

Color Texture Classification Using Rao Distance between Multivariate Copula Based Models

Ahmed Drissi El Maliani, Mohammed El Hassouni, Nour-Eddine Lasmar,
Yannick Berthoumieu, Driss Aboutajdine

► **To cite this version:**

Ahmed Drissi El Maliani, Mohammed El Hassouni, Nour-Eddine Lasmar, Yannick Berthoumieu, Driss Aboutajdine. Color Texture Classification Using Rao Distance between Multivariate Copula Based Models. Pedro Real and Daniel Diaz-Pernil and Helena Molina-Abril and Ainhoa Berciano and Walter Kropatsch. Computer Analysis of Images and Patterns, Springer Berlin Heidelberg, pp.498-505, 2011, 978-3-642-23677-8. <10.1007/978-3-642-23678-5>. <hal-00727122>

HAL Id: hal-00727122

<https://hal.inria.fr/hal-00727122>

Submitted on 2 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Color Texture Classification Using Rao Distance Between Multivariate Copula Based Models

Ahmed Drissi El Maliani¹, Mohammed El Hassouni², Nour-Eddine Lasmar³,
Yannick Berthoumieu³ and Driss Aboutajdine¹

¹ LRIT, URAC 29, Mohammed V University, Agdal, Morocco,

² DESTEC, FLSHR, Mohammed V University, Agdal, Morocco,

³ IMS- Groupe Signal- UMR 5218 CNRS, ENSEIRB, University Bordeaux, France,

Abstract. This paper presents a new similarity measure based on Rao distance for color texture classification or retrieval. Textures are characterized by a joint model of complex wavelet coefficients. This model is based on a Gaussian Copula in order to consider the dependency between color components. Then, a closed form of Rao distance is computed to measure the difference between two Gaussian Copula based probability density functions on the corresponding manifold. Results in term of classification rates, show the effectiveness of the Rao geodesic distance when applied on the manifold of Gaussian Copula based probability distributions, in comparison with the Kullback-Leibler divergence.

Keywords: Color texture classification; Gaussian Copula; Rao distance

1 Introduction

Texture classification is an important and challenging task in image analysis. Efficient texture classification is essentially based on a pertinent feature extraction and similarity measurement steps, especially when choosing the K-Nearest Neighbor approach. The feature extraction step consists of figuring out a set of attributes that best describe the texture, and best discriminate this latter from different textures. Many works, stressed that treating the texture in the wavelet domain allows accurate characterization, by modeling histograms of the wavelet subbands with appropriate models.

In this context, the Generalized Gaussian Density (GGD) proposed by Do and Vetterli [1] showed a good ability in modeling the heavily tailed and pickly pronounced behavior of the histograms. In the case of color textures, and more specifically when they are represented in the RGB color space, a big correlation exists between the color bands, thus considering univariate models leads to a considerable loss of information in the characterization. For this, joint models were proposed to describe the dependence across color bands. Verdoolaege et al. [2], proposed a multivariate Generalized Gaussian distribution (MGGD)

³ This work was supported by franco-moroccan CNRS-CNRST project STIC 05/10

for multiscale color texture retrieval, modeling dependence across color components. A t-Student Copula based multivariate Weibull distribution was proposed by Kwitt et al. [3], describing dependence between complex wavelet coefficient magnitudes also in the context of color texture retrieval. The second step of a KNN based classification process, is the similarity measurement, which consists of measuring distance between textures according to the extracted features. In their pioneering work, Do and Vetterli [1] showed that the Kullback-Leibler (KL) divergence is suitable for comparing textures reposing on the estimated model parameters. Recently KL divergence has been used to measure similarity between Copula based distributions, in the context of texture retrieval [3]. A Monte-carlo approach was adopted to deal with the lack of a closed form of the KL divergence between Copula based models. However, the KL divergence cannot be considered as a metric on the space of probability distributions, since it is not symmetric and does not obey the triangle inequality.

To remedy to this problem, a Riemannian metric namely the Rao or Geodesic distance can be used instead of the commonly used Kullback-Leibler divergence. The Rao distance was first proposed by Rao [4] exploiting that the Fisher information for a set of probability density functions (pdfs) is a Riemannian metric on the corresponding manifold. Thus, the Rao distance is achieved reposing on the Fisher information metric and by resolving the corresponding geodesic equations. Rao distance was already used in different fields, such as segmentation and classification [5][6]. Closed forms of the Fisher-Rao metric were proposed for Extreme Value probability distributions namely the Gumbel, Cauchy-Fréchet, and Weibull [7]. In [2], the Rao distance was used as a similarity measure between zero-mean multivariate generalized Gaussian distributions (MGGD). A closed form of the Rao metric was computed in the case of a fixed shape parameter, while geodesic equations were solved numerically when assuming different shape parameters.

