# Bhattacharyya Probabilistic Distance of the Dirichlet Density and its Application to Split-And-Merge Image Segmentation

T. W. Rauber
Departamento de Informática, CT
Univ. Federal do Espírito Santo
Av. Fernando Ferrari s/n, Vitória,
29060-970, Espírito Santo, Brazil
Phone: (55) 27-3335.2654 Fax: (55) 27-3335.2850 E-mail: thomas@inf.ufes.br

A. Conci
Instituto de Computação,
Departamento de Informática
Universidade Federal Fluminense,
Rua Passo da Pátria 156, Niterói,
24210-240, Rio de Janeiro, Brazil
Phone: (55) 21- 2629.5479 Fax: (55) 21- 2629.5669 E-mail: aconci@ic.uff.br

T. Braun and K. Berns
Robotics Research Lab
Department of Computer Science
University of Kaiserslautern
Gottlieb-Daimler-Straße, Gebäude 48,
Kaiserslautern, 67663, Germany
Phone: (49) 631-205.2689 Fax: (49) 631-205.2640 E-mail: {braun,berns}@informatik.uni-kl.de

**Keywords: Probabilistic distance measures, Bhattacharyya distance, Dirichlet distribution**

**Abstract – We derive the Bhattacharyya distance between two Dirichlet densities. As an application we use image segmentation by a Split-and-Merge algorithm.**

## 1. INTRODUCTION

Pratt [1] emphasizes the importance of image features for the segmentation of contiguous regions. In the context of region characterization by texture features, Faugeras and Pratt [2] measure the difference between regions by a probabilistic distance measure [3], the Bhattacharyya distance [4], [5], [6]. They point out that the measure should be applicable to other image features. In this work we investigate the use of the Bhattacharyya distance between luminance based image features, typically three channel color information like RGB or HSV. As an underlying feature model we use a Dirichlet density of the intensity normalized color values. Compared to a Gaussian distribution it has advantages if the density is asymmetric and has a compact support. For instance if the expected value of some feature is close to the maximum and minimum value, together with a high variance, the Gaussian bell shaped density allows implausible values below zero or above the maximum permitted value.

Bouguila [7] have proposed mixture Dirichlet densities in feature-space based human skin segmentation. In an image- domain region growing segmentation approach the use of mixture models is limited due to the following facts: small regions usually do not have a multimodal density; the estimation of mixture parameters is difficult and does not guarantee success; the estimation of simple (i.e. non-mixture) Dirichlet parameters is relatively fast and straightforward and therefor appropriate for a Split-and-Merge segmentation process; most important is the fact the homogeneity of a region is by nature antagonistic to a mixed, heterogeneous probability density model.

The paper is organized in the following manner: Section 2 introduces homogeneity criteria in segmentation algorithms which guide the region grouping. Section 3 explains the Dirichlet distribution. The analytical definition in the case of a Dirichlet density of the Bhattacharyya probabilistic distance is given in section 4. Experimental results are shown in section 5 and finally the conclusions are drawn in section 6.

## 2. HOMOGENEITY CRITERIA IN IMAGE-DOMAIN BASED REGION SEGMENTATION

Lucchese and Mitra survey segmentation methods for color images [8]. We limit our considerations to what they call image-domain based segmentation which considers spatial relationships among image points. There exist basically two approaches: split-and-merge and region growing. They have in common that a homogeneity predicate must be defined which is true within the definition of a region and false for distinct regions. The most elementary criterion is a difference of the mean gray-level of an existing region and a candidate that falls below a certain threshold. If it is satisfied that candidate is incorporated into the existing region. The Dirichlet probability distribution has been used in a variety of supervised and unsupervised pattern recognition do- mains, for instance in the context of bioinformatics for protein sequence analysis [9], feature domain image segmentation [7] or text analysis [10]. In this work we will explore the Dirichlet model in an image-domain region segmentation algorithm.

