

# Geodesics on the Manifold of Multivariate Generalized Gaussian Distributions With an Application to Multicomponent Texture Discrimination

Geert Verdoolaege · Paul Scheunders

Received: date / Accepted: date

**Abstract** We consider the Rao geodesic distance (GD) based on the Fisher information as a similarity measure on the manifold of zero-mean multivariate generalized Gaussian distributions (MGGD). The MGGD is shown to be an adequate model for the heavy-tailed wavelet statistics in multicomponent images, such as color or multispectral images. We discuss the estimation of MGGD parameters using various methods. We apply the GD between MGGDs to color texture discrimination in several classification experiments, taking into account the correlation structure between the spectral bands in the wavelet domain. We compare the performance, both in terms of texture discrimination capability and computational load, of the GD and the Kullback-Leibler divergence (KLD). Likewise, both uni- and multivariate generalized Gaussian models are evaluated, characterized by a fixed or a variable shape parameter. The modeling of the interband correlation significantly improves classification efficiency, while the GD is shown to consistently outperform the KLD as a similarity measure.

**Keywords** geodesic distance · multivariate generalized Gaussian distribution · texture discrimination · multicomponent images

**Mathematics Subject Classification (2000)** MSC 62H35 · MSC 62H12 · MSC 68T10 · MSC 53C22

---

G. Verdoolaege  
Department of Applied Physics, Ghent University, J. Plateaustraat 22,  
9000 Gent, Belgium  
Tel.: +32-9-2643831  
Fax: +32-9-2644198  
E-mail: geert.verdoolaege@ugent.be

P. Scheunders  
IBBT-Visionlab, Department of Physics, University of Antwerp, Uni-  
versiteitsplein 1, 2610 Wilrijk, Belgium

## 1 Introduction

The visual description of texture has been a major research topic during more than twenty years and still is a matter of intensive ongoing research. An important challenge is the automated discrimination of textured images or regions. Texture discrimination techniques form a significant part of many classification and machine learning frameworks, with applications in medical imaging, remote sensing and numerous other practical domains. The present work deals with the discrimination and classification of multicomponent textured images, with a specific application in content-based image retrieval.

The advent of the Internet and extensive digital image libraries has entailed the development of rapid and efficient computer-based image searching and browsing techniques. The term content-based image retrieval (CBIR) (for a review, see Datta et al, 2008) refers to the automatic retrieval of images from a database based on a set of graphic features that qualify the images and that are, loosely speaking, similar to the characteristics of a given query image. The goal is to match the applied notion of similarity as well as possible with the human perception of image resemblance. The essence of the task of CBIR is captured by two main steps. One is the process of *feature extraction*, involving the generation of a set of features (the image *signature*) that characterize the image information accurately and concisely. The latter is an important quality for guaranteeing the computational efficiency of the retrieval system and for storage and transmission of images. In this work, texture and color information is used collectively for deriving image signatures. The second main step in CBIR, called *similarity measurement*, concerns the calculation of a distance function measuring the similarity of two images based on their respective set of features. This allows the ranking of the images in the database according to their similarity to the query image.

Naturally, the choice of feature set and similarity measure largely determines the performance of the retrieval system. Moreover, since querying an image database is often an on-line activity, the evaluation of the distance measure should be sufficiently fast. In practice, this typically means that a closed-form expression for the distance function should be used.

In many existing texture discrimination schemes the signatures are obtained from the images after some suitable linear transform. The rationale is the observation that the image information in the transformed domain is often simpler to model. For instance, the discrete wavelet transform applied to an image results in a set of multiscale oriented subbands that are sensitive to horizontal, vertical and diagonal edges in the original image (Mallat, 1999). Several approaches to texture characterization in the wavelet domain assume that the wavelet representation accurately characterizes texture (see e.g. Manjunath and Ma, 1996; Van de Wouwer et al, 1999). In addition, physiological studies of the visual cortex suggest that the wavelet decomposition is a natural way of structuring the image information (Daugman, 1980). The foregoing arguments provide the motivation to conduct our work in the wavelet domain.

