A Color Based Image Segmentation and its Application to Text Segmentation

Anandarup Roy, Swapan Kumar Parui
CVPR Unit
Indian Statistical Institute
Kolkata, India
{*ananda_t,swapan*}@*isical.ac.in*

Amitav Paul, Utpal Roy
DCSS
Siksha Bhavana
Visva-Bharati
{*paul.amitav,roy.utpal*}@*gmail.com*

Abstract

The goal of this article is twofold. First, it deals with color image segmentation in hue-saturation space. A model for circular data is provided by the vM-Gauss distribution, which is a joint distribution of von-Mises and Gaussian distributions. The mixture of vM-Gauss distributions is used to model hue-saturation data. After segmentation, a post processing based on both spectral and spatial similarity of clusters is applied to separate such identifiable objects in the image. The results and comparisons are shown on Berkeley segmentation dataset. The problem of text extraction from a color image is taken as an application of the proposed method. We use a laboratory made text image dataset to test the method.

1. Introduction

Color based image segmentation [4] 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 paper, we study the problem of clustering based color image segmentation. The 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. Park et. al [12] apply this algorithm to a pattern space represented by RGB coordinates. Image segmentation can also be viewed as a hidden variable problem. Segmenting an image into clusters involves determining which cluster generates the image pixels, which is the hidden information. In this framework, a mixture model based approach is suitable for segmentation. To this direction an attempt was taken by Carson et. al. [3], using Gaussian mixture model (GMM). Roy et. al. [13] considered skewed and independent distributions of the RGB color bands, and used a mixture of Beta distributions to model the image data. In [2] Bougulia et. al. applied mixture of Dirichlet distributions to approximate the joint distribution of RGB color spectrum. In this paper the color information is analyzed in the hue and saturation space of the HSV color model. The HSV model has a mixed space in the sense that it is a mixture of angular (hue) and linear (saturation and intensity) data. Nevertheless, it is important to take into account these mixed characteristics.

The von-Mises (vM) [9] distribution is defined on an unit circle and analogous to univariate Gaussian distribution in $\Re^2$. Several attempts are taken to estimate the parameters of the mixture of vM distributions. Mooney et al. [11] used a mixture of two circular vM distributions and estimated the parameters using a quasi-Newton procedure. Banerjee et. al [1] estimated the parameters of von-Mises Fisher distribution which is a generalization of vM distribution. The vM distribution also has application in image processing. Ludtke et. al. [7] used a mixture of vM distributions to model the orientation of the contour. Further this information is used to detect corners. In this paper, hue and saturation are considered independent and the joint distribution is formulated with a vM and a Gaussian distributions. The parameters are estimated using expectation algorithm (EM) algorithm. A post processing method based on both spatial and spectral similarity is used to merge the clusters in order to retrieve the meaningful objects present in a scene. The Berkeley segmentation dataset [10] is used to test the performance of the method. A comparison with GMM in RGB space is presented with respect to some images from the Berkeley dataset. Further we apply the method to extract color texts embedded in a color image. A number of efforts have been taken towards this problem. Hase et.al. [5] divide a full color image into several representative color images. Then, character strings are nominated from each image and refined by conflict resolution methodology. Mancas-Thillou and Gosselin [8] propose a selective metric-based clustering to extract textual information in real-world images. Liu et. al. [6] extract text through labeling connected components of the binary image as character or non-character according to its neighbors. The case of

978-0-7695-3476-3/08 $25.00 © 2008 IEEE
DOI 10.1109/ICVGIP.2008.69

three neighboring characters is represented as the GMM. In this paper we first segment the color image irrespective of whether there is text in the image. Further each of the connected components of the segmented image is studied and based on a measure of elongatedness the text parts are identified. We test this method on a laboratory made color text dataset.

The paper is organized as follows: Section 2 describes the vM-Gauss mixture distribution. In Section 3 we design an EM framework for parameter estimation. Section 4 describes the application of the mixture model to color segmentation problem. In Section 5 the object detection procedure is designed along with the similarity measures. Section 6.1 gives the results on Berkeley segmentation dataset. The results on color text image dataset are given in Section 6.2. Finally we draw the conclusions and suggest possible improvements in Section 7.