## 3. DIRICHLET PROBABILITY DENSITY

### *3.1 Definition*

The Dirichlet probability density function (pdf) of a multi-variate random variable $\mathbf{x}$ with parameters $\boldsymbol{\alpha}$ is defined as

$$p(\boldsymbol{x};\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})}\prod_{k=1}^{d} x_k^{\alpha_k-1} \quad (1)$$

with constraints $\alpha_k > 0$ and $\sum_k x_k = 1$. The normalizing constant $B(\boldsymbol{\alpha})$ which forces $p(\boldsymbol{x};\boldsymbol{\alpha})$ as a pdf to integrate to unity over the domain of x is the *multinomial beta function* with value $B(\boldsymbol{\alpha}) = \prod_{k=1}^{d}\Gamma(\alpha_k)\,/\Gamma(\sum_{k=1}^{d}\alpha_k)$. The density $p(\boldsymbol{x};\boldsymbol{\alpha})$ has a compact support over a sub-domain of $\mathbb{R}^d$ which is a $(d-1)$-simplex and possesses a great variety of shapes, U-shaped, bell-shaped, exponentially-shaped, depending on the values of the parameters $\alpha_k$. As a representative model we will use throughout the paper the $d = 3$ intensity normalized RGB values of a pixel as features, hence

$$x = [\,R \;\; G \;\; B\,]^T/(R+G+B) \quad (2)$$

which satisfies the sum-to-unity constraint of its component features $x_1$, $x_2$, $x_3$. Other feature models with this constraint are equally usable. The 2-simplex in this case is the triangle in the intensity normalized RGB space, spanned among the points (1, 0, 0), (0, 1, 0), (0, 0, 1). Any color triple RGB in projected onto the simplex by the normalization process, in a similar way as a XYZ color onto xy-space in the CIE chromaticity diagram [1].

### *3.2 Parameter estimation*

We want to know the *d* values $\alpha_k$, given as the only information *n* samples $\boldsymbol{x}_i \in \mathbb{R}^d$ which form the data set $\mathcal{D} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$. Minka [11] derives for the maximization of the Log-likelihood function $L(\mathcal{D}|\boldsymbol{\alpha})$ an iterative fixed-point estimation which will be used here.

In a segmentation algorithm, when we melt two regions $\mathcal{R}_a$, $\mathcal{R}_b$, characterized by parameters $\boldsymbol{\alpha_a}$, $\boldsymbol{\alpha_b}$, the parameters $\boldsymbol{\alpha_{a \cup b}}$ of the union of the two regions have to be re-estimated. In order to avoid the calculus of $\boldsymbol{\alpha_{a\cup b}} \in \mathbb{R}^d$ from scratch, an efficient, incremental estimation for updating the Dirichlet parameters is proposed. Suppose that from the *n* pixels of region $\mathcal{R}_a$ the parameters $\boldsymbol{\alpha_a} = \{\alpha_{\boldsymbol{a1}}, \dots, \alpha_{\boldsymbol{ad}}\}$ have been estimated. For the merged region $\mathcal{R}_{\boldsymbol{a\cup b}}$, we first update the mean of the logarithm of the *k*th component of the sample value which is needed in Minka's algorithm. If a single pixel of region $\mathcal{R}_b$ is added as the *(n+1)*th pixel of $\mathcal{R}_{\boldsymbol{a\cup b}}$ we have the update rule (3), based on the Robbins-Monroe algorithm [12], [13]

$$\overline{log\, x_k}^{(n+1)} = \frac{1}{n+1}\{n\, \overline{log\, x_k}^{(n)} + log\, x_{n+1,k}\}. \quad (3)$$

Since we have only to update a mean value of a union of two sets, given two mean values of the sets, a direct update of mean log values of the union $\overline{log\, x_k}$ is also possible:

$$\overline{log\, x_k}^{(n_a+n_b)} = \frac{n_a \overline{log\, x_k}^{(n_a)} + n_b \overline{log\, x_k}^{(n_b)}}{n_a+n_b}. \quad (4)$$

## 4. BHATTACHARYYA PROBABILISTIC DISTANCE MEASURE

We have assumed that a region of an image is characterized by features $x_1, \dots, x_d$ which are assembled into the multivariate feature vector $x = [\, x_1 \ \dots \ x_d]^T$. The most simple case is when the region is only described by the gray value *f* of its pixels at position $(u, v)$. In this case we have $d = 1$ and $\boldsymbol{x} = x = x_1 = f(u, v)$.