We will work within a parametric probabilistic framework, modeling the wavelet detail coefficients through appropriate probability distributions. This yields a compact description of the texture information in the images, preventing the entire image from having to be stored or transmitted. Now, the wavelet transforms of real-world images tend to be sparse, yielding a large number of small wavelet detail coefficients and a small number of large coefficients (Mallat, 1989). Therefore, the wavelet histograms can be characterized through unimodal probability distributions. In addition, the distributions have zero skewness (symmetry around the mean) and have mean zero because the wavelet detail coefficients are at the output of a high-pass filtering process of the original image. When considering single-valued images and assuming the wavelet coefficients from different subbands to be independent, univariate distributions can be used. In this case, the free parameters of the distributions for all subbands serve as image features. The sparseness of the wavelet coefficients prevents the Gaussian distribution to provide a satisfactory fit to the respective histograms in every single case. Indeed, often a distribution is needed marked by a more peaky and heavy-tailed behavior compared to the Gaussian distribution, i.e. a leptokurtic distribution. The *generalized Gaussian distribution* is frequently used in this setting (see Mallat, 1989; Van de Wouwer et al, 1999; Do and Vetterli, 2002), but some other possibilities are the Gaussian Scale Mixture (Scheunders and De Backer, 2007), the alpha-stable distributions (Tzagkarakis et al, 2006) and the Student-t distribution (Synyavskyy et al, 2001).

The intention of this work is to include both texture and color information in a single probabilistic model. The modeling of color features is an active subject in the field of texture discrimination. Color histograms have been used extensively for color characterization, within various color spaces. An overview of several color descriptors that have proved to lend themselves for image retrieval purposes was presented by Manjunath et al (2001). However, in addition we want to make use of the information residing in the rich correlation structure between the color bands, since we expect that the classification and retrieval tasks will benefit from these additional data. To this end, in a probabilistic framework we model the spectral bands jointly through a multivariate probability distribution. At the same time, we assume independence among the wavelet subbands belonging to a single color component. Moreover, our model is not restricted to color images, but can handle multicomponent images in general. For example, the techniques for multicomponent texture similarity measurement developed in this work, can also be used in applications involving multispectral and hyperspectral imagery, which is used increasingly in remote sensing of the earth's surface (Chang, 2007; Mercier and Lennon, 2002).

The joint modeling of multivalued wavelet images has been considered before. For instance, although in this work we neglect any existing correlation between wavelet subbands, other approaches have allowed interdependencies across subbands in a joint Gaussian model, e.g. in a denoising application (Micak et al, 1999), while Tzagkarakis et al (2006) fit a joint alpha-stable sub-Gaussian distribution to the wavelet coefficients in order to better capture the tail statistics in a retrieval experiment. In the latter work, with the aid of the KLD, good retrieval rates were obtained, but a computationally complex gaussianization step was required. In addition, the process of wavelet-based multivariate image denoising has been treated recently by several authors (Benazza-Benyahia and Pesquet, 2005; Pižurica and Philips, 2006; De Backer et al, 2008). In these works, multivariate probability density functions of the images were proposed that account for the correlations between the image bands. Particularly heavy-tailed models were found to be efficient. These models were applied as priors in a Bayesian framework.

In our work, for the joint modeling of multicomponent wavelet images we use a particular case of the multivariate Kotz-type distribution, which we call the (*multivariate*) *generalized Gaussian distribution ((M)GGD)*, also sometimes called the multivariate exponential power distribution. Other possible extensions of the univariate generalized Gaussian density towards multiple dimensions were given by Cho and Bui (2005) and Boubchir and Fadili (2005), but the definition that we use has the advantage that it is relatively well documented (see e.g. Gómez et al, 1998; Fang et al,

1990). The MGGD probabilistic model is characterized by a sufficient number of degrees of freedom to suitably handle both the interband correlation structure as well as the heavy-tailedness of the wavelet distributions. Thus, an MGGD serves as an excellent descriptor of texture information, since the tails of the wavelet distribution contain most of the relevant information on the occurrence of edges in an image, characterizing the texture over multiple scales. Furthermore, we present a fairly detailed discussion of several procedures for estimating the parameters of the best-fit MGGD from wavelet data. We treat both the method of moments and the maximum likelihood (ML) method, the latter implemented by recursive solution of the ML equations on the one hand, and on the other hand through the Fisher scoring algorithm. We also consider the measurement of the goodness of the MGGD fit.

Just as several candidate probabilistic models exist for describing wavelet detail statistics, there are also many possible choices for the distance function between textures. In the probabilistic setting, distance measures between probability distributions are required. It turns out that the Euclidean distance as a first common choice is not a natural similarity measure between probability distributions, although its application in a CBIR context may yield acceptable retrieval performances (Do and Vetterli, 2002). In addition, good retrieval performances for gray-level images have been reported using the Kullback-Leibler divergence (KLD) between generalized Gaussian densities (Do and Vetterli, 2002; Mathiassen et al, 2002; Bonet and Viola, 1998). In the experimentation section, Section 4, we will compare with the retrieval results obtained by Do and Vetterli (2002).