2. Mixture of vM-Gauss Distributions

A pair of independent random variables $(\Theta, X), \theta \in [0, 2\pi), x \in (-\infty, \infty)$ is said to follow vM-Gauss distribution if its probability density function is given by

$$f(\Theta, X|\mu, \kappa, \nu, \sigma) = f_1(\Theta|\mu, \kappa).f_2(X|\nu, \sigma), \qquad (1)$$

where

$$\begin{aligned} f_1(\Theta|\mu, \kappa) &= \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\mu - \theta)) \\ \text{and } f_2(X|\nu, \sigma) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\nu)^2}{2\sigma^2}\right). \end{aligned}$$

Here, $I_0(\kappa)$ is the modified bessel function of the first kind of order 0 and argument κ. f_1 is the von-Mises distribution with mean $\mu \in [0, 2\pi)$ and concentration parameter $\kappa \in \Re^+ \cup \{0\}$. When $\kappa = 0$, the density gives the uniform distribution on $[0, 2\pi)$. f_2 is the Gaussian distribution with parameters $\nu \in \Re$ and $\sigma \in \Re^+$. Note, the wrapped normal (WN) distribution has also similar shape as the vM distribution. The WN distribution has the form:

$$f_{WN}(y|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp \frac{-(y + 2\pi k - \mu)^2}{2\sigma^2} \qquad (2)$$

where, $0 \le y < 2\pi$. One may use WN distribution instead of vM distribution. With some minor conditions the WN can well approximate the vM distribution. However it involves infinite sum, hence the log-likelihood become tedious. We thus find vM more suitable for mixture model. The bessel function is defined as: $I_0(\kappa) = (1/2\pi) \int_0^{2\pi} \exp(\kappa \cos\alpha) d\alpha$. This seems to be difficult to evaluate. Fortunately we have a polynomial approximation to $I_0(\kappa)$. A mixture of K vM-Gauss distributions is given by:

$$f(\Theta, X|\Xi) = \sum_{h=1}^{K} P(h) f(\Theta, X|h, \Xi_h). \qquad (3)$$

where $P(h)(0 \le P(h) \le 1$ and $\sum_{h=1}^{K} P(h) = 1)$ are the mixing proportions, $f(\Theta, X|h, \Xi_h)$ is a vM-Gauss distribution representing the h^{th} component of the mixture and $\Xi_h = (\mu_h, \kappa_h, \nu_h, \sigma_h)$ is the set of parameters of the h^{th} component. The symbol $\Xi = (\Xi_1, \ldots, \Xi_K, P(1), \ldots, P(K))$ refers to the entire set of parameters to be estimated.

3. Maximum Likelihood Estimation

Let $\aleph = p_1, \ldots, p_N$ be a finite set of N samples drawn independently from the mixture of vM-Gauss distributions. Here $p_i = (\theta_i, x_i)$ is a pair of circular and linear data. To maximize the likelihood function, the expectation maximization (EM) is most widely used. The standard EM settings express the mixture model in terms of hidden variables represented by:

$$q(h|p_i) = \frac{P(h) f(p_i|h, \Xi_h)}{\sum_{l=1}^{K} P(l) f(p_i|l, \Xi_l)}. \qquad (4)$$

With this definition of hidden variables, the log likelihood function for mixture of vM-Gauss distribution can be written as:

$$\begin{aligned} \Phi(\aleph, \Xi, \lambda) = \sum_{i=1}^{N} \sum_{h=1}^{K} \left[\ln\left(P(h) f(p_i|h, \Xi_h)\right)\right] q(h|p_i) \\ + \lambda \left(1 - \sum_{h=1}^{K} P(h)\right). \end{aligned}$$

where λ is the lagrange multiplier.

