}$$

As it is mentioned earlier for more convenience, the complete log likelihood is used as

$$L(\mathcal{X}, \widehat{\mathcal{X}}, \mathcal{Z}|\theta) = \sum_{n=1}^{N} \sum_{j=1}^{M} z_{n,j} (2 \log P_j + \log p(\vec{X}_n|\vec{\theta}_j) + \log p(\vec{\widehat{X}}_n|\vec{\theta}_j)) \tag{7}$$

Using the EM algorithm, the parameters which maximize the completed log likelihood function are estimated iteratively in 2 different steps. The Expectation (E) step and Maximization (M) step. In E-step, the conditional expectation of the complete data likelihood is calculated as

$$\begin{aligned} E[L(\mathcal{X}, \widehat{\mathcal{X}}, \mathcal{Z}|\theta)] &= Q(\mathcal{X}, \widehat{\mathcal{X}}, \theta) = \sum_{n=1}^{N} \sum_{j=1}^{M} p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\alpha}_j^{(k)}) \\ &\times (2 \log P_j + \log p(\vec{X}_n|\vec{\alpha}_j) + \log p(\vec{\widehat{X}}_n|\vec{\alpha}_j)) \end{aligned} \tag{8}$$

where $p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\alpha}_j)$ is the posterior probability which indicates the probability that $\vec{X}_n$ and $\vec{\widehat{X}}_n$ are assigned to cluster j:

$$p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\alpha}_j) = \frac{P_j p(\vec{X}_n|\vec{\alpha}_j) P_j p(\vec{\widehat{X}}_n|\vec{\alpha}_j)}{\sum_{j'=1}^{M} P_{j'} p(\vec{X}_n|\vec{\alpha}_{j'}) P_{j'} p(\vec{\widehat{X}}_n|\vec{\alpha}_{j'})} \tag{9}$$

Then, in M-step, $Q(\mathcal{X}, \widehat{\mathcal{X}}, \theta)$ (equation 8) will be maximized which give us the following for the mixing proportions:

$$P_j^{(k+1)} = \frac{1}{N} \sum_{n=1}^{N} p(j|\vec{X}_n, \vec{\widehat{X}}_n, \alpha_j^{(k)}) \tag{10}$$

A closed-form solution does not exist for the α_j parameters. Thus, we shall employ a Newton-Raphson approach:

$$\vec{\alpha}_j^{(k+1)} = \vec{\alpha}_j^{(k)} - H^{-1}(\vec{\alpha}_j^{(k)}) \times (\frac{\partial Q(\mathcal{X}, \widehat{\mathcal{X}}, \theta)}{\partial \vec{\alpha}_j}) \tag{11}$$

where H is the Hessian matrix, which can be evaluated by using the second and mixed derivatives of $Q(\mathcal{X}, \widehat{\mathcal{X}}, \theta)$. For the first derivative with respect to $\vec{\alpha}_j$ we have

$$\begin{aligned} \frac{\partial Q(\mathcal{X}, \widehat{\mathcal{X}}, \theta)}{\partial \alpha_{jd}} &= \sum_{n=1}^{N} p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\alpha}_j) \frac{\partial}{\partial \alpha_{jd}} \log p(\vec{X}_n|\vec{\alpha}_j) + \frac{\partial}{\partial \alpha_{jd}} \log p(\vec{\widehat{X}}_n|\vec{\alpha}_j) \\ &= \sum_{n=1}^{N} p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\alpha}_j) \times \left[2(\Psi(|\vec{\alpha}_j|) - \Psi(\alpha_{jd})) + \log X_{nd} + \log \widehat{X}_{nd}\right] \end{aligned} \tag{12}$$

where $\Psi(.)$ is the digamma function. The second derivative is given by

$$\frac{\partial \mathcal{Q}(\mathcal{X}, \widehat{\mathcal{X}}, \theta)}{\partial^2 \alpha_{jd}} = [2(\Psi'(|\vec{\alpha}_j|) - \Psi'(\alpha_{jd}))] \sum_{n=1}^{N} p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\alpha}_j) \tag{13}$$

and the mixed derivative is

$$\frac{\partial \mathcal{Q}(\mathcal{X}, \widehat{\mathcal{X}}, \theta)}{\partial \alpha_{jd_1} \partial \alpha_{jd_2}} = 2(\Psi'(|\vec{\alpha}_j|)) \sum_{n=1}^{N} p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\alpha}_j) \tag{14}$$

where in both equations $\Psi'(.)$ is the trigamma function. Now by using the second derivatives as the diagonal of the Hessian matrix and mixing derivatives as the other non-diagonal elements of the Hessian matrix, we can make our Hessian matrix as follow

$$H_j = 2 \sum_{n=1}^{N} P\left(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\alpha}_j\right) \begin{pmatrix} \Psi'(|\vec{\alpha}_j|) - \Psi'(\alpha_{j1}) & \cdots & \Psi'(|\vec{\alpha}_j|) \\ \vdots & \ddots & \vdots \\ \Psi'(|\vec{\alpha}_j|) & \cdots & \Psi'(|\vec{\alpha}_j|) - \Psi'(\alpha_{jD+1}) \end{pmatrix} \tag{15}$$

Then, the inverse of Hessian can be easily calculated using the approach proposed in [21]. By having the inverse of Hessian matrix we can update the new values for $\vec{\theta}$ by using equation 11.

2.2.3 Initialization and Segmentation algorithm

Parameters initialization is an important issue for mixture models parameter estimation when using the EM algorithm. Our initialization algorithm is done through well known K-means and the method of moments (MM) algorithms. According to [21], the method of moments for Dirichlet can be calculated as

$$\alpha_d = \frac{(x'_{11} - x'_{21})x'_{1d}}{x'_{21} - \left(x'_{11}\right)^2} \quad d = 1, ..., D \tag{16}$$

$$\alpha_{D+1} = \frac{(x'_{11} - x'_{21})\left(1 - \sum_{d=1}^{D} x'_{1d}\right)}{x'_{21} - \left(x'_{11}\right)^2} \tag{17}$$

$$x'_{1d} = \frac{1}{N}\sum_{n=1}^{N} x_{nd} \quad d = 1, ..., D+1 \tag{18}$$

$$x'_{21} = \frac{1}{N}\sum_{n=1}^{N} x_{n1}^2 \tag{19}$$

Thus, the proposed segmentation algorithm can be summarized as follows:

1. Choose a large initial value for M as number of image regions (this value should be larger than the expected number of the regions in the image).
2. Initialize the algorithm using K-means and method of moments.
3. Use the image data points and related peer points to update generalized Dirichlet mixture parameters by alternating the following two steps:
   - E-Step: Compute the posterior probabilities using equation 9.
   - M-Step: Update the mixture parameters using equations 10 and 11.
4. Check the mixing parameters P_j values. If a value is close to zero its related cluster should be removed and the number of clusters, M, should be reduced by one.

5. Go to 3 until convergence.

Since for each pixel (r, c) there are 4 main neighbors that are likely to be in the same region, we can use one of them as the corresponding peer of the pixel. In our experiments, following [11] we shall use the pixel $(r + 1, c)$ as the corresponding peer.

2.3 Finite Generalized Dirichlet Mixture Model

Because of the known Dirichlet distribution limitations (such as its negative covariance matrix) [23], Connor et al [29] have generalized the Dirichlet distribution as follows:

$$p(\vec{X}|\vec{\alpha}) = \prod_{d=1}^{D} \frac{\Gamma(\alpha_d + \beta_d)}{\Gamma(\alpha_d)\Gamma(\beta_d)} X_d^{\alpha_d - 1} \left(1 - \sum_{i=1}^{d} X_i\right)^{\gamma_d} \tag{20}$$

for $\sum_{d=1}^{D} X_d < 1$ and $0 < X_d < 1$ for $d = 1, \ldots, D$ where $\gamma_d = \beta_d - \alpha_{d+1} - \beta_{d+1}$ for $d = 1, \ldots, D-1$ and $\gamma_D = \beta_D - 1$. Note that the generalized Dirichlet distribution is reduced to a Dirichlet distribution when $\beta_d = \alpha_{d+1} + \beta_{d+1}$. The mean, variance and covariance of generalize Dirichlet distribution are given by [23]:

$$E(X_d) = \frac{\alpha_d}{\alpha_d + \beta_d} \prod_{i=1}^{d-1} \frac{\beta_i + 1}{\alpha_i + \beta_i} \tag{21}$$

$$Var(X_d) = E(X_d) \left(\frac{\alpha_d + 1}{\alpha_d + \beta_d + 1} \prod_{i=1}^{d-1} \frac{\beta_i + 1}{\alpha_i + \beta_i} + 1 - E(X_d) \right) \tag{22}$$

$$Cov(X_i, X_j) = E(X_j) \left(\frac{\alpha_i}{\alpha_i + \beta_i + 1} \prod_{k=1}^{i-1} \frac{\beta_k + 1}{\alpha_k + \beta_k} + 1 - E(X_i) \right) \tag{23}$$

while some other interesting properties of this distribution can be found in [30].