In this work, as a similarity function we use the Rao geodesic distance (GD) derived from the Fisher information. The GD between probability distributions is defined in the context of information geometry, which expresses probability theory in terms of geometric concepts (see e.g. Amari and Nagaoka, 2000; Murray and Rice, 1993; Kass and Vos, 1997). Cramér (1946) and Rao (1945) observed that the Fisher information can be regarded as a Riemannian metric on a manifold of probability distributions. Čenkov (1982) showed that this Fisher-Rao metric is the unique intrinsic metric on such a manifold, invariant under some basic probabilistic transformations. Thus, probability theory can be described in terms of geometric structures invariant under coordinate transformations, to which the methods of differential geometry can be applied. For an introduction to differential geometry and Riemannian geometry the reader may refer to O’Neill (1982) and do Carmo and Flaherty (1992). The corresponding geodesics between probability distributions have a property of length minimization; they are the ‘straight lines’ of the geometry. The GD is a natural, intrinsic distance measure on the manifold of probability distributions.

Several authors have studied the Fisher-Rao metric and the geodesics for the multivariate normal model. Relevant publications include those by Burbea (1986), Burbea and Rao (1982), James (1973), Skovgaard (1981) and Skovgaard (1984). The GD between multivariate normal distributions with fixed mean and differing covariance matrices was first derived by Jensen<sup>1</sup> and independently by Skovgaard (1984). Mitchell (1989) derived the Fisher-Rao metric for general multivariate elliptical distributions, with the multivariate normal distribution as a special case. A closed expression for the associated GD, in the case of elliptical distributions differing only in their dispersion matrix, was obtained by James (1973) and Berkane et al (1997) (see also Calvo and Oller, 2002). We have derived a closed-form expression for the GD between MGGDs in the case of a fixed MGGD shape parameter and have proposed a suitable approximation to the geodesics on the manifold of MGGDs with varying shape parameters<sup>2</sup>. Although the latter approximation does not result in an analytic form for the GD, as we will show, it still allows to compute the distance relatively fast. It should be noted that the GD and KLD, as similarity measures between probability distributions, are useful in any texture discrimination application employing a statistical description of texture.

The GD as a similarity measure between probability distributions has several important advantages compared to, e.g. the KLD. First, since a geodesic is a geometric object with a rather clear intuitive meaning, it becomes possible to convey a measure of distance between points on the probabilistic manifold (i.e. between distributions), through visualization of the geodesic path by plotting its coordinate functions. In turn, this allows straightforward evaluation of a geodesic approximation strategy, in contrast with the KLD, which cannot be visualized in a similar way. Therefore, if computational demands of the application are an issue, as in the case of e.g. image retrieval, a trade-off between accuracy and speed of the similarity measurement can be readily made in the numerical calculations. In this work, for the case of variable MGGD shape parameters, we apply a linear approximation to the geodesic coordinate functions and show that this still yields good classification results, while maintaining a reasonable computational load. In addition, the visualization of the geodesics, possibly supplemented with the calculation of geometric quantities such as the curvature, allows one to develop a useful geometric intuition of the MGGD manifold.

Second, the GD is, in contrast to the KLD, a genuine distance measure, symmetric in its arguments and obeying

<sup>1</sup> S.T. Jensen, private communication in (Atkinson and Mitchell, 1981), 1976.

<sup>2</sup> G. Verdoolaege and P. Scheunders, On the Geometry of Multivariate Generalized Gaussian Models, submitted to Journal of Mathematical Imaging and Vision, 2010.

the triangle inequality. The latter is a useful property, not shared by the KLD, for significantly reducing the computational demands in an image retrieval application, by comparing a query image to a predefined set of key images that effectively summarize the visual characteristics of all images in the database (Berman and Shapiro, 1999; Burkhard and Keller, 1973). Indeed, suppose  $I_q$  represents a query image,  $I_d$  a database image and  $I_{k_i}$  the key images ( $i = 1, \dots, K$ ), then the triangle inequality requires

$$\text{GD}(I_q||I_d) \geq \max_{1 \leq i \leq K} |\text{GD}(I_q||I_{k_i}) - \text{GD}(I_d||I_{k_i})|, \quad (1)$$

