A Mixture Model of Circular-Linear Distributions for Color Image Segmentation

Anandarup Roy
CVPR Unit
Indian Statistical Institute
203 B. T. Road, Kolkata 700108

Swapan K. Parui
CVPR Unit
Indian Statistical Institute
203 B. T. Road, Kolkata 700108

Utpal Roy
Dept. of Computer & System Sc.
Visva-Bharati University
Santiniketan- 731235, India

ABSTRACT

This article deals with mixture model based color image segmentation in the LCH color space. In this space, one of the components (representing hue in particular) is circular in nature. Hence LCH image pixels are samples on a cylinder. A statistical model for such data needs to employ circular-linear joint distributions. Here such a model is designed using the "Independent von-Mises Gaussian" distribution. Further its mixture is used to approximate the distribution of the LCH data. The mixture parameters are estimated using standard EM algorithm. Comprehensive experiments are conducted on Berkeley segmentation data set to measure the performance of the algorithm in terms of a variety of quantitative indices for image segmentation. A comparison is further made with some existing mixture models. Our study reveals that the proposed mixture model performs satisfactorily in this regard.

General Terms:

Statistical model, Pattern recognition

Keywords:

Finite mixture model, Circular-linear distribution, Color image segmentationifx

1. INTRODUCTION

Color based image segmentation [1] is becoming increasingly important in many applications since color images are now easily available and can provide more information than gray level images. In this article, a clustering based segmentation approach is designed. The basic idea behind this approach is to directly cluster the pixels in a certain color space by employing some clustering algorithms. A classical technique for clustering is the K-Means algorithm [1]. Clustering can also be viewed as a hidden variable problem. Segmenting an image into clusters involves determining which clusters generate the image pixels, which is the hidden information. In this framework, a mixture model based approach is suitable for segmentation. This approach assumes a mixture distribution approximating the distribution of image pixels. The mixture distribution parameters can be estimated using standard expectation maximization (EM) algorithm. The EM algorithm utilizes the hidden variable to make an estimate. To this direction an attempt was taken by Carson et al. [2] by modeling the joint distribution of color and texture, with Gaussian Mixture Model (GMM). The image was represented in CIE-LUV space. GMM can be applied to the RGB color space also. The GMM assumes symmetric distribution of R, G and B color bands. However, these three bands may be skewed in either directions. Roy et al. [3] considered skewed distributions of the R, G, and B color bands, and used a mixture of Beta distributions to model the image data. The Beta distribution can approximate a skewed

distribution, thus can provide a more appropriate model for color bands. We here consider LCH color space instead of RGB space. The emphasis of this article is to obtain a suitable mixture distribution for hue, chroma and lightness data. The LCH model is a mixture of angular (hue) and linear (chroma and lightness) data. It is important to take into account these mixed characteristics. The angular data can be represented by a random variable $\Theta \in [0, 2\pi)$. A direct modeling of Θ can be avoided by transforming it into the representation $(\cos\theta, \sin\theta)$. Unlike Θ, the quantities $\cos\theta$ and $\sin\theta$ themselves are not circular. Hence a linear model e.g. GMM can be fit on the pair $(\cos\theta, \sin\theta)$. This approach has some disadvantages. First, the dimension of data is increased by one. So, it has a negative effect on computational complexity. A more serious drawback is the correlation between the pair $(\cos\theta, \sin\theta)$. This pair together represents a point on the circumference of a circle, and may be highly correlated. This may pose the problem of singularity of the covariance matrix for some mixture components. Therefore, we use the von-Mises distribution for Θ. The von-Mises [4] distribution is defined on an unit circle and analogous to univariate Gaussian distribution in $\Re^2$. Several attempts have been taken to estimate the parameters of the mixture of von-Mises distributions. Mooney et al. [5] used a mixture of two circular von-Mises distributions and estimated the parameters using a quasi-Newton procedure. Banerjee et al [6] estimated the parameters of von-Mises Fisher distribution which is a generalization of von-Mises distribution. The von-Mises distribution also has application in image processing. Ludtke et al. [7] used a mixture of von-Mises distributions to model the orientation of the contour.