2.3.1 Parameter Estimation

The integration of spatial information and complete data likelihood for generalized Dirichlet model follow the same structure as Dirichlet model, so we will start this section with parameter estimation

by using EM algorithm. But before applying EM algorithm, we use an interesting property of generalized Dirichlet distribution to refine the estimates. If a vector $\vec{X}_n$ follows a generalized Dirichlet distribution, then we can construct a vector $\vec{W}_n = (W_{n1}, \ldots, W_{nD})$ using the following geometric transformation $W_{nd} = T(X_{nd})$

$$T(X_{nd}) = \begin{cases} X_{nd}, & \text{if } d = 1 \\ \frac{X_{nd}}{1-X_{n1}-\cdots-X_{nd-1}}, & \text{if } d = 2, 3, \ldots, D \end{cases} \tag{24}$$

In this vector $\vec{W}_n$, each W_{nd}, $d = 1, \ldots, D$, has a Beta distribution with parameters α_{jd} and β_{jd}, and the parameters $\{\alpha_{jd}, \beta_{jd}, d = 1, \ldots, D\}$ define the generalized Dirichlet distribution which characterizes $\vec{X}_n$ [23]. Thus, the problem of estimating the parameters of a generalized Dirichlet mixture can be reduced to estimation of the parameters of d Beta mixtures. Now to accomplish this, the new form of equation (7) will be given by

$$L(\mathcal{W}, \widehat{\mathcal{W}}, \mathcal{Z}|\theta) = \sum_{n=1}^{N}\sum_{j=1}^{M} z_{n,j}(2\log P_j + \log p_{\text{beta}}(\vec{W}_{nd}|\vec{\theta}_{jd}) + \log p_{\text{beta}}(\widehat{W}_{nd}|\vec{\theta}_{jd})) \tag{25}$$

where $\mathcal{W} = (W_{1d}, \ldots, W_{Nd})$, $0 < d < D$, $\vec{\theta}_j = (\alpha_{j1}, \beta_{j1}, \ldots, \alpha_{jD}, \beta_{jD})$, and $\theta = (P_1, \ldots, P_M,$ $\vec{\theta}_1, \ldots, \vec{\theta}_M)$ is the set of all mixture parameters.

In E-step, the conditional expectation of the complete data likelihood is calculated as

$$E[L(\mathcal{W}, \widehat{\mathcal{W}}, \mathcal{Z}|\theta)] = Q(\mathcal{W}, \widehat{\mathcal{W}}, \theta) = \sum_{n=1}^{N}\sum_{j=1}^{M} p_{\text{beta}}(j|\vec{W}_{nd}, \vec{\widehat{W}}_{nd}, \vec{\theta}_{jd}) \tag{26}$$

$$\times \quad (2\log P_j + \log p_{\text{beta}}(\vec{W}_{nd}|\vec{\theta}_{jd}) + \log p_{\text{beta}}(\vec{\widehat{W}}_{nd}|\vec{\theta}_{jd}))$$

where $p_{\text{beta}}(j|\vec{W}_{nd}, \vec{\widehat{W}}_{nd}, \vec{\theta}_{jd})$ is the posterior probability which indicates the probability that $\vec{W}_{nd}$ and $\vec{\widehat{W}}_{nd}$ are assigned to cluster j:

$$p_{\text{beta}}(j|\vec{W}_{nd}, \vec{\widehat{W}}_{nd}, \vec{\theta}_{jd}) = \frac{P_j p_{\text{beta}}(\vec{W}_{nd}|\vec{\theta}_{jd}) P_j p_{\text{beta}}(\vec{\widehat{W}}_{nd}|\vec{\theta}_{jd})}{\sum_{j'=1}^{M} P_{j'} p_{\text{beta}}(\vec{W}_{nd}|\vec{\theta}_{j'd}) P_{j'} p_{\text{beta}}(\vec{\widehat{W}}_{nd}|\vec{\theta}_{j'd})} \tag{27}$$

Then, in M-step, $Q(\mathcal{W}, \widehat{\mathcal{W}}, \theta)$ (equation 26) will be maximized which give us the following for the mixing proportions:

$$P_j^{(k+1)} = \frac{1}{N} \sum_{n=1}^{N} p_{\text{beta}}(j|\vec{W}_{nd}, \vec{\widehat{W}}_{nd}, \theta_{jd}^{(k)}) \tag{28}$$

A closed-form solution does not exist for the θ_{jd} parameters. Thus, we shall employ a Newton-Raphson approach:

$$\vec{\theta}_{jd}^{(k+1)} = \vec{\theta}_{jd}^{(k)} - H^{-1}(\vec{\theta}_{jd}^{(k)}) \times (\frac{\partial Q(\mathcal{W}, \widehat{\mathcal{W}}, \theta)}{\partial \vec{\theta}_{jd}}) \tag{29}$$

where H is the Hessian matrix, which can be evaluated by using the second and mixed derivatives of $Q(\mathcal{W}, \widehat{\mathcal{W}}, \theta)$. For the first derivative with respect to $\vec{\alpha}_{jd}$ and $\vec{\beta}_{jd}$ we have

$$\begin{aligned} \frac{\partial Q(\mathcal{W}, \widehat{\mathcal{W}}, \theta)}{\partial \alpha_{jd}} &= \sum_{n=1}^{N} p_{\text{beta}}(j|\vec{W}_{nd}, \vec{\widehat{W}}_{nd}, \vec{\theta}_{jd}) \frac{\partial}{\partial \alpha_{jd}} \log p_{\text{beta}}(\vec{W}_{nd}|\vec{\theta}_{jd}) + \frac{\partial}{\partial \alpha_{jd}} \log p_{\text{beta}}(\vec{\widehat{W}}_{nd}|\vec{\theta}_{jd}) \\ = \quad & \sum_{n=1}^{N} p_{\text{beta}}(j|\vec{W}_{nd}, \vec{\widehat{W}}_{nd}, \vec{\theta}_{jd}) \times \left[2(\Psi(\alpha_{jd} + \beta_{jd}) - \Psi(\alpha_{jd})) + \log W_{nd} + \log \widehat{W}_{nd}\right] \end{aligned} \tag{30}$$

$$\begin{aligned} \frac{\partial Q(\mathcal{W}, \widehat{\mathcal{W}}, \theta)}{\partial \beta_{jd}} &= \sum_{n=1}^{N} p_{\text{beta}}(j|\vec{W}_{nd}, \vec{\widehat{W}}_{nd}, \vec{\theta}_{jd}) \frac{\partial}{\partial \beta_{jd}} \log p_{\text{beta}}(\vec{W}_{nd}|\vec{\theta}_{jd}) + \frac{\partial}{\partial \beta_{jd}} \log p_{\text{beta}}(\vec{\widehat{W}}_{nd}|\vec{\theta}_{jd}) \\ = \quad & \sum_{n=1}^{N} p_{\text{beta}}(j|\vec{W}_{nd}, \vec{\widehat{W}}_{nd}, \vec{\theta}_{jd}) \times \left[2(\Psi(\alpha_{jd} + \beta_{jd}) - \Psi(\beta_{jd})) + \log(1 - W_{nd}) + \log(1 - \widehat{W}_{nd})\right] \end{aligned} \tag{31}$$

where $\Psi(.)$ is the digamma function. The second derivatives are given by

$$\frac{\partial \mathcal{Q}(\mathcal{W}, \widehat{\mathcal{W}}, \theta)}{\partial^2 \alpha_{jd}} = [2(\Psi'(\alpha_{jd} + \beta_{jd}) - \Psi'(\alpha_{jd}))] \sum_{n=1}^{N} p_{\text{beta}}(j|\vec{W}_{nd}, \vec{\widehat{W}}_{nd}, \vec{\theta}_{jd}) \tag{32}$$