where  $\text{GD}(\cdot||\cdot)$  is a shorthand notation for the GD between two images, based on their respective probabilistic model. This yields a lower bound on the distance  $\text{GD}(I_q||I_d)$ . The distances  $\text{GD}(I_d||I_{k_i})$  can be precomputed for all database and key images. Now, if, for example, we wish to find all images  $I_d$  in the database that are closer to the query image (in terms of GD) than a certain threshold  $T$ , i.e.  $\text{GD}(I_q||I_d) \leq T$ , all we have to do is calculate the distances between the query image and each of the key images, apply (1) and reject the database images for which the lower bound on  $\text{GD}(I_q||I_d)$  is larger than  $T$ . Optionally, we can then conduct a further query among the reduced set of database images. The bottom line is that this way the triangle inequality permits to considerably decrease the number of required evaluations of the similarity measure at the time of the query submission.

Third, calculating the KLD involves solving a multidimensional integral over the data space. For this reason closed-form expressions for the KLD are difficult to find or may not exist at all. So far we did not obtain an analytic expression for the KLD between two MGGDs in general. One may attempt a numerical evaluation of the KLD integral, but it is clear that this drastically increases computational load and, in practice, renders e.g. a retrieval system impractical to be used in real-time. In contrast, as we will show, the above mentioned analytic expression for the GD and, in the case of variable shape parameter, the linear approximation to the geodesic coordinate functions, permit a practicable, relatively fast evaluation of the GD.

Finally, recent observations suggest that the GD is a more natural and more accurate similarity measure between probability distributions, compared to the KLD (Verdoolaege et al, 2008; Lenglet et al, 2006a,b; Castano-Moraga et al, 2007). This will be confirmed by the experiments outlined in the present work, where the (approximated) GD exhibits good results in a texture classification experiment, superior to those obtained via the KLD.

In this paper we illustrate the performance of both the generalized Gaussian model and the geodesic distance, as compared to the Gaussian distribution and the KLD, in several classification and retrieval experiments on textured color images. We start with the discussion of a simple experiment for texture classification on a small texture database and we

next conduct the same experiment in a texture retrieval context. We use  $k$ -nearest neighbor classification and we validate the classifier via the leave-one-out method. We gradually build up the complexity of the model for the wavelet statistics, first using univariate distributions on gray-level textures, obtained from the luminance of the RGB color images. Next we consider the three color bands in parallel, without modeling the interband correlation. We then demonstrate the performance of both the GD and the KLD in an experiment where the correlation between the three color bands of an image is modeled pairwise by a bivariate distribution. We recently obtained a closed expression for the KLD between bivariate GGDs (Verdoolaege et al, 2009), which we put into practice here. We proceed with the full joint modeling of the wavelet coefficients corresponding to the three color bands and we show that the information in the interband correlation structure leads to a significant increase in texture discrimination performance. In addition, we demonstrate the superior capability of a heavy-tailed distribution to adequately describe the image wavelet statistics. We compare the heavy-tailed Laplace distribution with the Gaussian model, both MGGDs with a fixed shape parameter. In turn we show that the enhanced flexibility of the MGGD model with its variable shape parameter results in several situations in still better classification performance. Finally, the GD is shown to outperform the KLD in terms of correct classification rate. The computational load of each of the methods is evaluated as well. The classification experiments are then repeated in the context of image retrieval, showing similar trends as for classification. We next show that our techniques are competitive with a state of the art method for gray-scale texture classification based on image patch exemplars. We examine the impact of corruption of the images by additive Gaussian noise as well as the influence of the color space that is used (RGB or HSV). Finally, we demonstrate our methods in a more realistic experimental setting for image retrieval. We use a considerably larger database than in the first series of experiments, the images consisting of multiple patches of several colored textures. Again we compare different statistical models and similarity measures.

This paper is organized as follows. In Section 2 we introduce the class of multivariate generalized Gaussian distributions and we discuss several methods to estimate the parameters. Section 3 summarizes the techniques that allow the calculation of the GD and KLD between MGGDs, while Section 4 presents the results of several classification and retrieval experiments on gray-level and color images. We also provide an outlook toward possible improvements of the methods presented in this work and suggest some topics for future research. Finally, Section 5 concludes the paper. The main contributions of this paper lie in the definition of the MGGD, its application to the modeling of multivari-

ate wavelet histograms, the discussion on model fitting and goodness of fit testing, the approximation of GDs between MGGDs with variable shape parameter and the application of the MGGD and the GD to texture classification.