In this study, the hue, chroma and lightness are assumed independent random variables. The hue information is modeled using von-Mises distribution. To model the linear variables (i.e., the chroma and the lightness) we use Gaussian and Beta distributions as two alternatives. The joint density can be obtained by the product of the two (von-Mises and Gaussian or von-Mises and Beta) distributions. A mixture of this joint distribution approximate the LCH distribution over the image pixels. An EM algorithm is designed to estimate the mixture parameters. The clustering is done on the basis of the maximum a posteriori probability.

The paper is organized as follows. First a brief overview on directional statistics is presented in Section 2. Later, in Section 3 we describe the joint distribution for the circular-linear data and construct the mixture model (IvMMM). The parameter estimation of this mixture model is considered in Section 4. This section also includes the clustering evaluation strategies. The details about applying IvMMM to color segmentation problem is discussed in Section 5. The post-processing to merge boundary pixels is also described here. Further, the color segmentation results are presented in Section 6. We use the Berkeley segmentation data set (Section 6.1) for the application. Here, the IvMMM is compared with two existing mixture model approaches for
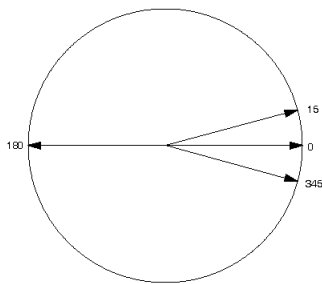
and distance measure are defined as follows.

$$\vec{\mu}^{cl} \;=\; \left(\overline{\Theta}, \overline{\vec{X}}\right)^T \qquad (5)$$

$$\text{and } d(\vec{p}_i, \vec{p}_j) \;=\; 1 - \cos(\theta_i - \theta_j) + \|\vec{x}_i - \vec{x}_j\|. \qquad (6)$$

Here the standard deviation of each $\vec{x}_k$ is assumed to be unity.

3. CIRCULAR-LINEAR JOINT DISTRIBUTION: DESCRIPTION AND MIXTURE

Let us consider a circular-linear random vector $\vec{p} = (\theta, x_1, \ldots, x_d)^T$, i.e., $\vec{p} \in [0, 2\pi) \times \Re^d$. Let us start by defining the marginal distributions of Θ and $\vec{X}$ prior to specify the joint distribution.

The von-Mises distribution is used here for the circular random variable Θ. The pdf of von-Mises distribution is given as follows.

$$f^{(v)}(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\theta - \mu)) \qquad (7)$$

Here, $I_0(\kappa)$ is the modified bessel function of the first kind of order zero and argument κ. The mean $\mu \in [0, 2\pi)$ and concentration parameter $\kappa \in \Re^+ \cup \{0\}$. When $\kappa = 0$, the density gives an uniform distribution on $[0, 2\pi)$. We use a polynomial approximation to $I_0(\kappa)$ discussed by Mardia and Jupp [4].

The linear variable ($\vec{X}$) here is described using two choices. Let us assume Gaussian distributions for $\vec{X}$. Let x be an element of $\vec{X}$. Its pdf is then given as:

$$g^{(g)}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\nu)^2}{2\sigma^2}\right). \qquad (8)$$

With an usual sense, ν and σ are the mean and standard deviation of $g^{(g)}(x)$.

One limitation of the Gaussian distribution is its symmetry. In general, the concerned data may not form a symmetric distribution. Then the Gaussian is not an appropriate choice to approximate the data distribution. To add asymmetry, let us now assume $\vec{X}$ follows a Beta distribution instead of a Gaussian. Then the pdf has the following form.

$$g^{(b)}(x) = \frac{1}{B(\alpha, \beta)} x^{1-\alpha}(1-x)^{1-\beta}, \qquad (9)$$

where $B(\alpha, \beta)$ is the Beta function. The Beta distribution may become symmetric or asymmetric depending upon the choice of the parameters α and β. In this context it is more powerful than the Gaussian, to represent underlaying data distribution. On the other hand, the Gaussian pdf is simpler than Beta distribution pdf in terms of computation. The Beta function $B(\alpha, \beta)$ needs approximation to evaluate. Nevertheless, the Beta margins provide a good alternative design as our results indicate.