$$\frac{\partial \mathcal{Q}(\mathcal{W}, \widehat{\mathcal{W}}, \theta)}{\partial^2 \beta_{jd}} = [2(\Psi'(\alpha_{jd} + \beta_{jd}) - \Psi'(\beta_{jd}))] \sum_{n=1}^{N} p_{\text{beta}}(j|\vec{W}_{nd}, \vec{\widehat{W}}_{nd}, \vec{\theta}_{jd}) \tag{33}$$

and the mixed derivative is

$$\frac{\partial \mathcal{Q}(\mathcal{W}, \widehat{\mathcal{W}}, \theta)}{\partial \alpha_{jd} \partial \beta_{jd}} = 2(\Psi'(\alpha_{jd} + \beta_{jd})) \sum_{n=1}^{N} p_{\text{beta}}(j|\vec{W}_{nd}, \vec{\widehat{W}}_{nd}, \vec{\theta}_{jd}) \tag{34}$$

where in both equations $\Psi'(.)$ is the trigamma function. Then, the inverse of Hessian can be easily calculated using the approach proposed in [21]. By having the inverse of Hessian matrix we can update the new values for $\vec{\theta}$ by using equation 29.

2.3.2 Initialization and Segmentation algorithm

Parameters initialization is an important issue for mixture models parameter estimation when using the EM algorithm. Our initialization algorithm is done through well known K-means and the method of moments (MM) algorithms [23]. Thus, the proposed segmentation algorithm can be summarized as follows:

1. Choose a large initial value for M as number of image regions (this value should be larger than the expected number of the regions in the image).
2. Initialize the algorithm using the approach in [23].
3. Use the image data points and related peer points to update generalized Dirichlet mixture parameters by alternating the following two steps:
   - E-Step: Compute the posterior probabilities using equation 27.
   - M-Step: Update the mixture parameters using equations 28 and 29.
4. Check the mixing parameters P_j values. If a value is close to zero its related cluster should be removed and the number of clusters, M, should be reduced by one.
5. Go to 3 until convergence.

2.4 Finite Beta-Liouville Mixture Model

Although generalized Dirichlet distribution can overcome the disadvantages of Dirichlet distribution (for instance, the covariance matrix is not restricted to be negative anymore) [23], it involves a large number of parameters (It has $2D$ parameters in dimension D). Another good choice for random vector $\vec{X}_n = (X_{n1}, \ldots, X_{nD})$ would be Liouville distribution of second kind, which has $D + 2$ parameters in dimension D.

If random vector $\vec{X}$ follows a Liouville distribution of second kind with positive parameters $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_D)$, with density function $f(.)$, then [31]

$$p(\vec{X}|\vec{\alpha}) = \frac{\Gamma(\sum_{d=1}^{D} \alpha_d)}{u^{\sum_{d=1}^{D} \alpha_d - 1}} f(u) \prod_{d=1}^{D} \frac{X_d^{\alpha_d - 1}}{\Gamma(\alpha_d)} \tag{35}$$

where $u = \sum_{d=1}^{D} X_d < 1$,$0 < X_d$,$d = 1, \ldots, D$. A flexible univariate distribution that can be a suitable choice as our density function is Beta distribution [26] which has positive parameters α and β

$$f(u|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha - 1}(1 - u)^{\beta - 1} \tag{36}$$

Replacing the equation 36 into the equation 35 gives us the following:

$$p\left(\vec{X}|\alpha_1, \alpha_2, \ldots, \alpha_D, \alpha, \beta\right) = \frac{\Gamma\left(\sum_{d}^{D} \alpha_d\right) \Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \prod_{d=1}^{D} \frac{X_d^{\alpha_d - 1}}{\Gamma(\alpha_d)} \left(\sum_{d=1}^{D} X_d\right)^{\alpha - \sum_{d=1}^{D} \alpha_d} \left(1 - \sum_{d=1}^{D} X_d\right)^{\beta - 1} \tag{37}$$

which has positive parameters $\theta = (\alpha_1, \alpha_2, \ldots, \alpha_D, \alpha, \beta)$ and is called Beta-Liouville distribution [31]. It is noteworthy to mention that Beta-Liouville distribution of second kind can be reduced to Dirichlet distribution with parameters $\alpha_1, ..., \alpha_{D+1}$ by choosing $\sum_{d=1}^{D} \alpha_d$ and α_{D+1} as parameters of Beta component of the distribution. The mean and variance of the Beta-Liouville distribution of second kind can be calculated as

$$E(X_d) = \frac{\alpha}{\alpha + \beta} \frac{\alpha_d}{\sum_{d=1}^{D} \alpha_d} \tag{38}$$

Table 2.1: Parameters and number of parameters for Multivariate Gaussian Distribution (MGD), Dirichlet Distribution (DD), Generalized Dirichlet Distribution (GDD) and Beta-Liouville Distribution (BLD).

	MGD	DD	GDD	BLD
Parameters	$\mu_1, ..., \mu_D$ Σ	$\alpha_1, ..., \alpha_{D+1}$	$\alpha_1, ..., \alpha_D$ $\beta_1, ..., \beta_D$	$\alpha_1, ..., \alpha_D$ α, β
Number of Parameters	$D + \frac{D(D+1)}{2}$	$D + 1$	$2D$	$D + 2$

$$Var\left(X_d\right) = E\left(X_d\right) \frac{\alpha + 1}{\alpha + \beta + 1} \frac{\alpha_d + 1}{\sum_{d=1}^{D} \alpha_d + 1} - \frac{\alpha^2}{\left(\alpha + \beta\right)^2} \frac{\alpha_d^4}{\left(\sum_{d=1}^{D} \alpha_d\right)^4} \tag{39}$$

Beta-Liouville distribution of second kind like generalized Dirichlet distribution is not restricted to have negative covariance matrix [26] and has fewer parameters for estimation, than the generalized Dirichlet distribution. More details about the parameters and number of parameters of discussed distributions in this thesis (Parameters and number of parameters for Multivariate Gaussian Distribution (MGD), Dirichlet Distribution (DD), Generalized Dirichlet Distribution (GDD) and Beta-Liouville Distribution (BLD) can be found in table 2.1.

2.4.1 Parameter Estimation

For this new model, the integration of spatial information and the log likelihood of complete data are like the Dirichlet model, so now we start by using of EM algorithm directly on equation 7 for our new distribution and estimate the related parameters. So in E-step we have

$$E[L(\mathcal{X}, \widehat{\mathcal{X}}, \mathcal{Z}|\theta)] = Q(\mathcal{X}, \widehat{\mathcal{X}}, \theta) = \sum_{n=1}^{N} \sum_{j=1}^{M} p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\theta}_j) \times (2 \log P_j + \log p(\vec{X}_n|\vec{\theta}_j) + \log p(\vec{\widehat{X}}_n|\vec{\theta}_j)) \tag{40}$$

with posterior

$$p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\theta}_j) = \frac{P_j p(\vec{X}_n|\vec{\theta}_j) P_j p(\vec{\widehat{X}}_n|\vec{\theta}_j)}{\sum_{j'=1}^{M} P_{j'} p(\vec{X}_n|\vec{\theta}_{j'}) P_{j'} p(\vec{\widehat{X}}_n|\vec{\theta}_{j'})} \tag{41}$$

Then by maximizing equation 40, the new mixing proportions can be derived as

$$P_j^{(k+1)} = \frac{1}{N}\sum_{n=1}^{N} p(j|\vec{X}_n, \vec{\widehat{X}}_n, \theta_j^{(k)}) \tag{42}$$

The new values of model parameters can be estimated by Newton-Raphson approach:

$$\vec{\theta}_j^{(k+1)} = \vec{\theta}_j^{(k)} - H^{-1}(\vec{\theta}_j^{(k)}) \times (\frac{\partial Q(\mathcal{X}, \widehat{\mathcal{X}}, \theta)}{\partial \vec{\theta}_j}) \tag{43}$$