## 2 Modeling of wavelet statistics through a (multivariate) generalized Gaussian distribution

We start by introducing the distribution that we use to model wavelet statistics, namely the (multivariate) generalized Gaussian distribution. We proceed with a discussion on the estimation of the MGGD parameters from a wavelet data set, treating also the goodness of the resulting fit.

### 2.1 The multivariate generalized Gaussian distribution

With a view to modeling wavelet detail coefficients we consider only zero-mean distributions. For the definition of a multivariate GGD, we seek inspiration with the univariate zero-mean GGD, which has the following density function:

$$f(x|\alpha, \beta) = \frac{\beta}{2\alpha\Gamma[1/\beta]} \exp\left[-(|x|/\alpha)^\beta\right],$$

where  $\Gamma(\cdot)$  denotes the Gamma function. Clearly,  $\alpha$  is a scale parameter, playing the role of a variance that determines the ‘width’ of the PDF, while  $\beta > 0$ , called the *shape parameter*, controls the fall-off rate in the vicinity of the mode (the higher  $\beta$ , the lower the fall-off rate). Note that  $\beta = 2$  results in the Gaussian distribution and  $\beta = 1$  yields the Laplacian PDF. There does not appear to exist a generally agreed upon multivariate extension of the univariate generalized Gaussian distribution. However, here we define a multivariate generalized Gaussian distribution through the following density function:

$$f(\mathbf{x}|\Sigma, \beta) = \frac{\Gamma\left(\frac{m}{2}\right)}{\pi^{\frac{m}{2}}\Gamma\left(\frac{m}{2\beta}\right)2^{\frac{m}{2\beta}}|\Sigma|^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2}[\mathbf{x}'\Sigma^{-1}\mathbf{x}]^\beta\right\}, \quad (2)$$

where  $m$  is the dimensionality of the probability space, e.g.  $m = 3$  for three-band color images. This is a particular case of the multivariate Kotz-type distribution (see Gómez et al, 1998 and Fang et al, 1990 for details) and it is also sometimes called the *multivariate exponential power distribution*. Again, we call  $\beta$  the shape parameter, which controls the peakedness of the distribution and the heaviness of its tails. Note however that now the Gaussian PDF is retrieved by setting  $\beta = 1$ , while by analogy with the univariate case, we will call the distribution with  $\beta = 1/2$  the multivariate Laplace distribution.  $\Sigma$  is a dispersion matrix, equal to the distribution covariance only in the Gaussian case. If  $\beta < 1$ ,

the distribution (2) has heavier tails (leptokurtic) compared to the multivariate Gaussian distribution. As an example, Fig. 1 shows the histogram for the wavelet coefficients corresponding to a real world gray-scale image from the MIT Vision Texture (VisTex) database (MIT Vision and Modeling Group, 2010) (used in our experiments, see Section 4), together with the fit via maximum likelihood of a Gaussian and a generalized Gaussian (see also the next subsection). The latter clearly describes best the peakedness and the heavy tails of the histogram.

### 2.2 MGGD parameter estimation

The estimation of the mean, dispersion and shape parameter of a *univariate* generalized Gaussian distribution was treated by Varanasi and Aazhang (1989). Three parameter estimation methods were compared in that work, viz the method of moments (MM), the maximum likelihood method (ML) and a moment method followed by an optimization through a single Fisher scoring (Newton) step (MM-Fisher). It was observed that for small values of the shape parameter  $\beta$  (strongly leptokurtic distributions) the ML method performs best, while as  $\beta$  increases, both the moment method and the MM-Fisher technique perform increasingly well compared to the ML method. For large values of the shape parameter ( $\beta > 1$ ), the MM-Fisher algorithm yields the best results. We now proceed with an investigation of the three above mentioned methods for the case of the estimation of the multivariate generalized Gaussian model.