To maximize $\Phi(\aleph, \Xi, \lambda)$ we may go independently for a priory probabilities $P(h)$ and the parameters Ξ_h. The a priori probabilities can be found out with:

$$P(h) = \frac{1}{N} \sum_{i=1}^{N} q(h|p_i). \qquad (5)$$

Differentiating $\Phi(\aleph, \Xi, \lambda)$ with respect to μ_h we get the expression for μ_h as:

$$\mu_h = tan^{-1} \left(\frac{\sum_{i=1}^{N} \sin\theta_i q(h|p_i)}{\sum_{i=1}^{N} \cos\theta_i q(h|p_i)} \right). \qquad (6)$$

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

$$A_2(\kappa_h) = \frac{I_1(\kappa_h)}{I_0(\kappa_h)} = \frac{\sum_{i=1}^{N} \cos(\mu_h - \theta_i) q(h|p_i)}{N_h}. \quad (7)$$

with $N_h = \sum_{i=1}^{N} q(h|p_i)$. 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 κ is the best choice for estimating κ. Such approaches also take constant computation time unlike any iterative numerical method. We here use the approximation in Banerjee et. al [1] to find κ. Thus κ can be estimated by:

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

where $A_2(\kappa) = \overline{r}$. Other approximations of κ are discussed by Mardia and Jupp [9].

The expressions for ν_h and σ_h are given by:

$$\nu_h = \frac{1}{N_h} \sum_{i=1}^{N} q(h|p_i) x_i \quad (8)$$

$$\sigma_h^2 = \frac{1}{N_h} \sum_{i=1}^{N} q(h|p_i)(x_i - \nu_h)^2. \quad (9)$$

4. Color Image Segmentation

The HSV system separates color information from intensity information and thus is more suitable for color segmentation than the RGB system. Color information is presented by hue and saturation values. Since hue represents circular data, circular distributions should be considered for its statistical study. We assume the hue and saturation values of pixels in a color image arise from a finite mixture of vM-Gauss distributions (Eq. 1). The parameters of such a mixture are estimated using the EM method. The expectation (E) step computes $q(h|p_i)$ by Eq. 4 using initially estimated parameters. The maximization (M) step estimates Ξ using Eqs. 5, 6, 7, 8 and 9. The steps iterate until the log-likelihood stabilizes. We apply K-Means algorithm in hue and saturation space to obtain initial clustering. Initially, μ_h, ν_h and σ_h^2 are the circular mean, the arithmetic mean and the variance respectively. The concentration parameter κ_h is found by:

$$\kappa_h = A_2^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\cos(\mu_h - \theta_i)\right).$$

The a priori probabilities are set with $P(h) = \frac{N_h}{N}$ where N_h is the number of elements in h^{th} cluster. In order to detect the number of clusters automatically we use *Minimum Mutual Description Length (MMDL)* criteria. The value of K at the first local minimum of *MMDL* is chosen as the optimum value of K. The gray portions of the images need special consideration since the hue becomes undefined for the gray portions. We separate out the gray portions before employing EM algorithm. A separate clustering should be employed with the gray portions. Yet, here we encounter only a few gray pixels mostly in the background and thus leave them to minimize time complexity.

5. Object Detection using Similarity Measures

A Berkeley dataset image contains at least one object that can be identified from the background. Though the object has spectral (hue and saturation) values that are reasonably close, their distribution is not in general unimodal and normally contains several unimodal components that lie close to one another. Thus, after clustering with EM, each of the object and background contains multiple color clusters. In the post processing stage, the task is to merge individual clusters to achieve proper image segmentation. For this, we consider both spectral similarity and spatial closeness between clusters. Let $q \in N(p)$ if q is a 4-neighbor of p. Let the clusters obtained with EM algorithm be denoted by $C_1, C_2 \ldots$. Consider a matrix S with entries

$$s_{ij} = |\{p | p \in C_i \text{ and } q \in C_j \text{ and } q \in N(p)\}|$$

The value of s_{ij} is normalized by dividing it by the total number of boundary pixels of C_i, and then multiplying it by 200. s_{ij} indicates how spatially close C_i is to C_j, that is, the higher the value of s_{ij}, closer is C_i to C_j. For example, $s_{ij} = 200$ means C_i is completely surrounded by C_j. One may consider 8-neighborhood instead of 4-neighborhood, with some added complexity. Now, for a spectral similarity measure, an upper triangular matrix D is defined as