For evaluating the Hessian matrix, the first derivative with respect to α_j, β_j and α_{jd} would be

$$\begin{aligned}\frac{\partial Q(\mathcal{X}, \widehat{\mathcal{X}}, \theta)}{\partial \alpha_j} &= \sum_{n=1}^{N} p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\theta}_j)\frac{\partial}{\partial \alpha_j}\log p(\vec{X}_n|\vec{\theta}_j) + \frac{\partial}{\partial \alpha_j}\log p(\vec{\widehat{X}}_n|\vec{\theta}_j) \\ &= \sum_{n=1}^{N} p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\theta}_j)\left[2\left(\Psi(\alpha_j+\beta_j) - \Psi(\alpha_j)\right) + \log\sum_{d=1}^{D} X_{nd} + \log\sum_{d=1}^{D}\widehat{X}_{nd}\right]\end{aligned} \tag{44}$$

$$\begin{aligned}\frac{\partial Q(\mathcal{X}, \widehat{\mathcal{X}}, \theta)}{\partial \beta_j} &= \sum_{n=1}^{N} p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\theta}_j)\frac{\partial}{\partial \beta_{jd}}\log p(\vec{X}_n|\vec{\theta}_j) + \frac{\partial}{\partial \beta_{jd}}\log p(\vec{\widehat{X}}_n|\vec{\theta}_j) \\ &= \sum_{n=1}^{N} p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\theta}_j)\left[2\left(\Psi(\alpha_j+\beta_j) - \Psi(\beta_j)\right) + \log\left(1 - \sum_{d=1}^{D} X_{nd}\right)\right. \\ &\quad \left. + \log\left(1 - \sum_{d=1}^{D}\widehat{X}_{nd}\right)\right]\end{aligned} \tag{45}$$

$$\begin{aligned}&\frac{\partial Q(\mathcal{X}, \widehat{\mathcal{X}}, \theta)}{\partial \alpha_{jd}} = \sum_{n=1}^{N} p\left(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\theta}_j\right)\frac{\partial}{\partial \alpha_{jd}}\log p\left(\vec{X}_n|\vec{\theta}_j\right) + \frac{\partial}{\partial \alpha_{jd}}\log p\left(\vec{\widehat{X}}_n|\vec{\theta}_j\right) \\ &= \sum_{n=1}^{N} p\left(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\theta}_j\right)\left[2\left(\Psi\left(\sum_{d=1}^{D}\alpha_{jd}\right)\right) + \left(\log X_{nd} - \Psi\left(\alpha_{jd}\right) - \log\sum_{d=1}^{D} X_{nd}\right)\right. \\ &\left. + \left(\log\widehat{X}_{nd} - \Psi\left(\alpha_{jd}\right) - \log\sum_{d=1}^{D}\widehat{X}_{nd}\right)\right]\end{aligned} \tag{46}$$

The second and mixed derivatives are given by

$$\frac{\partial Q(\mathcal{X}, \widehat{\mathcal{X}}, \theta)}{\partial^2 \alpha_j} = \left[2\left(\Psi^{'}(\alpha_j+\beta_j) - \Psi^{'}(\alpha_j)\right)\right]\sum_{n=1}^{N} p(j|\vec{X}_n, \vec{\widehat{X}}_n, \vec{\theta}_j) \tag{47}$$

$$\frac{\partial Q(\mathcal{X},\widehat{\mathcal{X}},\theta)}{\partial^2\beta_j}=\left[2\left(\Psi^{'}(\alpha_j+\beta_j)-\Psi^{'}(\beta_j)\right)\right]\sum_{n=1}^{N}p(j|\vec{X}_n,\vec{\widehat{X}}_n,\vec{\theta}_j) \tag{48}$$

$$\frac{\partial Q\left(\mathcal{X},\widehat{\mathcal{X}},\theta\right)}{\partial\alpha_{jd_1}\partial\alpha_{jd_2}}=\begin{cases}\left[2\left(\Psi^{'}\left(\sum_{d=1}^{D}\alpha_{jd}\right)\right)-2\left(\Psi^{'}\left(\alpha_{jd}\right)\right)\right]\sum_{n=1}^{N}p\left(j|\vec{X}_n,\vec{\widehat{X}}_n,\vec{\theta}_j\right), & \text{if } \alpha_{jd_1}=\alpha_{jd_2}\\ \left[2\left(\Psi^{'}\left(\sum_{d=1}^{D}\alpha_{jd}\right)\right)\right]\sum_{n=1}^{N}p\left(j|\vec{X}_n,\vec{\widehat{X}}_n,\vec{\theta}_j\right), & \text{otherwise}\end{cases} \tag{49}$$

$$\frac{\partial Q(\mathcal{X},\widehat{\mathcal{X}},\theta)}{\partial\alpha_j\partial\beta_j}=\frac{\partial Q(\mathcal{X},\widehat{\mathcal{X}},\theta)}{\partial\beta_j\partial\alpha_j}=\left[2\left(\Psi^{'}(\alpha_j+\beta_j)\right)\right]\sum_{n=1}^{N}p(j|\vec{X}_n,\vec{\widehat{X}}_n,\vec{\theta}_j) \tag{50}$$

$$\frac{\partial Q(\mathcal{X},\widehat{\mathcal{X}},\theta)}{\partial\alpha_j\partial\alpha_{jd}}=\frac{\partial Q(\mathcal{X},\widehat{\mathcal{X}},\theta)}{\partial\beta_j\partial\alpha_{jd}}=\frac{\partial Q(\mathcal{X},\widehat{\mathcal{X}},\theta)}{\partial\alpha_{jd}\partial\beta_j}=\frac{\partial Q(\mathcal{X},\widehat{\mathcal{X}},\theta)}{\partial\alpha_{jd}\partial\beta_j}=0 \tag{51}$$

We can show that the Hessian has a block diagonal matrix format

$$H\left(\theta_j\right)=blockdiag\left(H_a,H_b\right)=\begin{bmatrix}H_a & 0\\ 0 & H_b\end{bmatrix} \tag{52}$$

where

$$H_a=H\left(\alpha_j,\beta_j\right)=\begin{bmatrix}\frac{\partial Q(\mathcal{X},\hat{\mathcal{X}},\theta)}{\partial^2\alpha_j} & \frac{\partial Q(\mathcal{X},\hat{\mathcal{X}},\theta)}{\partial\alpha_j\partial\beta_j}\\ \frac{\partial Q(\mathcal{X},\hat{\mathcal{X}},\theta)}{\partial\beta_j\partial\alpha_j} & \frac{\partial Q(\mathcal{X},\hat{\mathcal{X}},\theta)}{\partial^2\beta_j}\end{bmatrix} \tag{53}$$

$$H_b=H\left(\alpha_{j1},\ldots,\alpha_{jD}\right)=\frac{\partial Q\left(\mathcal{X},\hat{\mathcal{X}},\theta\right)}{\partial\alpha_{jd_i}\partial\alpha_{jd_k}},\text{where } i,k\in\{1,\ldots,D\} \tag{54}$$

Then, the inverse of $H\left(\theta_j\right)$ can be easily computed as

$$H^{(-1)}(\theta_j)=(blockdiag(H_a,H_b))^{(-1)}=blockdiag\left((H_a)^{(-1)},(H_b)^{(-1)}\right) \tag{55}$$

By having the inverse of Hessian, we can update the new values for $\vec{\theta}_j$ by using equation 43

2.4.2 Initialization and Segmentation algorithm

For initialization phase, the k-means and method of moments are used again, with considering the fact, that Beta-Liouville distribution of second kind can be reduced to Dirichlet distribution with parameters $\alpha_1,...,\alpha_{D+1}$ by using $\sum_{d=1}^{D}\alpha_d$ and α_{D+1} as the parameters of the Beta component of

the distribution. Thus, the same method of Dirichlet can be used for initialization of Beta-Liouville model.

Then, the proposed segmentation algorithm can be summarized as follows:

1. Choose a large initial value for M as number of image regions (this value should be larger than the expected number of the regions in the image).
2. Initialize the algorithm using the approach in [21].
3. Use the image data points and related peer points to update Beta-Liouville mixture parameters by alternating the following two steps:
   - E-Step: Compute the posterior probabilities using equation 41.
   - M-Step: Update the mixture parameters using equations 42 and 43.
4. Check the mixing parameters P_j values. If a value is close to zero its related cluster should be removed and the number of clusters, M, should be reduced by one.
5. Go to 3 until convergence.

CHAPTER 3

Experimental Results

3.1 Introduction

In this chapter we validate our models, and we also investigate how the choice of distinctive color spaces can make difference in color image segmentation.