### *4.1 Analytical definition*

The values of the feature vectors of two regions $\mathcal{R}_\mathrm{a}$ and $\mathcal{R}_\mathrm{b}$ form conditional probability density functions $\mathrm{p_a}(\mathrm{x}|\mathcal{R}_\mathrm{a})$ and $\mathrm{p_b}(\mathrm{x}|\mathcal{R}_\mathrm{b})$, where the subscripts in p emphasize the fact the we deal with two different functional forms. The *Bhattacharyya* coefficient $\rho \in [0,1]$ between two probability distributions described by the functional forms $p_a$, $p_b$ and their respective parameters $\theta_a$, $\theta_b$ is defined [4], [5] as

$$\rho(\theta_a, \theta_b) = \int \sqrt{p_a(\boldsymbol{x}|\mathcal{R}_a;\theta_a)p_b(\boldsymbol{x}|\mathcal{R}_b;\theta_b)}\, d\boldsymbol{x}. \quad (5)$$

For reasons of clarity the functional forms of the pdfs $p_a$ and $p_b$ have be omitted in the argument list. Note however that the Bhattacharyya coefficient is defined between densities that do not necessarily have to come from the same functional family.

From $\rho$ the Bhattacharyya distance (or B-distance for short) can be defined as

$$B = -\ln \rho, \qquad B \in [0, \infty]. \quad (6)$$

An alternative which obeys the triangle inequality [9] is

$$B' = \sqrt{1-\rho}, \qquad B' \in [0,1]. \quad (7)$$

### *4.2 Bhattacharyya distance between Dirichlet densities*

If we apply eq. (5) to two Dirichlet distributed densities, we obtain for the Bhattacharyya coefficient

$$\rho(\boldsymbol{\alpha_a}, \boldsymbol{\alpha_b}) = \frac{1}{\sqrt{B(\boldsymbol{\alpha_a})B(\boldsymbol{\alpha_b})}} \int \prod_{k=1}^d x_k^{\frac{\alpha_{ak}}{2}+\frac{\alpha_{bk}}{2}-1}\, d\boldsymbol{x}. \quad (8)$$

Let $\beta_k := \frac{\alpha_{ak}}{2} + \frac{\alpha_{bk}}{2}$, $k = 1, \dots, d$. Then $\beta_k$ satisfies the constraint $\beta_k > 0$ and $p(\boldsymbol{x};\boldsymbol{\beta})$ is a Dirichlet distribution with parameters $\boldsymbol{\beta}$. Hence $\int p(\boldsymbol{x};\boldsymbol{\beta}) = 1$ or equivalently from the definition of eq. (1) we find that $\int \prod_{k=1}^d x_k^{\beta_k - 1} = B(\boldsymbol{\beta})$. Plugging this into eq. (8) we define the Bhattacharyya coefficient for the probabilistic distance between two Dirichlet densities as

$$\rho(\boldsymbol{\alpha_a}, \boldsymbol{\alpha_b}) = \frac{B\left(\frac{\boldsymbol{\alpha_a}}{\boldsymbol{2}} + \frac{\boldsymbol{\alpha_b}}{\boldsymbol{2}}\right)}{\sqrt{B(\boldsymbol{\alpha_a})B(\boldsymbol{\alpha_b})}} =$$

$$\frac{\prod_{k=1}^d \Gamma\left(\frac{\alpha_{ak}}{2}+\frac{\alpha_{bk}}{2}\right)}{\Gamma\left(\frac{1}{2}\sum_{k=1}^d(\alpha_{ak}+\alpha_{bk})\right)} \frac{\sqrt{\Gamma(|\boldsymbol{\alpha_a}|)\Gamma(|\boldsymbol{\alpha_b}|)}}{\sqrt{\prod_{k=1}^d \Gamma(\alpha_{ak})}\sqrt{\prod_{k=1}^d \Gamma(\alpha_{bk})}} \quad (9)$$