Considering the two alternative margins for the linear variables, the joint distributions are defined as follows.

IvMG: The independent von-Mises Gaussian joint distribution uses Gaussian margins for $\vec{X}$. It has the following pdf

$$f_{IG}(\vec{p}) = f^{(v)}(\theta) \prod_{i=1}^{d} g_i^{(g)}(x_i). \qquad (10)$$

IvMB: The independent von-Mises Beta joint distribution assumes Beta margins for $\vec{X}$ instead of Gaussian. So its pdf becomes

$$f_{IB}(\vec{p}) = f^{(v)}(\theta) \prod_{i=1}^{d} g_i^{(b)}(x_i). \qquad (11)$$

To make it brief, further we use the term "IvM" to denote both IvMG and IvMB distributions. For the rest of this article, f_I denotes IvM distribution.

Fig. 1. The arithmetic mean points to the wrong way.

color segmentation. The findings are described in Section 6.2. Our results show an improvement over the considered two existing mixture models.

2. ON DIRECTIONAL STATISTICS

Directional data may be visualized as points on the surface of a hypersphere, or in two dimensions, on the circumference of a circle. A circular random variable Θ has domain $[0, 2\pi)$ and can be represented as a point on an unit circle. The problem that arises with the directional data is that the linear statistics may not be directly applicable to directional variables. Consider the *arithmetic mean* for a circular variable. Let us have two samples $15°$ and $345°$ for θ. These samples are displayed in Fig. 1 (with East as zero direction and anti-clockwise as positive motion). Measured in a standard way, the mean of these two samples becomes $180°$ directed West, whereas the two samples point towards East. From this example it is clear that the usual approach to computing the arithmetic mean cannot be applied to directional data. An appropriate measure of the mean direction can be obtained by treating the data as unit vectors and using the direction of their resultant vector. Suppose, we have $\theta_1, \ldots, \theta_n$ circular observations. Consider the rectangular transformation for each observation θ_i, i.e., the pair $(\cos\theta_i, \sin\theta_i)$. Then the resultant vector of these unit vectors is given by:

$$\mathbf{R} = \left(\sum_{i=1}^{n} \cos\theta_i, \sum_{i=1}^{n} \sin\theta_i\right) = (C, S), \text{ say.} \qquad (1)$$

The direction of this resultant vector, i.e., the mean direction is denoted by $\overline{\Theta}$ and defined as:

$$\overline{\Theta} = \arg\left(\sum_{i=1}^{n} \cos\theta_i + i\sum_{i=1}^{n} \sin\theta_i\right) \qquad (2)$$

$$= \arctan\left(\frac{S}{C}\right). \qquad (3)$$

Note, since $0 \le \theta < 2\pi$, we should employ a quadrant-specific inverse of tangent so that $\overline{\Theta} \in [0, 2\pi)$.

Besides the circular mean we may be interested in measuring the *circular distance*. A reasonable measure of circular distance between two points A and B with angles θ_1 and θ_2 is given as follows.

$$d(A, B) = 1 - \cos(\theta_1 - \theta_2). \qquad (4)$$

If α is the angle between A and B, then clearly $d(A, B)$ is monotonically increasing from 0 (when $\alpha = 0$) to 2 (when $\alpha = \pi$).

After having the definition of circular mean and distance, they may be extended to cope the circular-linear case. This could be done by the insertion of linear measures of statistics along with circular measures. Let $\vec{p}$ be the vector $(\theta, x_1, \ldots, x_d)^T$, i.e., $\vec{p} \in [0, 2\pi) \times \Re^d$ (hence $\vec{X}$ is linear while Θ is circular). When we have n such vectors $\vec{p}_1, \ldots, \vec{p}_n$, the circular-linear mean ($\vec{\mu}^{cl}$)