3.2 Design of Experiments

The main goal of this section is to investigate the performance of the proposed approaches (Dirichlet , generalized Dirichlet and Beta-Liouville mixtures) as compared to the one developed in [11] which has been based on the integration of the spatial information into Gaussian mixture models. It is noteworthy that an important problem when dealing with color images is the choice of the color space. In the case of image segmentation, it is highly desirable that the chosen color space be robust against varying illumination, concise, discriminatory and robust to noise. Some of such color spaces have been analyzed, evaluated and discussed in [24]. Among these spaces, we have the RGB normalized color space which rgb planes are defined by [24, 32]

$$r(R,G,B) = \frac{R}{R+G+B} \tag{1}$$

$$g(R,G,B) = \frac{G}{R+G+B} \tag{2}$$

$$b(R,G,B) = \frac{B}{R+G+B} \tag{3}$$

and the $l_1l_2l_3$ color space defined by [24]

$$l_1(R,G,B) = \frac{(R-G)^2}{(R-G)^2+(R-B)^2+(G-B)^2} \tag{4}$$

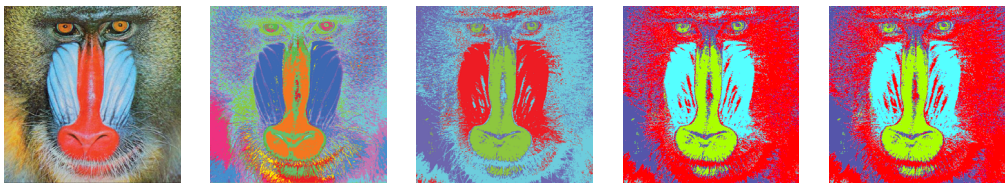
$$l_2(R,G,B) = \frac{(R-B)^2}{(R-G)^2+(R-B)^2+(G-B)^2} \tag{5}$$

$$l_3(R,G,B) = \frac{(G-B)^2}{(R-G)^2+(R-B)^2+(G-B)^2} \tag{6}$$

which is a photometric color invariant for matte and shiny spaces [24]. The rgb and $l_1l_2l_3$ have been shown to outperform the widely used RGB space [24] and thus will be considered in our experiments. To have a fair comparison, in all cases (i.e. Gaussian, Dirichlet, generalized Dirichlet and Beta-Liouville mixtures), the initial values for the number of clusters M are set to 30.

3.3 Experiment 1

In addition to the famous Baboon image, which is widely used to evaluate image segmentation algorithms, we have employed 300 images from the well-known publicly available Berkeley segmentation data set [33]. This database is composed of a variety of natural color images generally used as a reliable way to compare image segmentation algorithms. Figure 3.1 shows a comparison between the segmentation results obtained by our approach and the technique developed in [11] when we consider the rgb color space. Figure 3.1(a) shows the original Baboon image, while figures 3.1(b), 3.1(c), 3.1(d) and 3.1(e) show the results obtained with the Gaussian mixture, the Dirichlet mixture, generalized Dirichlet mixture and Beta-Liouville mixture, respectively. The algorithm in [11] selected 12 regions for the baboon image while our proposed algorithms considered 4 regions for all Dirichlet, generalized Dirichlet and Beta-Liouville mixture models. As the figure indicates, in addition to the less number of regions preferred by our algorithms, the regions provided are more meaningful. In this image the nose of baboon is almost composed of two clear



(a) (b) (c) (d) (e)

Figure 3.1: Baboon image segmentation in the rgb color space. (a) Original image, (b) Segmentation using the Gaussian mixture ($M = 12$), (c) Segmentation using the Dirichlet mixture ($M = 4$), (d) Segmentation using the Generalized Dirichlet mixture ($M = 4$), (e) Segmentation using the Beta-Liouville mixture ($M = 4$).

regions while the hair is divided to light and dark regions.

The images in figure 3.2 are chosen from the Berkeley database. The estimated number of regions, in these images when considering the rgb color space, selected by all the Gaussian, the Dirichlet, the generalized Dirichlet and the Beta-Liouville mixtures are mentioned in the related images sub-captions. According to this figure we can see clearly that the proposed segmentation algorithms generate both quantitatively (less and better regions) and qualitatively (more meaningful) better results when compared to the Gaussian mixture approach.

To allow a principled comparison between image segmentation approaches and measure the differences between them, the authors in [34] have proposed the Normalized Probabilistic Rand (NPR) index as a neutral scale for quantitative comparison between image segmentation algorithms. The NPR index has a value up to 1 where a higher value shows better segmentation results. The calculation of the NPR index requests the availability of a hand-labeled segmentation used as a ground truth to score the segmentation algorithm. Unfortunately there are not many databases which have the ground truth information, but Berkeley database has provided at least 5 ground truth segmentation results for all its 300 natural public images.The NPR index can be calculated as follow

$$\text{NPR Index} = \frac{\text{PR Index} - \text{Expected Index}}{\text{Maximum Index} - \text{Expected Index}} \tag{7}$$

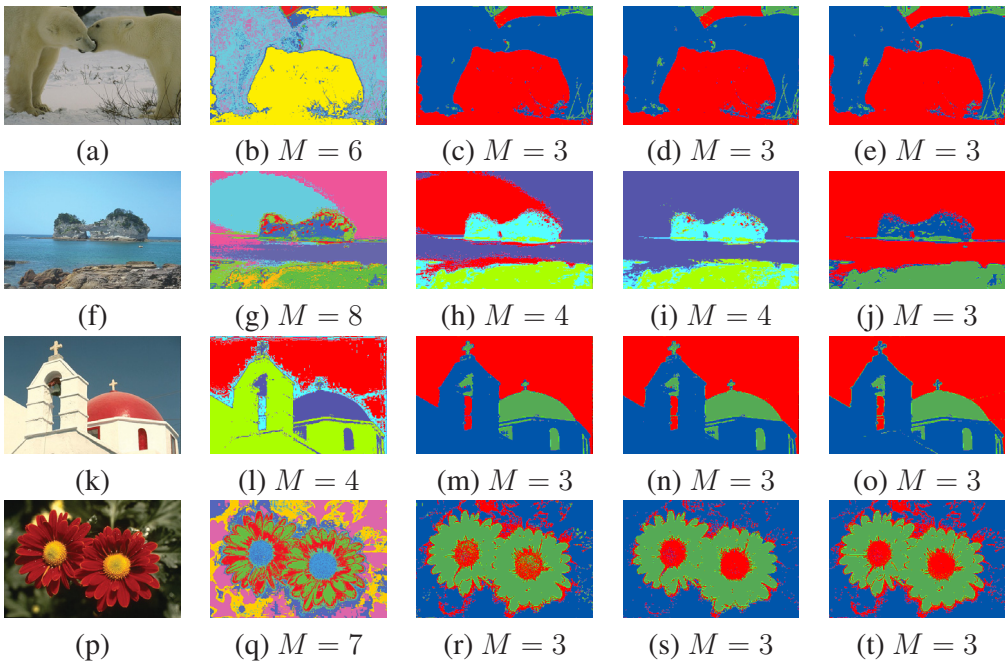


Figure 3.2: Examples of images segmentation results in the rgb color space. (a,f,k,p) Original images from the Berkeley Database. (b,g,l,q) Segmentation results using the Gaussian mixture model. (c,h,m,r) Segmentation results using the Dirichlet mixture model. (d,i,n,s) Segmentation results using the generalized Dirichlet mixture model. (e,j,o,t) Segmentation results using the Beta-Liouville mixture model.

where the Probabilistic Rand (PR) index is defined as

$$\text{PR}(S_{test}, \{S_k\}) = \frac{1}{\binom{N}{2}} \sum_{\substack{i,j \\ i<j}} \left[p_{ij}^{c_{ij}} (1 - p_{ij})^{1-c_{ij}} \right] \tag{8}$$

which has a value between 0 and 1. In equation 8, S_{test} defines the segmentation which has to be compared with the ground truth segmentations, S_K is set of K hand-labeled ground truth segmentations, N defines the number of pixels in the image, p_{ij} shows how likely the unordered pair of pixels (i, j) are in the same segment (Based on ground truth segmentations) and c_{ij} is the event that (i, j) pixels are in the same segment in the test image (So $c_{ij} \in \{0, 1\}$). In other words equation 8 shows that the sets of all segmentations will follow a Bernoulli distribution over the

Table 3.1: NPR index sample mean for Gaussian mixture model (GMM), Dirichlet mixture model (DMM), generalized Dirichlet mixture model (GDMM) and Beta-Liouville mixture model (BLMM) in rgb color space.