$$\begin{aligned} d_{ij} &= 0 \text{ if } i \geq j \\ &= dist(C_i, C_j), \text{ otherwise.} \end{aligned}$$

where,

$$dist(C_i, C_j) = \frac{\min(|h_i - h_j|, 2\pi - |h_i - h_j|)}{\pi} + |t_i - t_j|$$

where division by π is for normalization. $dist(C_i, C_j)$ is the hue-saturation distance between C_i and C_j. h_i and t_i are hue and saturation of cluster C_i. Here, lower values of d_{ij} indicate more spectral similarity between C_i and C_j. Merging of clusters is done on the basis of both s_{ij} and d_{ij} using the following strategy. Identify the pair of clusters with minimum d_{ij} and merge the pair C_i and C_j if

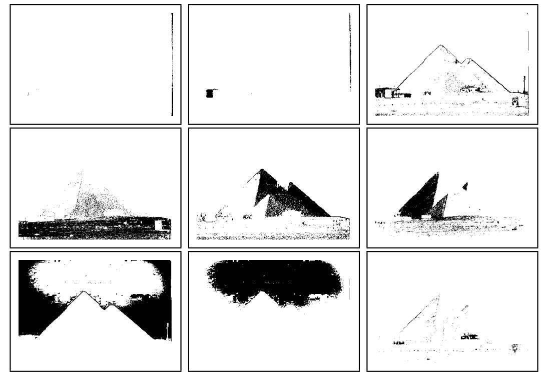
$max(s_{ij}, s_{ji}) > T_1$ and $min(d_{ij}) < T_2$ where T_1 is a threshold value (max is considered since s_{ij} is asymmetric and merging is done on the basis of single linkage principle) and T_2 is another threshold value. Next consider the pair of clusters with second minimum d_{ij} and merge them provided the above conditions are satisfied.

6. Results and Discussions

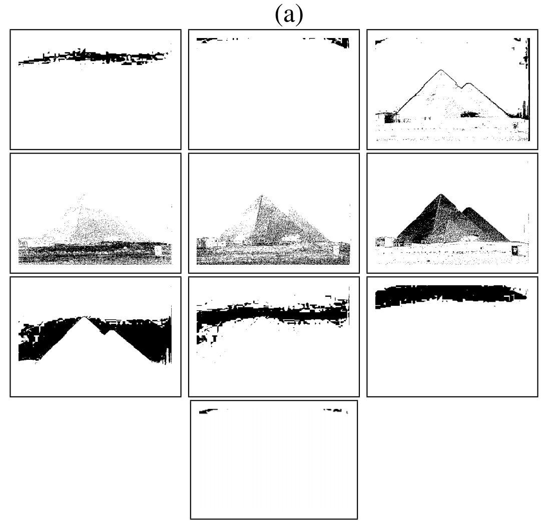
In this section we describe the results obtained by applying the vM-Gauss mixture distribution. The results are presented for two different datasets. The *Berkeley segmentation dataset* contains color images of natural scenes. A laboratory made color text dataset is used to find an application of the segmentation algorithm in text extraction. The goal here is to separate text portions from color images. In order to identify the text portions we apply a text identification method based on elongatedness of the connected components emerging from each cluster.