In this work, we propose a Rao metric as a similarity measure on the manifold of Gaussian Copula based multivariate probability distributions. This assumes that we know Rao distance for the marginal distributions, before computing the metric for the joint model. To test effectiveness of the Rao distance based similarity measure, we show results in comparison with the KL divergence based similarity measure in term of classification rates.

This paper is organized as follows. In the next section, we give a review of Copula theory and construct the multivariate models reposing on the Copula approach. In section 3, we derive the general form of the Rao distance on the manifold of Gaussian Copula based probability distributions. In section 4, we present results of texture classification using the Rao distance and the Kullback-Leibler divergence, before concluding in section 5.

2 Gaussian Copula Based Probability Distributions Manifold

2.1 Review of the Copula theory

We draw on the Copula theory to incorporate the information of dependency between color components in the RGB color space. Copulas are an elegant tool for merging a set of marginal pdfs into a multivariate pdf with a particular dependence structure. A Copula is a multivariate cumulative distribution function defined on the d -dimensional unite cube $[0, 1]^d$ [8], with uniform one dimensional marginals. Given a d -dimensional vector $x = [x_1, \dots, x_d]$ on the unit cube $[0, 1]$, with a cumulative distribution function F and marginal cumulative distribution functions (cdf) F_1, \dots, F_d . The multivariate cdf is:

$$F(x_1, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d) \quad (1)$$

Sklar theorem [9] shows that there exists a d -dimensional Copula C such that:

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (2)$$

Further, if C is continuous and differentiable, the Copula density is given by:

$$c(u_1, \dots, u_d) = \frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1 \dots \partial u_d} \quad (3)$$

The joint pdf is uniquely deduced from the margins and the Copula density as follows:

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{i=1}^d f_i(x_i) \quad (4)$$

where $f_i, i = 1, \dots, d$, represent the marginal densities. It appears that the Gaussian Copula is suitable to model linear dependence, which is the most popular in texture modeling. Gaussian Copula density is defined by:

$$c(u, \Sigma) = \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2} \vartheta^T (\Sigma^{-1} - I) \vartheta\right] \quad (5)$$

with $\vartheta_i = \phi^{-1}(F_i(u_i))$, and ϕ represents the standard normal cumulative distribution function. Σ denotes the correlation matrix, and I denotes the d -dimensional matrix identity. From this we can derive the joint model as:

$$f(x, \theta) = \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2} \vartheta^T (\Sigma^{-1} - I) \vartheta\right] \prod_{i=1}^d f_i(x_i) \quad (6)$$

where $\theta = (\eta, \Sigma)$ denotes the hyperparameters of the joint model, where $\eta = (\eta^{(1)}, \eta^{(2)}, \dots, \eta^{(k)})$ represents a set of the marginal parameters and Σ denotes the covariance matrix of the Gaussian vector ϑ .

For estimating parameters of the Gaussian Copula based model we use the IFM (Inference From Margins) method [10]. In a first time, this consists of estimating

the parameters of the marginals using the Maximum Likelihood (ML) procedure. ML estimators $\hat{\eta}_i$ are deduced as:

$$\hat{\eta}_k = \underset{\eta_k}{\operatorname{argmax}} \sum_{i=1}^n \log(f_k(x_{ik}, \eta_k)) \quad (7)$$

Secondly, the log-likelihood function for the joint distribution is maximized using the margins estimators $\hat{\eta}=(\hat{\eta}^{(1)}, \dots, \hat{\eta}^{(k)})$:

$$\hat{\Sigma} = \underset{\Sigma}{\operatorname{argmax}} \sum_{i=1}^n \log c(F_1(x_{1i}; \hat{\eta}_1), \dots, F_d(x_{di}; \hat{\eta}_d); \Sigma) \quad (8)$$

2.2 Multivariate Weibull Distribution

From (6), the pdf of multivariate Weibull (Mwbl) distribution is defined as:

$$f_{Mwbl}(x, \theta) = \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\vartheta^T(\Sigma^{-1} - I)\vartheta\right] \times \left(\frac{\tau}{\lambda}\right)^d \prod_{i=1}^d x_i^{\tau-1} \exp\left[-\sum_{i=1}^d \left(\frac{x_i}{\lambda}\right)^\tau\right] \quad (9)$$

with $\theta = (\tau, \lambda, \Sigma)$, τ represents the shape parameter, λ represents the scale parameter, and Σ denotes the covariance matrix.