	GMM	DMM	GDMM	BLMM
NPR Index Sample Mean	0.2667	0.5376	0.5523	0.5595

number of pairs.

The expected value of PR index can also be computed as

$$\mathrm{E}\left[\mathrm{PR}(S_{test}, \{S_k\})\right] = \frac{1}{\binom{N}{2}} \sum_{\substack{i,j \\ i<j}} \left[{p'}_{ij} p_{ij} \left(1 - {p'}_{ij}\right) \left(1 - p_{ij}\right) \right] \tag{9}$$

where ${p'}_{ij}$ can be translated as the weight proportion of (i, j) unordered pairs for all images in the database (More details can be found in [34]).

Because of the "expensive" calculation of NPR Index [34], we have calculated the NPR index for a reasonable number of images. NPR sample mean of each model for rgb color space is given in table 3.1. It is clear that all proposed algorithms improved the NPR index enormously, and also there is a difference between the NPR index sample mean of Dirichlet model and two other generalization models (generalize Dirichlet and Beta-Liouville). Figure 3.3 shows some of the original images used to calculate the NPR index, the Gaussian mixture's segmentation results, and our segmentation results. It shows also five ground truth segmentations, from Berkeley database, for each selected image. The NPR index for each algorithm is mentioned in the images sub-captions. Figure 3.4 illustrates the effect of choosing the $l_1 l_2 l_3$ color space on the segmentation of the Baboon image. According to this image, the new color space has improved the segmentation result in the case of the Gaussian mixture by decreasing the number of regions to 10 as compared to the 12 regions found when the rgb color space has been considered. Changing the color space has not affected the numbers of regions in the case of the Dirichlet, generalized Dirichlet and Beta-Liouville mixtures and the result is still better than the Gaussian. For the Dirichlet, generalized Dirichlet and Beta-Liouville models the change of color space made the image smoother. Figure 3.5 displays

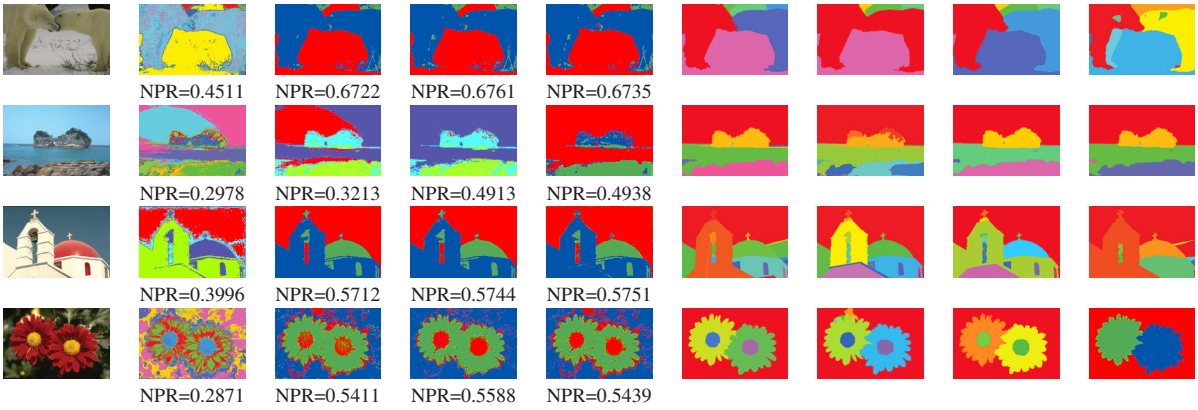


Figure 3.3: Examples of images used to calculate the NPR index of each segmentation approach in the rgb color space. First column contains the original images. Second column contains the segmentation results using the Gaussian mixture model. Third column contains the segmentation results using the Dirichlet mixture model. Forth column contains segmentation results using generalized Dirichlet model. Fifth column contains segmentation results using Beta-Liouville model. Columns 6, 7, 8 and 9 contains the ground truth segmentations.

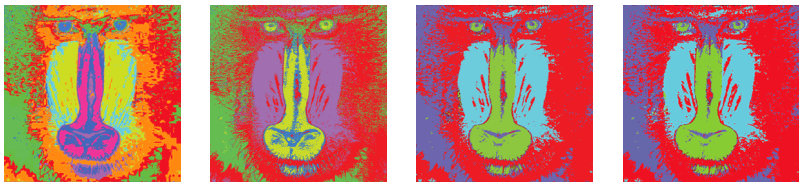


Figure 3.4: Baboon image segmentation in the $l_1l_2l_3$ color space. (a) Segmentation using the Gaussian mixture ($M = 10$), (b) Segmentation using the Dirichlet mixture ($M = 4$), (c) Segmentation using the generalized Dirichlet mixture ($M = 4$), (d) Segmentation using the Beta-Liouville mixture ($M = 4$).

the segmentation results in the $l_1l_2l_3$ color space when considering the images from the Berkeley database for the Gaussian, the Dirichlet, the generalized Dirichlet and Beta-Liouville mixtures. The NPR index values are mentioned in the sub-captions. As we can see from this figure, choosing the $l_1l_2l_3$ space provides generally smoother and more meaningful regions. This is actually clear

by comparing the NPR indexes in figures 3.5 and 3.3.

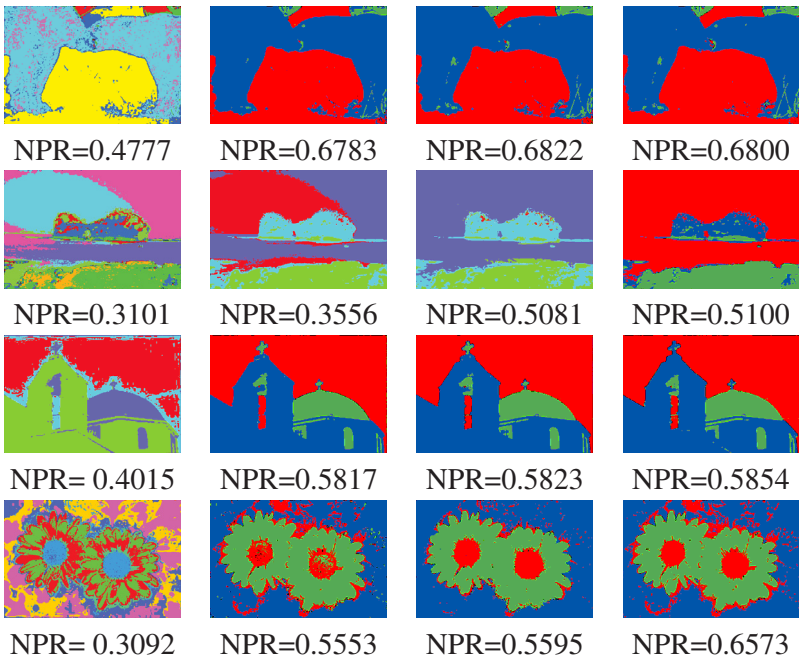


Figure 3.5: Images segmentation in the $l_1l_2l_3$ color space. First column: segmentation using the Gaussian mixture model. Second column: segmentation using the Dirichlet mixture. Third column: segmentation using the generalized Dirichlet mixture. Forth column: segmentation using the Beta-Liouville mixture.

The NPR index sample mean of all models for $l_1l_2l_3$ color space is shown table 3.2. Comparing the results in table 3.1 and table 3.2 indicate a segmentation improvement where the $l_1l_2l_3$ color space is considered.

3.4 Experiment 2

In this experiment, we mainly focus on the effect of distinctive color spaces on color image segmentation while we can still compare the effectiveness of the proposed approaches comparing to

Table 3.2: NPR index sample mean for Gaussian mixture model (GMM), Dirichlet mixture model (DMM), generalized Dirichlet mixture model (GDMM) and Beta-Liouville mixture model (BLMM) in $l_1l_2l_3$ color space.

	GMM	DMM	GDMM	BLMM
NPR Index Sample Mean	0.2755	0.5512	0.5732	0.5803

Gaussian based model subjectively. For this reason, each model is evaluated within rgb and $l_1l_2l_3$ color spaces and then for each mixture model, the proper color space will be chosen by using different well-known metrics.