The definition of the Bhattacharyya distance $B$ in the case of Dirichlet distributions is motivated by the fact the Gamma function $\Gamma(.)$ produces numerical overflows if its argument comes near the value 171.61. Especially if the variance of the provided samples is small this threshold is easily reached. Therefor the logarithmic form of eq. (6) for $B$ is preferred because it permits high values for the $\alpha_k$ parameters. The Bhattacharyya distance between two Dirichlet densities is then

$$B(\boldsymbol{\alpha_a}, \boldsymbol{\alpha_b}) = -ln\, \rho(\boldsymbol{\alpha_a}, \boldsymbol{\alpha_b}) =$$

$$= -ln\, \Gamma\left(\sum_{k=1}^d \frac{\alpha_{ak}+\alpha_{bk}}{2}\right) + \frac{1}{2}\left\{\sum_{k=1}^d ln\, \Gamma(\alpha_{ak}) + \sum_{k=1}^d ln\, \Gamma(\alpha_{bk})\right\} - \sum_{k=1}^d ln\, \Gamma\left(\frac{\alpha_{ak}+\alpha_{bk}}{2}\right) - \frac{1}{2}\{ln\, \Gamma(|\boldsymbol{\alpha_a}|) + ln\, \Gamma(|\boldsymbol{\alpha_b}|)\} \quad (10)$$

Since the Dirichlet distribution is a generalization of the Beta distribution, the Bhattacharyya coefficients and distances for this specializations is easily derived from (9) and (10).

## 5. EXPERIMENTAL RESULTS

It should be emphasized that the principal result of this work is the theoretical analysis of the Bhattacharyya probabilistic distances of the Dirichlet distribution. The following experiments do not claim to propose methods that are superior to others in the application area mentioned. They illustrate the process of statistical modeling with the Dirichlet distribution, parameter estimation and probabilistic distance measuring. We divide the evaluation of the proposed methods into the theoretical parameter estimation and an a image segmentation application. First we present the behaviour of the learning of the parameters for artificially generated Dirichlet distributed data samples. Then the learning is observed in the unsupervised case when the true distribution of the data assumes a Dirichlet distribution and the true parameters are not accessible. Finally a Split-And-Merge image segmentation is performed.

### *5.1 Parameter estimation for synthetic Dirichlet distribution*

A Dirichlet distributed data sample $\boldsymbol{x}$ with $d$ parameters $\alpha_k$ can be generated by independently generating $d$ variables $x_k$ that are Gamma distributed with density function $p(x_k\,;\alpha_k,1) = x_k^{\alpha_k-1}e^{-x_k}/\Gamma(\alpha_k)$. Then $x := [\,x_1\ \dots\ x_d]^T/\sum_{k=1}^{d} x_k$ is Dirichlet distributed with parameters $\boldsymbol{\alpha}$.

Table 1 shows the result of parameter estimation experiments with different number of samples. The number of iterations means that from that iteration on subsequent alphas are identical up to double length floating point precision. The true parameters are randomly chosen from a uniform distribution of the interval [0.1, 1.1]. Then the samples are generated with the true parameters. The initial estimated parameters are obtained by the method of moments [11]. For a higher number of samples there seems to be higher stability in the values as in the case of 10.000 samples. Besides, the initial estimated values $\boldsymbol{\alpha}^{(\mathbf{0})}$ are notably near the true values. The experiment suggests that the parameter learning is achieved after a few iteration steps and that the curves enter their saturation very quickly.