2.3 Multivariate Gamma Distribution

The joint pdf of multivariate Gamma (Mgam) distribution is defined as:

$$f_{Mgam}(x, \theta) = \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\vartheta^T(\Sigma^{-1} - I)\vartheta\right] \times \left(\frac{\beta^{-\alpha}}{\Gamma(\alpha)}\right)^d \prod_{i=1}^d x_i^{\alpha-1} \exp\left[-\sum_{i=1}^d \left(\frac{x_i}{\beta}\right)\right] \quad (10)$$

with $\theta = (\alpha, \beta, \Sigma)$, α represents the shape parameter, β represents the scale parameter, and Σ denotes the covariance matrix.

2.4 Multivariate Laplacian Distribution

The pdf of multivariate Laplacian (Mlap) Distribution is defined as:

$$f_{Mlap}(x, \theta) = \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\vartheta^T(\Sigma^{-1} - I)\vartheta\right] \times \frac{1}{b^d} \exp\left[-\sum_{i=1}^d \frac{(x_i - a)}{b}\right] \quad (11)$$

with $\theta = (a, b, \Sigma)$, a represents the location parameter, b represents the scale parameter, and Σ denotes the covariance matrix.

3 Rao Geodesic Distance On The Manifold Of Gaussian Copula Based Distributions

Let us consider M_θ as the statistical manifold of Gaussian Copula based probability distributions, and $f(x; \theta)$ a pdf from this manifold, where θ a vector of parameters of f . Rao distance is a Riemannian metric defined by the fisher information matrix as:

$$ds^2 = \sum_{i,j=1}^d g_{ij}(\theta) d\theta_i d\theta_j \quad (12)$$

where g_{ij} represents the Fisher matrix elements:

$$g_{ij}(\theta) = E \left[\frac{\partial}{\partial \theta^i} \log f(X; \theta) \frac{\partial}{\partial \theta^j} \log f(X; \theta) \right] \quad (13)$$

$$= -E \left[\frac{\partial^2}{\partial \theta^i \partial \theta^j} \log f(X; \theta) \right] \quad (14)$$

Hence, given two probability distributions $f(x; \theta_1)$ and $f(x; \theta_2)$ on M_θ , we can compute the Rao geodesic distance as:

$$L = \int_{\theta_1}^{\theta_2} ds = \int_0^1 \sqrt{\sum_{i,j} g_{ij} \dot{\theta}^i \dot{\theta}^j} dt \quad (15)$$

In M_θ , probability distributions are defined as:

$$f(x, \theta) = c_\Phi(u_1, \dots, u_d; \Sigma) \prod_{i=1}^d f_i(x_i; \eta_i) \quad (16)$$

$$\text{So, } g_{ij}(\theta) = -E \left[\frac{\partial^2}{\partial \theta^i \partial \theta^j} \log c_\Phi(u_1, \dots, u_d; \Sigma) \prod_{i=1}^d f_i(x_i; \eta_i) \right] \quad (17)$$

$$= -E \left[\frac{\partial^2}{\partial \theta^i \partial \theta^j} \log c_\Phi(u_1, \dots, u_d; \Sigma) + \log \prod_{i=1}^d f_i(x_i; \eta_i) \right] \quad (18)$$

The vector (u_1, \dots, u_d) , is constructed by transforming the vector (x_1, \dots, x_d) empirically with known parameters η_i via $u_i = F(x_i; \hat{\eta}_i)$, thus we consider these observations without assumptions on the parametric form of the marginal distributions: $(u_1, \dots, u_d) = (F(x_1, \hat{\eta}_1), \dots, F(x_d, \hat{\eta}_d))$.

This approach has been already considered for estimating the parameters of the copula without estimating the marginals parameters as an alternative of the IFM method, and was named as CML (Canonical Maximum Likelihood) [14]. When this is supposed, we have:

$$\frac{\partial}{\partial \eta} c_\Phi(u; \Sigma) = 0 \quad (19)$$

so then, $g_{\eta\Sigma}(\theta) = g_{\Sigma\eta}(\theta) = 0$.