3.1 Mixture model of the joint distributions

Let us now define the mixture of K circular-linear (i.e., IvM) distributions. The pdf of the mixture is given by:

$$f(\vec{p}|\vec{\Xi}) = \sum_{h=1}^{K} P_h f_I(\vec{p}|h, \vec{\Xi}_h), \qquad (12)$$

where P_h ($0 \le P_h \le 1$ and $\sum_{h=1}^{K} P_h = 1$) is the hth component mixing proportion (prior probability), $f_I(\vec{p}|h, \vec{\Xi}_h)$ is the IvM distribution representing the hth component of the mixture and $\vec{\Xi}_h$ is the set of parameters of the hth component. The symbol $\vec{\Xi} = (\vec{\Xi}_1, \ldots, \vec{\Xi}_K, P_1, \ldots, P_K)$ refers to the entire set of parameters of the mixture distribution to be estimated. This mixture model is further referred to as "IvMMM". To make it specific we use the term "IvMGMM" and "IvMBMM" to denote the mixture distributions involving IvMG and IvMB distributions respectively.

4. MAXIMUM LIKELIHOOD ESTIMATION

Let $\vec{\aleph} = \vec{p}_1, \ldots, \vec{p}_N$ be a finite set of N samples drawn independently from IvMMM. Here $\vec{p}_i$ is a circular-linear data. The standard method to fit finite mixture model to data is the EM algorithm which converges to maximum likelihood estimate of the mixture parameters. The EM algorithm is based on the assumption that $\vec{\aleph}$ is incomplete data. The missing information is the label associated with each $\vec{p}_i$ and indicates which mixture component produces that $\vec{p}_i$. From data clustering viewpoint, these labels are simply the cluster information associated with each data point. The standard EM settings express the distribution of these missing variables as follows.

$$q(h|\vec{p}_i) = \frac{P_h f_I(\vec{p}_i|h, \vec{\Xi}_h)}{\sum_{l=1}^{K} P_l f_I(\vec{p}_i|l, \vec{\Xi}_l)}. \qquad (13)$$

With the inclusion of the missing variables, the log likelihood function of the mixture distribution becomes:

$$\Phi(\vec{\aleph}, \vec{\Xi}, \lambda) = \sum_{i=1}^{N} \sum_{h=1}^{K} \left[\ln \left(P_h f_I(\vec{p}_i|h, \vec{\Xi}_h) \right) \right] q(h|\vec{p}_i) +$$
$$\lambda \left(1 - \sum_{h=1}^{K} P_h \right), \qquad (14)$$

where λ is the lagrange multiplier.

To maximize $\Phi(\vec{\aleph}, \vec{\Xi}, \lambda)$ one may go independently for a priori probabilities P_h and the parameters $\vec{\Xi}_h$. The a priori probabilities can be found out with:

$$P_h = \frac{1}{N} \sum_{i=1}^{N} q(h|\vec{p}_i). \qquad (15)$$

After differentiating $\Phi(\vec{\aleph}, \vec{\Xi}, \lambda)$ with respect to μ_h, the expression for μ_h is obtained as:

$$\mu_h = tan^{-1} \left(\frac{\sum_{i=1}^{N} q(h|\vec{p}_i) \sin \theta_i}{\sum_{i=1}^{N} q(h|\vec{p}_i) \cos \theta_i} \right). \qquad (16)$$

The expression of κ_h is however given in terms of the ratio of modified bessel functions:

$$A_2(\kappa_h) = \frac{I_1(\kappa_h)}{I_0(\kappa_h)} = \frac{\sum_{i=1}^{N} \cos(\mu_h - \theta_i) q(h|\vec{p}_i)}{\sum_{i=1}^{N} q(h|\vec{p}_i)}. \qquad (17)$$

Since $A_2(\kappa_h)$ involves ratio of Bessel functions, it is not possible to get an analytical solution. $A_2(\kappa_h)$ is a non-decreasing function. Thus one may obtain κ by applying Newton-Raphson. The numerical methods, however, often causes a problem of overflow. Therefore, an asymptotic approximation of $A_2(\kappa)$ is more suitable for estimating κ. Such approaches also take constant computation time unlike any iterative numerical method. In this paper, an approximation, proposed by Banerjee et al [6], is used to find κ. Thus κ can be estimated by:

$$\kappa = \frac{2\bar{r} - \bar{r}^3}{1 - \bar{r}^2},$$

where $A_2(\kappa) = \bar{r}$. Other approximations of κ are discussed by Mardia and Jupp [4].