To evaluate our approaches with more and different images and not to be restricted to a specific database, considerable number of images have been chosen from "Urban and Natural Scene Categories" of MIT Computational Visual Cognition Laboratory [35], this database has eight different categories (for instance, forests, highways, coasts and beaches and ...) and there exists few hundreds images in each category.

Authors in [36] have proposed four criteria for investigating the effectiveness of image segmentation methods (intra region uniformity, inter region disparity, simplicity in regions and simplicity in boundaries). Note that the last criterion may not be held for segmentation of natural images. Thus, the four best metrics (Q, V_{CP}, Zeb and F_{RC}) of [37] are considered to choose the appropriate color space within each mixture model. The Q metric can be calculated as

$$\frac{\sqrt{N}}{1000S_I}\sum_{j=1}^{N}\left[\frac{e_j^2}{1+\log S_j}+\left(\frac{N(S_j)}{S_j}\right)^2\right] \tag{10}$$

where N is the number of segments, S_I is the image number of pixels and e_j^2 is the square color error of region j. The V_{CP} metric is given by

$$\sqrt{\frac{1}{N}\sum_j\sum_k sobel_j^2-\left(\frac{1}{N}\sum_j\sum_k sobel_j\right)^2} \tag{11}$$

Table 3.3: Color space selection percentage by different metrics for Gaussian mixture model (GMM), Dirichlet mixture model (DMM), generalized Dirichlet mixture model (GDMM) and Beta-Liouville mixture model (BLMM).

	GMM		DMM		GDMM		BLMM	
	rgb	$l_1l_2l_3$	rgb	$l_1l_2l_3$	rgb	$l_1l_2l_3$	rgb	$l_1l_2l_3$
Q	47.0%	53.0%	27.7%	72.3%	24.9%	75.1%	26.4%	73.6%
V_{CP}	36.8%	63.2%	23.7%	76.3%	26.1%	73.9%	23.6%	76.4%
Zeb	22.7%	77.3%	20.1%	79.9%	13.0%	87.0%	13.5%	86.5%
F_{RC}	18.5%	81.5%	5.9%	94.1%	7.8%	92.2%	7.0%	93.0%

where in other words, the standard deviation of the Sobel coefficients of region j is used here as a metric. The Zeb metric is as follow

$$\frac{1}{S_j}\sum_{s\in R_j}\max\left\{contrast(s,t), t\in W(s)\cap R_j\right\} \tag{12}$$

where $W(s)$ is the neighbor of pixel s,R_j is the j-th region and the $contrast(s,t)$ calculate the contrast between the two pixels s and t. The last metric is F_{RC} where is define as

$$\frac{1}{N}\sum_{j=1}^{N}\frac{S_j}{S_I}e^2(R_j) \tag{13}$$

Table 3.3 demonstrates the selection percentage for each color space within the mixture models. The results indicate that the choice of $l_1l_2l_3$ color space will lead us to smoother and more meaningful regions while the disparity between the distinct regions are reserved as well. It is interesting to point out that the selection rate of $l_1l_2l_3$ color space is higher among Dirichlet based mixture models. As a subjective comparison, figure 3.6 shows some original images used to calculate the selected metrics, the Gaussian results, the Dirichlet results, the generalized Dirichlet results and Beta-Liouville results for both rgb and $l_1l_2l_3$ color spaces. The advantage of $l_1l_2l_3$ over rgb color space and also the supremacy of Dirichlet based models over Gaussian based model can be settled easily.

Since dealing with natural images is always a challenging problem for image segmentation algorithms because of the environmental noises, we have evaluated our models with more examples

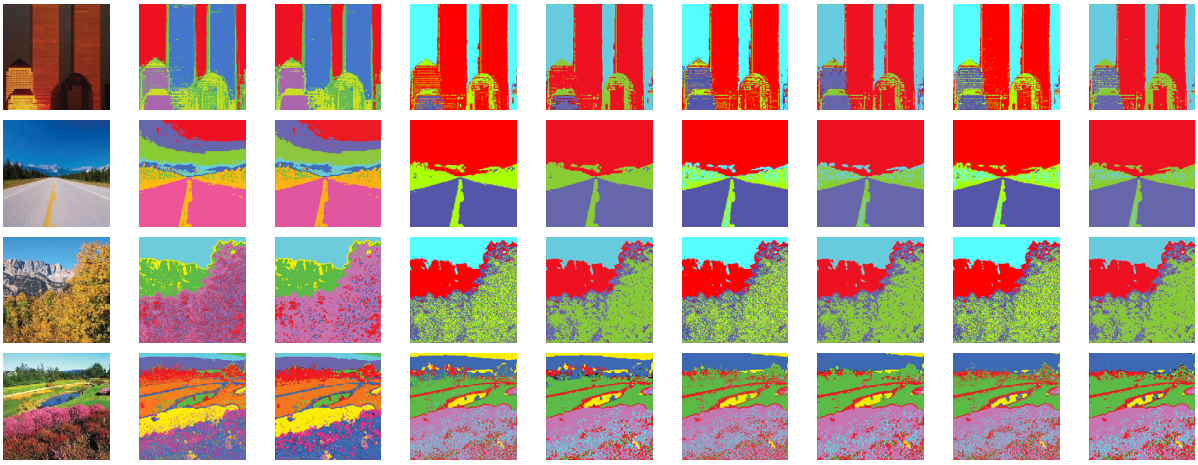


Figure 3.6: Images segmentation in each set for rgb and $l_1l_2l_3$ color spaces for all 4 mixture models. Column 1: original image. Columns 2 and 3: segmentation using the Gaussian mixture model. Columns 4 and 5: segmentation using the Dirichlet mixture. Columns 6 and 7: segmentation using the generalized Dirichlet mixture. Columns 8 and 9: segmentation using the Beta-Liouville mixture.

from SkyFlash [38] database. This database contains plenty of images of animals and military equipments which which segmentation is not an easy task. Some samples form this database in addition to the results of image segmentation for each color space within the distinct mixture model is shown in figure 3.7.

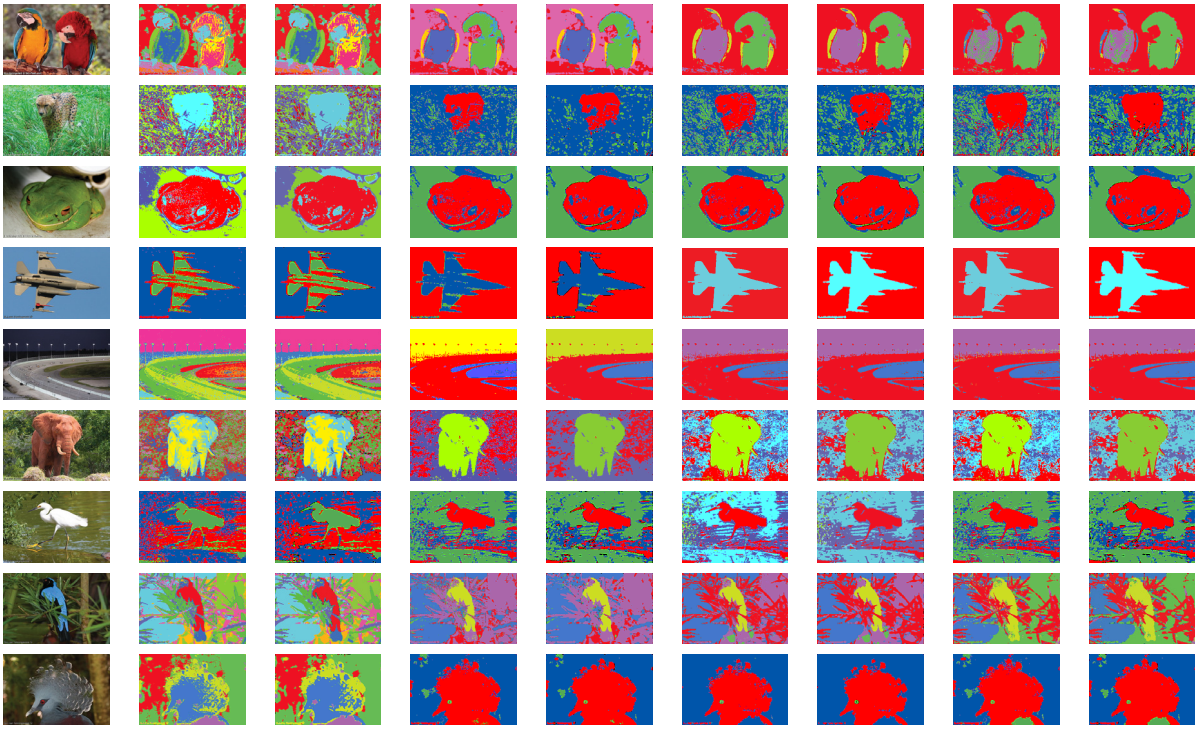


Figure 3.7: Images segmentation in each set for rgb and $l_1l_2l_3$ color spaces for all 4 mixture models. Column 1: original image. Columns 2 and 3: segmentation using the Gaussian mixture model. Columns 4 and 5: segmentation using the Dirichlet mixture. Columns 6 and 7: segmentation using the generalized Dirichlet mixture. Columns 8 and 9: segmentation using the Beta-Liouville mixture.