Thus

$$g_{\Sigma\Sigma}(\theta) = -E \left[\frac{\partial^2}{\partial \Sigma \partial \Sigma} \log c_\Phi(u; \Sigma) \right] \quad (20)$$

and

$$g_{\mu\nu}(\theta) = -E \left[\frac{\partial^2}{\partial \mu \partial \nu} \log \prod_{i=1}^d f_i(x_i; \eta_i) \right] \quad (21)$$

with $\mu, \nu \in \{\eta^{(1)}, \dots, \eta^{(k)}\}$

$$g_{\mu\nu}(\theta) = -E\left[\frac{\partial^2}{\partial\mu\partial\nu}\sum_{i=1}^d\log f_i(x_i;\eta_i)\right] \quad (22)$$

$$= -E\left[\sum_{i=1}^d\frac{\partial^2}{\partial\mu\partial\nu}\log f_i(x_i;\eta_i)\right] \quad (23)$$

$$= \sum_{i=1}^d -E\left[\frac{\partial^2}{\partial\mu\partial\nu}\log f_i(x_i;\eta_i)\right] \quad (24)$$

Thus, from (12):

$$ds^2 = g_{\Sigma\Sigma}d\Sigma d\Sigma + \sum_{i=1}^d\sum_{\mu,\nu}g_{\mu\nu}\dot{\mu}\dot{\nu}dt \quad (25)$$

$$= ds_{Gauss}^2 + \sum_{i=1}^d ds_{Margins}^2 \quad (26)$$

Thus, the Rao distance between two probability density functions $f(x;\theta_1)$ and $f(x;\theta_2)$ on M_θ is defined from equation (15) as follows:

$$L(f(x;\theta_1)||f(x;\theta_2)) = L_{Gauss}(f(x;\Sigma_1)||f(x;\Sigma_2)) + \sum_{i=1}^d L_{Margins}(f(x;\eta_i)||f(x;\eta_2)) \quad (27)$$

and then, the Rao distance between two Mwbl pdfs is:

$$L(f_{Mwbl}(x;\theta_1)||f_{Mwbl}(x;\theta_1)) = \left[\frac{1}{2}\sum_{i=1}^d(\ln r^i)^2\right]^{1/2} + \left(\frac{[\log(\tau_2/\tau_1) - s(\lambda_2 - \lambda_1)^2/\lambda_1^2\lambda_2^2]^2 + q^2(\lambda_2 - \lambda_1)^2/\lambda_1^2\lambda_2^2}{[\log(\tau_2/\tau_1) - s(\lambda_2 - \lambda_1)^2/\lambda_1^2\lambda_2^2]^2 + q^2(\lambda_2 + \lambda_1)^2/\lambda_1^2\lambda_2^2}\right)^{1/2} \quad (28)$$

where $r^i, i = 1, \dots, d$ represents the eigenvalues of $\Sigma_1^{-1}\Sigma_2$, $s = 1 - \gamma$ and $q = \frac{\pi}{\sqrt{6}}$. The Rao distance between two Mlap pdfs is:

$$L(f_{Mlap}(x;\theta_1)||f_{Mlap}(x;\theta_1)) = \left[\frac{1}{2}\sum_{i=1}^d(\ln r^i)^2\right]^{1/2} + \frac{1}{2}\log\frac{\{a_1 - (K + R)\}/\{a_1 - (K - R)\}}{\{a_2 - (K + R)\}/\{a_2 - (K - R)\}} \quad (29)$$

with, $K = \frac{1}{2}\frac{b_2^2 - b_1^2}{a_2^2 - a_1^2} + a_1 + a_2$, and $R^2 = \frac{b_1^2 + b_2^2}{2} + \frac{(a_2 - a_1)^2}{4} + \frac{(b_2^2 - b_1^2)^2}{4(a_2 - a_1)^2}$.

Sometimes, the geodesic equations of the marginal probability distribution are difficult to solve as the case of the Gamma distribution. In such cases, we proceed by numerical approximations for the geodesic equations [13].

4 Experimental Results

Experiments are conducted on the MIT vision texture database (Vistex) [11]. We consider 30 texture classes as used in [12]. Each 512×512 texture image is splitted into sixteen subimages of size 128×128 resulting on a database of 480 samples. Then color bands of each subimage are decomposed via the Dual Tree Complex Wavelet Transform (DTCWT). DTCWT presents advantages over the classic DWT (Discret Wavelet Transform), in terms of translation invariance and directional selectivity, since DTCWT provides complex subbands in six orientations at each decomposition level. Every color band (R, G and B) is then decomposed via DTCWT resulting on six subbands $r_i, g_i, b_i, i \in \{1, \dots, 6\}$ for each color component (considering only subbands of the second decomposition level). Thus, the dataset to be modeled is constructed by arranging the absolute values of the subbands coefficients in an $n \times 18$ matrix, where n is the number of subband coefficients.

As said in the introduction we choose the K-Nearest Neighbor approach for the classification purpose. In the KNN approach, an instance is classified reposing on a similarity measure, and is accorded the label of the majority of its K-Nearest Neighbors.