Now consider the linear variable $\vec{X}$. Let x be an element of $\vec{X}$. We assume Gaussian distribution for x. Then the estimates for ν_h and σ_h are then given by:

$$\nu_h = \frac{1}{\sum_{i=1}^{N} q(h|\vec{p}_i)} \sum_{i=1}^{N} q(h|\vec{p}_i) x_i, \qquad (18)$$

$$\sigma_h^2 = \frac{1}{\sum_{i=1}^{N} q(h|\vec{p}_i)} \sum_{i=1}^{N} q(h|\vec{p}_i)(x_i - \nu_h)^2. \qquad (19)$$

Besides Gaussian, Beta distribution is also considered for x. The method of moments can be used to estimate the Beta distribution parameters α and β. Given the estimated moments ν_h and σ_h, the parameters α_h and β_h are estimated by the following expressions.

$$\alpha_h = \nu_h \left(\frac{\nu_h(1 - \nu_h)}{\sigma_h} - 1 \right), \qquad (20)$$

$$\beta_h = (1 - \nu_h) \left(\frac{\nu_h(1 - \nu_h)}{\sigma_h} - 1 \right). \qquad (21)$$

According to EM algorithm, the E-Step involves the computation of $q(h|\vec{p}_i)$ using equation 13. The M-Step, on the other hand, maximizes the complete log-likelihood $\Phi(\vec{\aleph}, \vec{\Xi}, \lambda)$ by estimating the mixture parameters. These two steps iterate untill the log-likelihood stabilizes.

After EM, the different components indicate different clusters in the data cloud. Let these clusters be denoted by $C_1, C_2, \ldots, C_K$. After estimating all the parameters of the mixture, the cluster assignment of a data point $\vec{p}_i$ is performed using the following *maximum-a-posteriori* (MAP) condition.

$$\vec{p}_i \in C_j \text{ if } q(j|\vec{p}_i) > q(h|\vec{p}_i) \ \forall h \ne j. \qquad (22)$$

4.1 Initialization of the EM

The K-Means algorithm is applied with $(\Theta, \vec{X})$ values to obtain an initial clustering of the data. This K-Means procedure uses the circular-linear version of mean and distance (i.e. Eq. 5 and 6). The number of mixture components is same as the number of clusters. Further, to initialize the model we proceed as follows. Let the hth cluster have N_h data points. Then the parameters of the corresponding von-Mises marginal distribution (i.e. μ_h and

κ_h) are estimated with:

$$\tan \mu_h = \frac{\sum_{i=1}^{N_h} \sin \theta_i}{\sum_{i=1}^{N_h} \cos \theta_i} \qquad (23)$$

$$A_2(\kappa_h) = \frac{1}{N_h} \sum_{i=1}^{N_h} \cos(\theta_i - \mu_h). \qquad (24)$$

Next, let us consider the linear variable $\vec{X}$. Assuming Gaussian margin, the parameters are obtained using the following equations.

$$\nu_h = \frac{1}{N_h} \sum_{i=1}^{N_h} x_i, \ \sigma_h^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (x_i - \nu_h)^2. \qquad (25)$$

Considering Beta margin, given ν_h and σ_h^2 (Eq. 25), α_h and β_h are estimated using Eq. 20 and 21.

4.2 Estimating the Number of Mixture Components

Several model selection methods have been proposed to estimate the possible number of mixture components. These methods can be broadly categorized into two classes: deterministic and stochastic. The deterministic methods start with a set of candidate models for minimum to maximum number of components. The optimum number of components is then selected from this set by minimizing certain model selection criterion. We here use the BIC type approximation of integrated classification likelihood (ICL-BIC) [8] criterion for model selection. A detailed review of ICL-BIC and other criteria is presented by Mclachlan and Peel [9] where ICL-BIC seems to outperform the others.