| | True parameters | Initial estimated parameters | Final estimated parameters |
|---|---|---|---|
| 10000 sampl. | $\alpha_1 = 0.2522$ | $\hat{\alpha}_1 = 0.2541$ | $\hat{\alpha}_1 = 0.2536$ |
| 42 iterations | $\alpha_2 = 0.1084$ | $\hat{\alpha}_2 = 0.1090$ | $\hat{\alpha}_2 = 0.1072$ |
| | $\alpha_3 = 0.1066$ | $\hat{\alpha}_3 = 0.1106$ | $\hat{\alpha}_3 = 0.1083$ |
| 1000 samples | $\alpha_1 = 0.9645$ | $\hat{\alpha}_1 = 0.9729$ | $\hat{\alpha}_1 = 0.9464$ |
| 92 iterations | $\alpha_2 = 1.0173$ | $\hat{\alpha}_2 = 1.0345$ | $\hat{\alpha}_2 = 1.0090$ |
| | $\alpha_3 = 0.2440$ | $\hat{\alpha}_3 = 0.2609$ | $\hat{\alpha}_3 = 0.2518$ |
| 100 samples | $\alpha_1 = 0.5697$ | $\hat{\alpha}_1 = 0.5787$ | $\hat{\alpha}_1 = 0.5828$ |
| 74 iterations | $\alpha_2 = 0.4876$ | $\hat{\alpha}_2 = 0.4930$ | $\hat{\alpha}_2 = 0.5318$ |
| | $\alpha_3 = 0.8536$ | $\hat{\alpha}_3 = 0.9575$ | $\hat{\alpha}_3 = 0.8733$ |

Table 1. Estimated parameters $\hat{\boldsymbol{\alpha}}$ of Dirichlet Distribution with $d = 3$ for different number of samples

### *5.2 Parameter estimation for selected image regions*

We selected rectangular regions with different size of an image like they are typically used in the Split-and-Merge algorithm. For the determination of the Dirichlet parameters the only observable learning criterion is the log-likelihood $L(\mathcal{D}|\boldsymbol{\alpha})$ since we are dealing with an unsupervised case. Fig. 1 shows four selected ROIs of the intensity normalized Lenna-image: Region 1 is representative for a highly textured area, region 3 has approximately two different subregions (eyes and skin) and regions 2 and 4 are homogeneous with a low variance and covariance of the intensities.


Fig. 1. Original and intensity normalized Lenna-image with selected regions of interest.

The results of the parameter learning are shown in table 2. An obvious difference can be observed between the homogeneous regions {2,4} and the textured regions {1,3} relativ to the achieved likelihood, the dimension of the alphas and the number of iterations to reach a saturation. The homogeneous regions are very slowly converging to very high values of the Dirichlet parameters and have a very small likelihood. The textured regions converge fast to moderate alphas and present high values for the likelihood. A theoretical interpretation of these results is outside the scope of this paper and needed a more profound analysis of the Dirichlet probability distribution and the its parameter learning. The learning suggests however that homogeneous regions with small variation of the intensity values are more difficult to model by a Dirichlet probability distribution.

| ROI | Iterations | Bounding Box | | | | Initial estimated parameters | | | Log-Likelihood | Final estimated parameters | | | Log-Likelihood |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $x_{min}$ | $x_{max}$ | $y_{min}$ | $y_{max}$ | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\alpha}_3$ | $L(ROI\|\boldsymbol{\alpha}^{(\mathbf{0})})$ | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\alpha}_3$ | $L(ROI\|\boldsymbol{\alpha}^{(\mathbf{0})})$ |
| 1 | 2344 | 70 | 400 | 150 | 450 | 46.34 | 16.64 | 38.52 | 13280.6 | 36.51 | 12.79 | 30.55 | 13438.3 |
| 2 | 3000 | 270 | 400 | 330 | 500 | 84.7 | 47.98 | 43.8 | -126491.7 | 157.9 | 89.6 | 81.98 | -5188923.2 |
| 3 | 1469 | 246 | 243 | 353 | 287 | 27.73 | 13.04 | 15.67 | 13879.2 | 27.49 | 12.65 | 15.81 | 13906.7 |
| 4 | 15000 | 120 | 2 | 162 | 45 | 869.42 | 404.1 | 394. | -18037157.7 | 1213. | 564.2 | 550.3 | -27094869. |

Table 2. details for the learning of the dirichlet parameters for the selected ROIs of fig. 1. The number of iterations for regions 1 and 3 mean that from that iteration on subsequent alphas are identical up to double length floating point precision. in the case of regions 2 and 4 the number of iterations was chosen such that the log-likelihood reached a visible saturation.