#### 2.2.1 Method of moments

In the multivariate setting the parameters of a zero-mean MGGD to be estimated from a sample of  $n$   $m$ -dimensional vectors  $\mathbf{x}_i$ ,  $i = 1, \dots, n$ , are (the non-redundant elements of) the dispersion matrix  $\Sigma$  and the shape parameter  $\beta$ . In the present work the moment method was implemented by equating the population variance and kurtosis of an MGGD distributed random vector  $\mathbf{X}$  to the sample variance and sample kurtosis, respectively. The population variance and kurtosis (according to Mardia et al (1982)) are given by

$$\begin{aligned} \text{Var}(\mathbf{X}) &= \frac{2^{1/\beta}\Gamma\left(\frac{m+2}{2\beta}\right)}{m\Gamma\left(\frac{m}{2\beta}\right)}\Sigma, \\ \gamma_2(\mathbf{X}) &= \frac{m^2\Gamma\left(\frac{m}{2\beta}\right)\Gamma\left(\frac{m+4}{2\beta}\right)}{\left[\Gamma\left(\frac{m+2}{2\beta}\right)\right]^2} - m(m+2), \end{aligned} \quad (3)$$

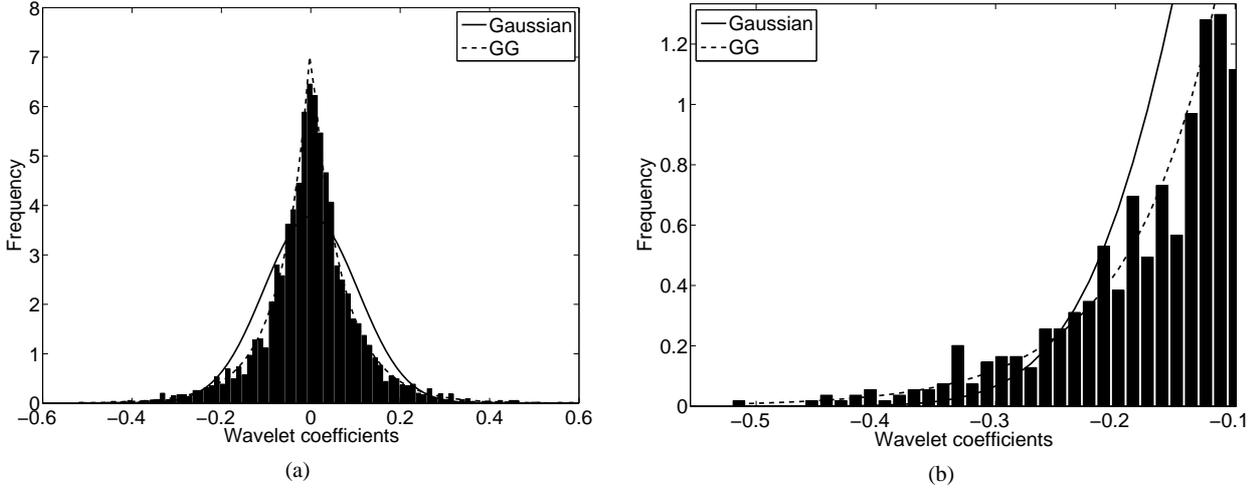


Fig. 1: (a) Histogram for the wavelet coefficients for a real world VisTex gray-scale image, together with the best-fit Gaussian and GGD ( $\beta = 0.48$ ) obtained via ML. (b) Zoomed display of (a), clearly showing the better fit of the tails by the GGD compared to the Gaussian fit.

whereas estimators of the variance and kurtosis from the sample  $\mathbf{x}_i$ ,  $i = 1, \dots, n$  are calculated as

$$\widehat{\text{Var}}(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i',$$

$$\widehat{\gamma}_2(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \left\{ \mathbf{x}_i' \left[ \widehat{\text{Var}}(\mathbf{X}) \right]^{-1} \mathbf{x}_i \right\}^2 - m(m+2).$$

This procedure results in a nonlinear equation for  $\beta$ , which can be solved numerically. Using this result together with the sample variance directly allows to calculate  $\Sigma$  via (3).

In practice, the estimates found using the MM were used as an initialization for both the ML and Fisher methods, resulting in MM-ML and MM-Fisher methods, respectively.

### 2.2.2 Maximum likelihood

The ML method for estimating the parameters of an MGGD proceeds by setting the differential to zero of the logarithm of  $f$  in (2). This results in the following equations for  $\Sigma$  and  $\beta$ , respectively:

$$\Sigma = \frac{\beta}{n} \sum_{i=1}^n u_i^{\beta-1} \mathbf{x}_i \cdot \mathbf{x}_i', \quad (4)$$

$$\sum_{i=1}^n \left\{ \frac{\beta}{2} \ln(u_i) u_i^\beta - \frac{m}{2\beta} \left[ \ln(2) + \Psi \left( \frac{m}{2\beta} \right) \right] - 1 \right\} = 0.$$

Here,  $u_i \equiv \mathbf{x}_i' \Sigma^{-1} \mathbf{x}_i$  and  $\Psi(\cdot)$  denotes the digamma function. These equations can be solved recursively, although we found convergence to be rather slow, especially when no proper initialization using the moment method was applied.