4.3 Evaluation of Segmentation

An obvious concern in clustering is to qualify the clustering algorithm. The data set we used here is an image data set for segmentation purpose. This data set accompanies a set of ground-truths suggesting different orderings of cluster labels. In this paper, four metrics are used to quantify the consistency between an image segmentation and its ground-truths. These metrics are: the probabilistic Rand index proposed by Unnikrishnan et al. [10], the global consistency error [11], the variation of information [12] and the boundary displacement error [13]. Let's take a brief overview of these measures.

(1) The probabilistic Rand index (PRI) is a classical metric that measures the probability that an arbitrary pair of samples have consistent labels in the two partitions. The PRI metric is in the range $[0, 1]$, with higher values indicating greater similarity between two partitions.

(2) The Global Consistency Error (GCE) measures the extent to which one segmentation can be viewed as a refinement of the other. Segmentations which are related in this manner are considered to be consistent, since they could represent the same natural image segmented at different scales. A lower value for GCE implies better performance.

(3) The variation of information (VoI) measures the sum of information loss and information gain between the two clusterings, and thus it roughly measures the extent to which one clustering can explain the other. The VoI metric is nonnegative, with lower values indicating greater similarity.

(4) The boundary displacement error (BDE) measures the average displacement error of boundary pixels between two segmented images. Particularly, it defines the error as the distance between the pixel in one boundary image and the closest pixel in the other boundary image.

The GCE and BDE, by definition, penalize under-segmentation more heavily than over-segmentation. In particular, GCE does not at all penalize over-segmentation, i.e., the best score is achieved by assigning each pixel as an individual segment. But then the VoI score can be poor. As Yang et al. [14] experienced, the PRI and the VoI are supposed to be good measures for human perception.

5. COLOR IMAGE SEGMENTATION

Let us assume that the LCH image pixels are generated from IvMMM (equation 12). Here, since we have only color channels as features, the dimensionality of the circular-linear vector $\vec{p}$ is three. The EM method estimates the mixture parameters by iterating expectation and maximization steps. As an extreme condition we must look at the gray portions inside an image. The gray portions can be represented using only the lightness component of LCH image. If we consider the LCH color cylinder, we find the lightness component varies parallel to z-axis (perpendicular to hue axis). Thus the hue becomes undefined for gray portions. Since IvMMM assumes the presence of hue, it could no longer be applicable. Here, we separate out the gray portions before employing EM. A separate clustering (say, univariate GMM) may be employed with the gray portions. However, for the present image data set, we encounter only a few gray pixels and thus ignore them to minimize time complexity.

After EM, a single spectral cluster, when mapped to an image, may contain several spatially connected components. It is observed that between two such adjacent components from two different clusters, there is a thin boundary consisting of pixels with spectral values that are different from both these clusters. In fact, a pixel on such a boundary really comes from any one the two adjacent components. These thin boundaries do not correspond to any particular physical entities in the original image. Hence, these boundary pixels may be merged with any of the adjacent clusters without harm. This issue is addressed by applying a spatial similarity based smoothing proposed in the following subsection.

5.1 Smoothing Towards Merging of Boundary Pixels

We observed that if there is a thin boundary between two adjacent components, its thickness varies from one pixel to three pixels. So, the 5×5 (or larger) neighborhood around a pixel of such a boundary will necessarily contain pixels of at least two clusters other than the boundary. Let there be K clusters $C_1, C_2, \ldots, C_K$. Let p_i be a pixel belonging to cluster C_j. We examine the $n \times n$ neighborhood around p_i, where $n \ (\geq 5)$ is the size of the window. Further, we construct a set $H = \{h_1, h_2, \ldots, h_K\}$ where h_l denotes the number of pixels surrounding p_i and belonging to C_l. So, at least two elements of H, namely, h_l and $h_n \ (l \neq n \neq j)$ become positive. If we do not have at least two such surrounding clusters C_l and C_n, we may ignore the current pixel p_i and proceed to the next. However, even if we find a pixel with two or more different surrounding clusters, it may not be a boundary pixel. A thin object may have a pixel having a few pixels of two or more different clusters around it. Hence, we impose a condition that, if we find h_l and $h_n \ (l \neq n \neq j)$ both greater than $\lfloor h_j/2 \rfloor$, then p_i is treated as a boundary pixel. Such a pixel is to be merged with one of the two surrounding components corresponding to h_l and h_n. Next we find the posterior probabilities of p_i for these two clusters. We assign p_i to the cluster corresponding to the higher posterior probability.