6.1 Results on Berkeley segmentation dataset

The *Berkeley segmentation dataset* [10] contains several color images along with human segmentation results. The images contain at least one identifiable object. This dataset is used to measure the performance of our algorithm. In Fig. 1 we present segmentation results on four sample images taken from the Berkeley dataset. The gray portions in the original images are shown in black in the segmented images. Let us consider the seventh image of Fig. 1. We execute GMM in RGB space and the vM-Gauss mixture to segment the image. In Fig. 2 we give individual clusters obtained after executing the GMM and the vM-Gauss method. The optimal number of clusters is 9 as suggested by the MMDL criteria. Here the sample image contains an identifiable pyramid object. After executing GMM, the pyramid object is split into two parts as in Fig. 2 (a). On the other hand, when we apply the vM-Gauss model, the pyramid is almost retained as shown in the Fig. 2(b). Clearly the clusters in Fig. 2 (b) are more desirable. The reason behind the behaviors of GMM and vM-Gauss is that the pyramid object has more than one shade. As the GMM cannot avoid the brightness information, the pyramid is divided according to variation in brightness. The vM-Gauss model on the other hand, gets rid of the brightness information and uses only the color information. The color difference within the pyramid is not much prominent and hence the object preserved. Let us now consider the "Star Fish" image, i.e., the third image of Fig. 1. The clusters obtained by GMM is shown in Fig. 3, whereas in Fig. 4 we present the clusters by vM-Gauss method. In order to retrieve the object (the Star



(a)



(b)

Figure 2. Segmentation results with (a) GMM in RGB space and (b) with vM-Gauss mixture in HS space.

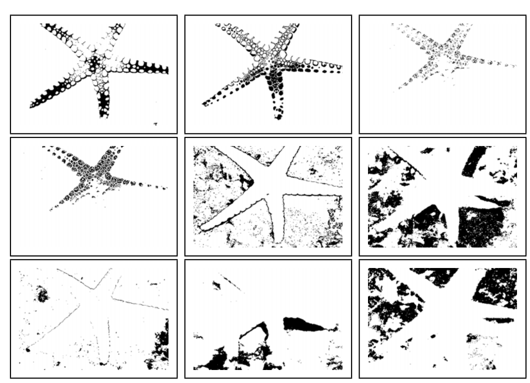


Figure 3. Initial clusters obtained by applying GMM on the "Star Fish" image.

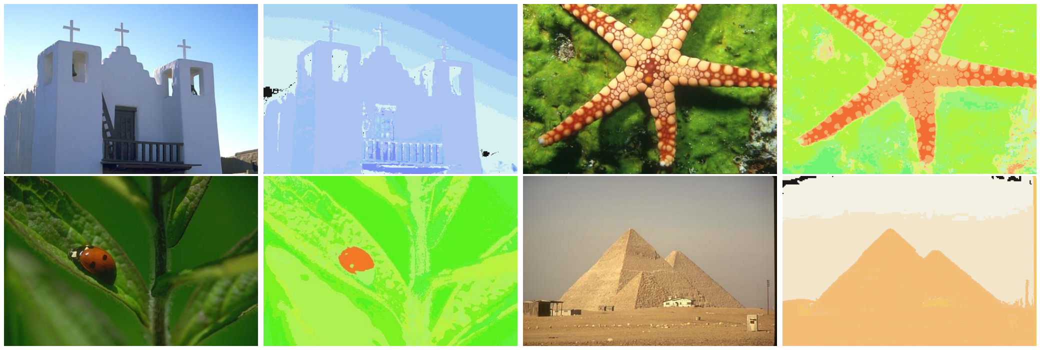


Figure 1. Sample images from Berkeley segmentation dataset: The original and the segmented (by vM-Gauss) images are shown side by side.

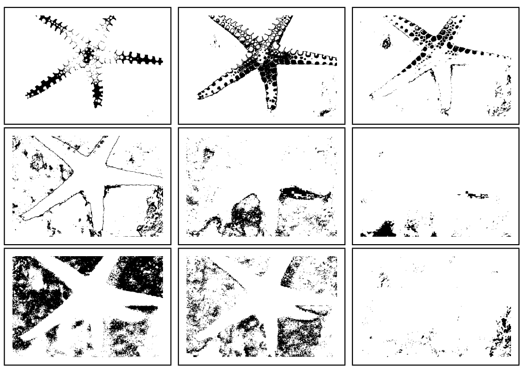


Figure 4. Initial clusters obtained by applying vM-Gauss on the "Star Fish" image.