We begin the classification process by dividing our dataset into training and testing sets. From each class of textures, we choose, randomly, Q samples to construct the training set. The rest of the samples are then considered as the testing set. This experiment is repeated 100 times before returning the average classification rate. In our experiments, we vary Q from 2 to 8 in order to test performances for different numbers of training samples. Table 1 shows the average

Table 1. Average classification rate for different number of training set.

Q	2	3	4	5	6	7	8
Mwbl	86.75	86.90	91.72	92.14	94.25	95	95.96
Mgam	84.83	85.74	91.76	91.93	92.70	93.62	95.33
Mwbl+KL [3]	85	86.14	90.65	91.33	93.75	94.81	95.45
MGGD [2]	84.54	84.98	88.42	90.14	92.49	92.95	94.56

classification rates for two submanifolds of M_θ namely Mwbl and Mgam with Rao distance as a similarity measure in comparison with the MGGD model [2], also with the same similarity measure, and then with the approach proposed in [3], which uses the Monte-carlo based KL divergence with Multivariate Weibull distribution. We observe that for Mwbl and Mgam higher classification rates are achieved for $Q = 8$, with *95.96%* for Mwbl and *95.33%* for Mgam, which is better than the rates achieved when using MGGD (*94.56%* for $Q = 8$). This is due the flexibility of the Gaussian Copula based models in describing the information of dependence. Also, it can be seen that, in comparison with the KL divergence based approach, results with the Rao distance are slightly better. These little improvements in term of classification rates are valuable, since one can replace the use of the Kullback-Leibler divergence by the Rao distance, and then benefits the advantages of this latter over the KL divergence, as being a distance in the

right sense of the word, respecting properties of symmetry and triangularity. Another advantage of the Rao distance, is that for multivariate distributions, and especially those based on copulas, the KL divergence is numerically computed using the Monte carlo method, known as being computationally expensive.

5 Conclusions

In this work, we have proposed to use the Rao distance as a similarity measure between Gaussian Copula based multivariate distributions. We derived a closed form for the Rao distance of these joint distributions when the Rao distance of the marginal distributions is known. Results in term of classification show the Rao distance can replace the use of the Kullback-Leibler divergence since rates are improved, and then benefit the advantages of the Rao distance.

References

1. M. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE transactions on image processing*, vol. 11, pp. 146–158, 2002.
2. G. Verdoolaege, S. De Backer, and P. Scheunders, "Multiscale colour texture retrieval using the geodesic distance between multivariate Generalized Gaussian models," in *Proceedings of the 15th IEEE International Conference On Image Processing (ICIP'08)*, San Diego, California, USA, 2008, pp. 169–172.
3. R. Kwitt and A. Uhl, "A joint model of complex wavelet coefficients for texture retrieval," in *Image Processing, 2009. ICIP 2009. 16th IEEE International Conference*, 2009, pp. 1877–1880.
4. C. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," in *Bull. Calcutta Math. Soc.*, vol. 37, pp. 81–89, 1945.
5. J. Pastore, E. Moler, and V. Ballarin, "Segmentation of brain magnetic resonance images through morphological operators and geodesic distance," in *Digital Signal Processing*, vol. 15, no. 2, pp. 153–160, 2005.
6. Q. Yong and Y. Jie, "Modified kernel functions by geodesic distance," *EURASIP Journal on Applied Signal Processing*, vol. 16, pp. 2515–2521, 2004.
7. J. M. Oller, "Information Metric for Extreme Value and Logistic Probability Distributions," *The Indian Journal of Statistics, Series A*, vol. 49, pp. 17–23, 1987.
8. R. B. Nelsen, *An Introduction to Copulas*. Springer Series in Statistics. Springer, second edition, 2006.
9. M. Sklar, "Fonctions de répartition à n dimensions et leurs marges,". *Publications de l'institut de Statistique de l'Université de Paris*. vol. 8, pp. 229–231, 1959.
10. H. Joe, *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability. Chapman & Hall, 1997.
11. "MIT vision and modeling group," [Online], Available from: <http://vismod.media.mit.edu>.
12. G. Van De Wouwer, P. Scheunders and D. Van Dyck, "Statistical Texture Characterization From Discrete Wavelet Representations," *IEEE transactions on image processing*, vol. 8, pp. 592 - 598, 1999.
13. F. Reverter and J. M. Oller, "Computing the Rao distance for Gamma distributions," *Journal of Computational and Applied Mathematics*, vol. 157, pp. 155–167, 2003.
14. V. Durrleman, A. N. Nikoghal, and T. Roncalli, "Which copula is the right one," *Groupe de Recherche Operationnelle, Credit Lyonnais*, 2000.