This smoothing algorithm only merges the boundary pixels. Any perturbations inside the image are left intact. Thus we perform only a minor change in the original clustering. We here intend to present the segmentation directly available after the clustering. Thus we compromise with the segmentation quality that may

be improved by applying certain sophisticated post processing methods.

6. RESULTS AND DISCUSSIONS

In this section, we present the segmentation results on the Berkeley segmentation data set (Section 6.1). Further, for comparison, we choose two other mixture approaches. Below, we give a brief description for each of them.

IGMM: The IvMMM has a similarity with the well-known GMM in linear space. Since the three color channels are assumed to be independent, the covariance matrix of the GMM reduces to a diagonal matrix. This version of GMM is termed as IGMM. We apply IGMM in LUV space, which is a linear version of LCH.

BMM: The Beta mixture model [3] can cope with skewed distributions of color channels. This way it becomes stronger than IGMM in modeling pixel distribution. Here, this mixture model is applied on LUV color space. According to the construction, the three channels are assumed independent.

We present here a detailed comparison both visually and with respect to the evaluation scores.

6.1 Berkeley Segmentation Data Set

The *Berkeley segmentation data set* [11] contains several color images (having unique Id) of natural scenes. Most images contain at least one distinguishable and identifiable object embedded in a natural scene. The images we consider have ground-truth in the form of manual segmentation by humans. Manual segmentation is performed by several users independently. Thus the images have multiple ground-truth segmentations. Also, the number of clusters generally varies from one user to another. These manual segmentations are done mostly based on human visual perception. Martin et al. [11] prepared this data set and used it to evaluate the performance of segmentation algorithms and to measure probability distributions associated with Gestalt grouping factors as well as statistics of image region properties.

6.2 Color segmentation results and evaluation

Let us now present the segmentation results produced by all the four algorithms, namely, IvMGMM, IvMBMM, IGMM and BMM. For each of these mixture approaches, we perform 9 independent trials of EM algorithm with $K = 2, 3, \ldots, 10$. Afterwards, we select the value of K at the first local minimum of ICL-BIC criterion, as optimum. After performing EM, the smoothing is applied to merge boundary pixels (Section 5.1). This smoothing is done using a sliding window of size 7×7. To make a visual comparison, we first present some sample results in Fig. 2. The general observation on Fig. 2 is that too many regions are generated by all the mixture approaches, though the actual number of clusters is small. These mixture approaches do not consider spatial information of pixels. Each pixel is assumed to be sampled independently from a model. Thus, these models employ a pixel color based clustering, producing several small regions inside an image. Since we apply ICL-BIC criterion, we only select the number of clusters instead of number of regions. A good post processing (preferably some spatial filters) may be applied after EM to limit the number of regions. We here keep the results intact to measure the relative strength of the original algorithms. For a more robust comparative study than the visual comparison, we accumulate all the scores (i.e., PRI, GCE, VoI and BDE) for all the images. The averages of all the scores are presented in Table 1. Now we are able to compare our algorithm quantitatively with the other two. Table 1 shows that IvMBMM outperforms the other algorithms in terms of GCE and VoI. IvMGMM becomes the best considering the PRI score. BMM outperforms the other algorithms with respect

Table 1. Average performance of different algorithms on the Berkeley Database. The best scores are in boldface.