### *5.3 Split-and-Merge Segmentation*

We use the Bhattacharyya distance between Dirichlet parameters of neighboring regions to control the splitting of the quadtree representation and the melting of neighboring regions in the merging phase. Fig. 2 shows a segmentation experiment with two images. We use a 256 pixel wide square image and limit the smallest region to 16 pixel to avoid a statistically irrelevant parameter estimation. For the initial training of the Dirichlet parameters, the similarity between ${\alpha_k}^{(\tau+1)}$ and ${\alpha_k}^{(\tau)}$ is set to $\epsilon = 0.001$ and the maximum number of iterations to $\tau_{max} = 1000$. When two regions are melted the iterative fixed-point estimation [11] is applied 10000 times. The training parameters to determine the Dirichlet parameters of the regions are identical as in the previous experiments.

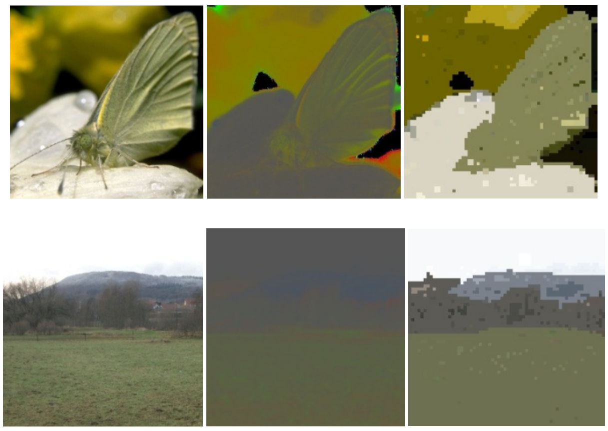

Fig. 2. Segmentation experiments for two 256 × 256 images. Each group shows the original image, the intensity normalized version and the segmentation result.

## 4. CONCLUSION AND FUTURE WORK

As a main result of this paper we present the formulation of the Bhattacharyya probabilistic distance between two Dirichlet distributions. An efficient update method for sequential parameter estimation when merging two data populations is also shown. As an application area we chose Split-And-Merge image segmentation. Other areas of application are possible as long as the data is modeled by a Dirichlet distribution. Mixture Dirichlet [4] modeling is a further possible field of work.


## ACKNOWLEDGMENTS

During the elaboration of the methods described in this paper Thomas W. Rauber was on sabbatical leave under grant 3168-04-2 from CAPES, the Brazilian government agency for the qualification of academic staff and the *Casadinho* research project sponsored by CNPq for financial support.


## REFERENCES


[1] W. K. Pratt. Digital Image Processing: PIKS Inside. John Wiley and Sons, New York, 3rd edition, 2001.

[2] O. D. Faugeras and W. K. Pratt. Decorrelation methods of texture feature extraction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2(4):323–332, 1980.

[3] P. A. Devijver and J. Kittler. Pattern Recognition: A Statistical Approach. Prentice/ Hall Int., London, 1982.

[4] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. Bull. Calcutta Math. Soc., 49:214–224, 1943.

[5] T. Kailath. The Divergence and the Bhattacharyya distance measures in signal selection. IEEE Transactions on Communication Technology, C-15(1):52–60, 1967.

[6] K. Fukunaga. Introduction to Statistical Pattern Recognition. Academic Press, San Diego, 2nd edition, 1990.

[7] 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.

[8] L. Lucchese and S. Mitra. Color image segmentation: A state-of-the- art survey. Proc. of the Indian National Science Academy (INSA-A), 67(2):207–221, 2001.

[9] K. Sjölander, K. Karplus, M. Brown, R. Hughey, A. Krogh, I. Saira Mian, and D. Haussler. Dirichlet mixtures: A method for improving de- tection of weak but signicant protein sequence homology. citeseer. ist.psu.edu/sj96dirichlet.html, 1996.

[10] M. Yamamoto and K. Sadamitsu. Dirichlet mixtures in text modeling. Technical Report CS-TR-05-1, University of Tsukuba, May 2005.

[11] T. P. Minka. Estimating a dirichlet distribution. research. microsoft.com/˜minka/papers/dirichlet, 2003.

[12] C. M. Bishop. Neural Networks for Pattern Recognition. Oxford University Press, USA, 1996.

[13] H. Robbins and S. Monroe. A stochastic approximation method. Annals of Mathematical statistics, 22:400–407, 1951.