Fish) we put the threshold $T_1 = 30$ and $T_2 = 0.25$. The resulting clusters are shown in Fig. 5. The merging sequence is $(7, 8)$, $(5, 6)$ and $(5, 8)$ where cluster numbers are same as window numbers in Fig. 4. Note, only the background portions are merged and still the object is divided into more than one clusters. Now, we increase $T_2 = 0.30$ taking same

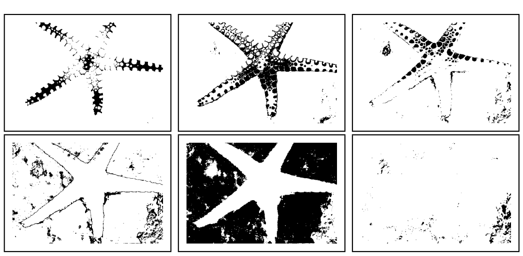


Figure 5. Merging results with $T_1 = 30$ **and** $T_2 = 0.25$**.**

T_1. The clusters with the object are now merged as shown in Fig. 6. Now, the Star Fish almost resides in a separate cluster. This is good enough. Yet, increasing T_2 to 0.35 yields only two clusters (Fig. 7). In Fig. 8 the merging re-

Figure 6. Merging results with $T_1 = 30$ **and** $T_2 = 0.30$**.**

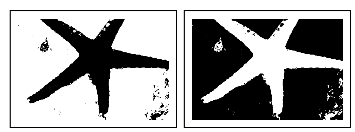


Figure 7. Merging up to two clusters $(T_1 = 30$ **and** $T_2 = 0.35)$**.**

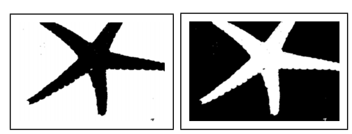


Figure 8. Merging results for clusters from GMM $(T_1 = 30$ **and** $T_2 = 0.50)$**.**

sults of the clusters obtained from the GMM method (Fig. 3) are shown. Note here after merging we get almost equivalent results from vM-Gauss and GMM. Yet one may notice that the object (the "Star Fish") is divided into several parts after GMM based process. As an effect the merging needs more attention and we have to wait till merging up to two clusters to separate the object. On the other hand merging after vM-Gauss yields the object before going to ultimate two clusters.

6.2 Results on the Color text dataset

Besides the Berkeley segmentation dataset we use a laboratory made *color text dataset* containing a number of color images having text. The text portions have homogenous color as compare to the background. The background, however, may have colors similar to the text portions and thus make it difficult to be separated. We apply the vM-Gauss mixture model and the similarity based object detection algorithm. Here the text portions are considered as objects to be separated. In Fig. 9 we show some examples from color text dataset along with the segmentation results. Several situations with respect to the text portions may occur after segmentation.

(a) (b)

Figure 9. Sample images from Color text dataset: (a) original images and (b) the segmented images.

1. Normally the text parts make one single cluster.
2. However, text parts may belong to several clusters. This is when texts in the image have more than one color.
3. A single cluster may have both text and non-text parts. This is when some text and some non-text region in the image have the same color.

The connected components are now obtained from each cluster. Sufficiently small components are removed. We take into account the degree of elongatedness of a connected component to distinguish between text and non-text regions of the image. The text like patterns are usually elongated. A measure of elongatedness of a component is now defined as:

$$\text{Elongatedness ratio(ER)} = \frac{\text{No. of boundary points}}{\sqrt{\text{Total no. of points}}}. \quad (10)$$

A connected component is text if ER $> T_3$, where T_3 is a threshold. We take the "Organic" image at the third row of Fig. 9 as an example. It has two separate text portions, the text "Organic" and the text "GARDENING". After applying the vM-Gauss mixture model we get 5 clusters (Fig. 10). The spatial and the spectral similarity matrices for these 5 clusters are shown in Table 1 and Table 2 respectively. Note we use -1 for invalid distances. We take $T_1 = 50$ and $T_2 = 0.30$ here. According to the merging procedure the second and the third clusters are most likely to be merged. The merging sequence is $(2, 3)$ and $(1, 3)$. After merging we obtain 3 clusters (Fig. 11(a)). The ER for each connected components are computed according to Eq. 10. The extracted text portions (with $T_3 = 5.00$) are shown in Fig. 11(b). Some more samples of extracted text are shown in Fig. 12 along with the original image. Note, in the "COASTAL" image (the first image in Fig. 9) has much confusion in between foreground (text) and background in