	PRI	GCE	VoI	BDE
IvMGMM	**0.7076**	0.3774	2.8698	15.7330
IvMBMM	0.7032	**0.3717**	**2.7893**	15.4740
IGMM	0.7037	0.3781	3.1896	10.7770
BMM	0.7062	0.3728	2.8581	**10.5131**

to BDE. Interestingly, IGMM falls behind BMM for all the criteria. The IGMM doest not incorporate skewness of the color channels whereas BMM does. In this way BMM is superior to IGMM in modeling the distribution of color channels. By the same argument, the IvMBMM (uses Beta distributions) is superior to IvMGMM (uses Gaussian distributions). The overall performance of IvMMM is thus satisfactory. We may conclude that IvMMM can provide a suitable mixture model for LCH image data. However, from image segmentation point of view, results may be improved by adding more features (other than L, C and H) and applying a good post processing.

7. SUMMARY AND FUTURE SCOPE

We study color image segmentation in LCH space. The hue is assumed to follow the von-Mises distribution whereas we use two alternative (Gaussian and Beta) distributions for the chroma and the lightness. Assuming the independence of the color channels, we design the IvMG and the IvMB joint distributions for hue, chroma and lightness. Further, the mixture of such joint distributions is used to approximate the pixel distribution. We define two alternative mixtures, namely, IvMGMM and IvMBMM for IvMG and IvMB distributions.
We apply the mixture models on Berkeley segmentation data set. For comparison, we use the existing IGMM and the BMM algorithms. The results show that the joint distributions (IvMGMM and IvMBMM) can well approximate the LCH distribution. Our next work aims to address the possible dependency among the three color channels.

8. REFERENCES

[1] H. D. Cheng, X. H. Jiang, Y. Sun, and J. L. Wang. Color image segmentation: advances and prospects. *Pattern Recognition*, 34:2259–2281, 2001.

[2] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. on PAMI*, 24(8):1026–1038, 2002.

[3] A. Roy, S. K. Parui, and U. Roy. A beta mixture model based approach to text extraction from color images. *Proc. of Int. Conf. on Advances in Pattern Recognition*, pages 321–326. World Scientific, 2007.

[4] K. V. Mardia and P. Jupp. *Directional Statistics*. John Wiley and Sons Ltd., 2000.

[5] J. A. Mooney, P. J. Helms, and I. T. Jolliffe. Fitting mixtures of von mises distributions: a case study involving sudden infant death syndrome. *Computational Statistics and Data Analysis*, 41:505–513, 2003.

[6] A. Banerjee, I. S Dhillon, J. Ghosh, and S. Sra. Clustering on the unit hypersphere using von mises-fisher distributions. *J. of Machine Learning Research*, 6:1345–1382, 2005.

Fig. 2. **(a) The original images and (b), (c), (d), (e) are color segmentation results using IvMGMM, IvMBMM, IGMM and BMM algorithms respectively.**

[7] N. Ludtke, B. Luo, E. Hancock, and R. C. Wilson. Corner detection using a mixture model of edge orientation. *Proc. of Int. Conf. on Pattern Recognition*, pages 574–577, 2002.

[8] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. on PAMI*, 22(7):719–725, 2000.

[9] G. J. Mclachlan and D. Peel. *Finite Mixture Models*. John Wiley & Sons, 2000.

[10] R. Unnikrishnan, C. Pantofaru, and M. Hebert. Toward objective evaluation of image segmentation algorithms. *IEEE Trans. on PAMI*, 29(6):929–944, 2007.

[11] 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 ecolog-

ical statistics. *Proc. of Int Conf. Computer Vision*, pages 416–423, 2001.

[12] M. Meilă. Comparing clusterings: an axiomatic view. In *Proc. of Int. conf. on Machine learning*, pages 577–584, 2005.

[13] J. Freixenet, X. Muñoz, D. Raba, J. Martí, X. Cufí. Yet another survey on image segmentation: Region and boundary information integration. *Proc. of European Conf. on Computer Vision-Part III*, pages 408–422, 2002.

[14] A. Y. Yang, J. Wright, Y. Ma, and S. S. Sastry. Unsupervised segmentation of natural images via lossy data compression. *Computer Vision and Image Understanding*, 110:212–225, 2008.