CHAPTER 4

Conclusions

In this thesis, we have presented different algorithms for color image segmentation by integrating spatial information into finite mixture models. The selection of these mixture models is motivated by their flexibility in approximation of data points in different shapes in contrast to the well known gaussian mixture model which always keeps the symmetric bell shape. First we chose Dirichlet mixture model for its flexibility in data modeling and its few number of parameters for estimation, but its restrictive covariance matrix was the negative point. This disadvantage has been handled by generalized Dirichlet mixture model in cost of an increase in the number of parameters. So finally the motivation for choosing the Beta-Liouville mixture model was its flexibility in shapes and its few number of parameters as compared to Gaussian and Generalized Dirichlet distributions. Then the spatial information is included in this model by considering pixels neighborhoods and by using this information as a prior knowledge in our model, we got the ability to estimate the number of segmentation regions automatically. The resulted image segmentation statistical model has been learned using maximum likelihood estimation within an expectation maximization framework. Results, which have concerned the segmentation of an important number of images from the well-known Berkeley images database, show that the proposed algorithms perform better than an approach which has been based on finite Gaussian mixture models. The effect of distinct color spaces on color image segmentation was also investigated by using different metrics over the famous MIT images database. Future works can be devoted to the application of the developed

segmentation algorithm for object detection and recognition and also a promising extension of this work would be on the integration of more visual features to improve further the segmentation results. Video segmentation could be considered as another interesting application which has to be done in an online fashion. So a potential future work could be the extension of the proposed approach to segment frames in a real-time stream.

List of References

[1] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1026 – 1038, 2002.

[2] M. Ozden and E. Polat. A Color Image Segmentation Approach for Content-Based Image Retrieval. *Pattern Recognition*, 40(4):1318–1325, 2007.

[3] D. Ziou, N. Bouguila, M. S. Allili, A. El Zaart. Finite Gamma Mixture Modeling Using Minimum Message Length Inference: Application to SAR Image Analysis. *International Journal of Remote Sensing*, 30(3):771–792, 2009.

[4] W. M. Wells, W. E. L. Grimson, R. Kikinis, F. A. Jolesz. Adaptive Segmentation of MRI Data. *IEEE Transactions on Medical Imaging*, 15(4):429–442, 1996.

[5] Liang Shen and R.M. Rangayyan. A segmentation-based lossless image coding method for high-resolution medical image compression. *IEEE Transactions on Medical Imaging*, 16(3):301–307, 1997.

[6] T. Bucher, C. Curio, J. Edelbrunner, C. Igel, D. Kastrup, I. Leefken, G. Lorenz, A. Steinhage, and W. von Seelen. Image processing and behavior planning for intelligent vehicles. *IEEE Transactions on Industrial Electronics*, 50(1):62–75, 2003.

[7] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888 –905, 2000.

[8] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.

[9] Nikhil R Pal and Sankar K Pal. A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277 – 1294, 1993.

[10] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45 –57, 2001.

[11] X. Yang and Sh. M. Krishnan. Image Segmentation using Finite Mixtures and Spatial Information. *Image and Vision Computing*, 22(9):735–745, 2004.

[12] C. Nikou, N. P. Galatsanos, and A. C. Likas. A Class-Adaptive Spatially Variant Mixture Model for Image Segmentation. *IEEE Transactions on Image Processing*, 16:1121–1130, 2007.

[13] W. L. Hung, M. Sh. Yang, and D. H. Chen. Bootstrapping approach to feature-weight selection in fuzzy c-means algorithms with an application in color image segmentation. *Pattern Recognition Letters*, 29(9):1317 – 1325, 2008.

[14] N. Bouguila and D. Ziou. Unsupervised Selection of a Finite Dirichlet Mixture Model: An MML-Based Approach. *IEEE Transactions on Knowledge and Data Engineering*, 18(8):993–1009, 2006.

[15] N. Bouguila and D.Ziou. On Fitting Finite Dirichlet Mixture Using ECM and MML. In S. Singh, M. Singh, C. Apté and P. Perner, editor, *Pattern Recognition and Data Mining, Third*

International Conference on Advances in Pattern Recognition, ICAPR (1), pages 172–182. Springer, LNCS 3686, 2005.

[16] T. N. Pappas. An adaptive clustering algorithm for image segmentation. *IEEE Transactions on Signal Processing*, 40(4):901–914, 1992.

[17] J. Luo, C. W. Chen and K. J. Parker. On the Application of Gibbs Random Field in Image Processing: From Segmentation to Enhancement. *Journal of Electronic Imaging*, 4(2):187–198, 1995.

[18] K. Sh. Chuang, H. L. Tzeng, Sh. Chen, J. Wu, and T. J. Chen. Fuzzy c-means clustering with spatial information for image segmentation. *Computerized Medical Imaging and Graphics*, 30(1):9–15, 2006.

[19] K. Held, E. R. Kops, B. J. Krause, W. M. Welles III, R. Kikinis, H-W. Müller-Gärtner. Markov Random Field Segmentation of Brain MR Images. *IEEE Transactions on Medical Imaging*, 16(6):878–886, 1997.

[20] N. Bouguila, D. Ziou and J. Vaillancourt. Novel Mixtures Based on the Dirichlet Distribution: Application to Data and Image Classification. In *Machine Learning and Data Mining in Pattern Recognition (MLDM)*, pages 172–181. Springer, LNAI 2734, 2003.

[21] 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*, 13(11):1533–1543, 2004.

[22] N. Bouguila and D. Ziou. Using unsupervised learning of a finite Dirichlet mixture model to improve pattern recognition applications. *Pattern Recognition Letters*, 26(12):1916–1925, 2005.

[23] 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*, 15(9):2657–2668, 2006.

[24] T. Gevers, A. W. M. Smeulders. Color-Based Object Recognition. *Pattern Recognition*, 32(3):453–464, 1999.

[25] N. Bouguila and D. Ziou. A Probabilistic Approach for Shadows Modeling and Detection. In *Proc. of the IEEE International Conference on Image Processing (ICIP)*, pages 329–332, 2005.

[26] N. Bouguila. Bayesian hybrid generative discriminative learning based on finite liouville mixture models. *Pattern Recognition*, 44(6):1183 – 1200, 2011.

[27] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. New York: Wiley-Interscience, 1997.

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

[29] R. J. Connor and J. E. Mosimann. Concepts of independence for proportions with a generalization of the dirichlet distribution. *Journal of the American Statistical Association*, 64(325):194–206, 1969.

[30] T. T. Wong. Generalized dirichlet distribution in bayesian analysis. *Applied Mathematics and Computation*, 97(2-3):165 – 181, 1998.

[31] K.T. Fang, S. Kotz, and K.W. Ng. *Symmetric multivariate and related distributions*. Monographs on statistics and applied probability. Chapman and Hall, 1990.

[32] N. Bouguila and D. Ziou. Dirichlet-Based Probability Model Applied to Human Skin Detection. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 521–524, 2004.

[33] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision Proc. 8th IEEE Int'l Conf.*, volume 2, pages 416–423, 2001.

[34] R. Unnikrishnan, C. Pantofaru, and M. Hebert. Toward objective evaluation of image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:929–944, 2007.

[35] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42:145–175, 2001. 10.1023/A:1011139631724.

[36] R. M. Haralick and L. G. Shapiro. Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, 29(1):100 – 132, 1985.

[37] H. Zhang, J. E. Fritts, and S. A. Goldman. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110(2):260 – 280, 2008.

[38] http://www.sky-flash.com/.