### 2.2.3 Fisher scoring

The Fisher scoring method is in principle a Newton-Raphson numerical optimization algorithm for maximizing the likelihood by searching for a zero of the gradient of the log-likelihood  $L$  (see e.g. Lehmann and Casella, 2003). However, instead of using the negative Hessian of the log-likelihood (observed information matrix) for the optimization, the expected negative Hessian, or Fisher information matrix, is applied. The Fisher information matrix  $g_{\mu\nu}$  is defined through the relations

$$g_{\mu\nu}(\theta) = -\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^\mu \partial \theta^\nu} \ln f(\mathbf{X}|\theta) \right], \quad \mu, \nu = 1 \dots N,$$

with  $N$  the number of non-redundant parameters, arranged in a vector  $\theta$ , parametrizing the distribution. Given a starting value of the parameters, arranged in an  $N$ -dimensional vector  $\theta_k$ , the next approximation  $\theta_{k+1}$  can be found by solving the following equations:

$$n \sum_{v=1}^N (\theta_{k+1}^v - \theta_k^v) g_{\mu\nu}(\theta_k) = \frac{\partial L(\theta_k)}{\partial \theta^\mu}, \quad (5)$$

with  $\theta_k^\mu$  denoting component  $\mu$  of the vector  $\theta_k$ . Since this is a linear set of equations, the solution can be obtained very fast. The disadvantage is the possibility for the algorithm to leave the allowed part of parameter space, returning e.g. a negative definite dispersion matrix. In the case of MGGD estimation, this was observed in our experiments to occur primarily for small shape parameters (practically  $\beta < 0.5$ ).

Since  $\Sigma$  is symmetric and hence contains  $m(m+1)/2$  non-redundant elements, the corresponding Fisher information is a matrix of size  $N = 1 + m(m+1)/2$ . In Section

3.1 we discuss the calculation of geodesic curves between MGGDs, deduced from the Fisher-Rao metric tensor on the  $N$ -dimensional MGGD manifold. We have derived expressions for the entries of the MGGD Fisher matrix and the details of this calculation will be published elsewhere<sup>3</sup>. The components can be obtained by calculating the line element, defined by

$$ds^2 = \sum_{\mu, \nu} g_{\mu\nu} d\theta^\mu d\theta^\nu.$$

This results in the following expressions in terms of the differential forms  $d\beta$  and  $d\Sigma$ :

$$\begin{aligned} \beta\beta\text{-component: } & \frac{1}{\beta^2} \left\{ 1 + \left(\frac{m}{2\beta}\right)^2 \Psi_1\left(\frac{m}{2\beta}\right) \right. \\ & + \frac{m}{\beta} \left[ \ln(2) + \Psi\left(\frac{m}{2\beta}\right) \right] + \frac{m}{2\beta} \left[ \ln(2) \right]^2 \\ & + \Psi\left(1 + \frac{m}{2\beta}\right) \left[ \ln(4) + \Psi\left(1 + \frac{m}{2\beta}\right) \right] \\ & \left. + \Psi_1\left(1 + \frac{m}{2\beta}\right) \right\} d\beta^2, \quad (6) \end{aligned}$$

$$\beta\Sigma\text{-component: } -\frac{1}{\beta} \left[ 1 + \ln(2) + \Psi\left(1 + \frac{m}{2\beta}\right) \right]$$

$$\times \text{tr}(\Sigma^{-1} d\Sigma) d\beta,$$

$$\Sigma\Sigma\text{-component: } 2b_h \text{tr}(\Sigma^{-1} d\Sigma \Sigma^{-1} d\Sigma)$$

$$+ \left(b_h - \frac{1}{4}\right) [\text{tr}(\Sigma^{-1} d\Sigma)]^2.$$

In the expression for the  $\beta\beta$ -component,  $\Psi_1(\cdot)$  is the trigamma function. In order to derive the equations (5), we need to express the information metric into the differential forms  $d\beta$  and  $d\Sigma_{ij}$ , with  $\Sigma_{ij}$  the  $(i, j)$ -th non-redundant (i.e.  $i \leq j$ ) element of  $\Sigma$ . Introducing, for all  $i$  and  $j$ , the matrix  $E_{(i,j)}$  by