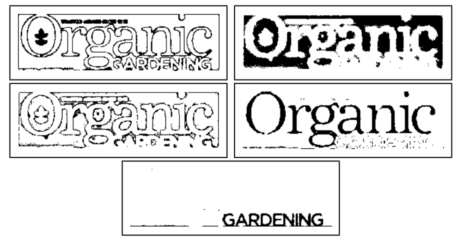


Figure 10. Clusters obtained from "Organic" image.

Table 1. The spatial similarity matrix for the "Organic" image.

0.00	3.10	101.37	75.00	20.54
5.81	0.00	194.19	0.00	0.00
98.47	100.43	0.00	0.72	0.38
171.72	0.00	1.70	0.00	26.58
126.26	0.00	2.38	71.36	0.00

Table 2. The spectral similarity matrix for the "Organic" image.

-1.00	0.38	0.27	0.31	0.53
-1.00	-1.00	0.08	0.65	0.86
-1.00	-1.00	-1.00	0.57	0.81
-1.00	-1.00	-1.00	-1.00	0.34
-1.00	-1.00	-1.00	-1.00	-1.00

terms of hue. As we use saturation component we can successfully separate the text portion.

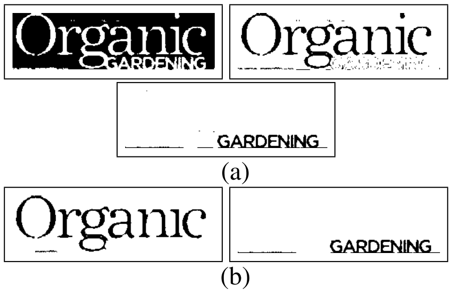


Figure 11. (a) Clusters obtained from "Organic" image after merging and (b) the two text portions.

(a) (b)

Figure 12. Results of text extraction (a) Original images and (b) the extracted text portions.

7. Conclusions and Future Scope

We study color image segmentation in HS space. The hue values are assumed to follow the vM distribution whereas the saturation values follow the Gaussian distribution independently. Results show the joint distribution (vM-Gauss) can well approximate the hue and saturation distribution. Further works should address possible dependency between hue and saturation. The post processing methodologies can be improved by introducing an objective function instead of the empirical thresholds.

8. Acknowledgements

One of the authors (Utpal Roy) acknowledges the UGC project Ref. No. F-33-437/2007(SR) for financial support.

References

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

[2] N. Bougulia, D. Ziou, and J. Vaillancourt. Unsupervised learning of a finite mixture model based on the dirichlet distribution and its application. *IEEE Trans. on Image Processing*, 13(13), 2004.

[3] 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, 2002.

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

[5] H. Hase, T. Shinokawa, M. Yoneda, and C. Y. Suen. Character string extraction from color documents. *Pattern Recognition*, 34:1349–1365, 2001.

[6] X. Liu, H. Fu, and Y. Jia. Gaussian mixture modeling and learning of neighboring characters for multilingual text extraction in images. *Pattern Recognition*, 41:484–493, 2008.

[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. Mancas-Thillou and B. Gosselin. Color text extraction with selective metric-based clustering. *Computer Vision and Image Understanding*, 107:97–107, 2007.

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

[10] 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. *Proc.* 8^{th} *Int'l Conf. Computer Vision*, pages 416–423, 2001.

[11] 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.

[12] S. Park, I. Yun, and S. Lee. Color image segmentation based on 3-d clustering: Morphological approach. *Pattern Recognition*, 31(8):1061–1076, 1998.

[13] 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, 2007.