$$E_{(i,j)} = \begin{cases} \bar{E}_{(i,i)} & i = j \\ \bar{E}_{(i,j)} + \bar{E}_{(j,i)} & i \neq j \end{cases},$$

where  $\bar{E}_{(i,j)}$  denotes the  $m \times m$  matrix with the  $(i, j)$ -th entry 1 and 0 elsewhere, we obtain the  $\beta\Sigma_{ij}$  components ( $i \leq j$ ) of the Fisher information:

$$-\frac{1}{2\beta} \text{tr}[\Sigma^{-1} E_{(i,j)}] \left[ 1 + \ln(2) + \Psi\left(1 + \frac{m}{2\beta}\right) \right], \quad i \leq j.$$

Similarly, the  $\Sigma_{ij}\Sigma_{rs}$  components ( $i \leq j, r \leq s$ ) can be derived as

$$\begin{aligned} & 2b_h \text{tr}[\Sigma^{-1} E_{(i,j)} \Sigma^{-1} E_{(r,s)}] \\ & + \left(b_h - \frac{1}{4}\right) \text{tr}[\Sigma^{-1} E_{(i,j)}] \text{tr}[\Sigma^{-1} E_{(r,s)}], \quad i \leq j, r \leq s. \end{aligned}$$

Finally, the gradient of the log-likelihood for  $n$  observations is given by

$$\begin{aligned} \frac{\partial L}{\partial \beta} &= \frac{n}{\beta} + \frac{nm}{2\beta^2} \left[ \ln(2) + \Psi\left(\frac{m}{2\beta}\right) \right] \\ & - \frac{1}{2} \sum_{i=1}^n \left[ \ln(\mathbf{x}'_i \Sigma^{-1} \mathbf{x}_i) (\mathbf{x}'_i \Sigma^{-1} \mathbf{x}_i)^\beta \right] \end{aligned}$$

and

$$\begin{aligned} \frac{\partial L}{\partial \Sigma_{r,s}} &= -\frac{n}{2} \text{tr}(\Sigma^{-1} E_{(r,s)}) \\ & + \frac{1}{2} \beta \sum_{i=1}^n \left[ (\mathbf{x}'_i \Sigma^{-1} \mathbf{x}_i)^{\beta-1} (\mathbf{x}'_i \Sigma^{-1} E_{(r,s)} \Sigma^{-1} \mathbf{x}_i) \right], \quad r \leq s. \end{aligned}$$

This allows us to solve the equations (5) for the case of an MGGD. In practice, we noticed that an initial estimate by the moment method followed by a single step of the Fisher scoring algorithm suffices for obtaining sufficiently accurate parameter estimates.

### 2.3 Goodness of fit

In order to assess the performance of the various estimation methods, we measured the goodness of the MGGD fits through classical null hypothesis significance testing. However, we split up the main goodness of fit hypothesis into several subhypotheses on the basis of the stochastic representation of an MGGD random vector  $\mathbf{X}$ . Indeed, it can be proved that if  $\mathbf{X}$  is distributed according to a zero-mean MGGD with parameters  $\beta$  and  $\Sigma$ , then  $\mathbf{X}$  can be written as

$$\mathbf{X} = R A' \mathbf{V}, \quad (7)$$

where  $R^2 \equiv U \equiv \mathbf{X}' \Sigma^{-1} \mathbf{X}$ ,  $A'A \equiv \Sigma$  (i.e.  $A$  is the matrix square root of  $\Sigma$ ) and  $\mathbf{V}$  is an  $m$ -dimensional random vector independent of  $R$  and distributed uniformly on the unit sphere (Fang and Zhang, 1990). It is easily shown that  $R^{2\beta}$  has a gamma distribution with shape parameter  $m/(2\beta)$  and scale parameter 2 (Gómez et al, 1998). Given a set of data  $\mathbf{X}_i, i = 1, \dots, n$ , and estimated MGGD parameters  $\Sigma$  and  $\beta$ , one may derive the associated  $R_i^{2\beta}$  and  $\mathbf{V}_i$  as follows:

$$R_i^{2\beta} = (\mathbf{X}'_i \Sigma^{-1} \mathbf{X}_i)^\beta,$$

$$\mathbf{V}_i = \frac{A^{-1} \mathbf{X}_i}{\|A^{-1} \mathbf{X}_i\|}.$$

Then, to test the goodness of fit of an MGGD in the particular case of  $m = 3$ , we carried out a set of  $m + 2 = 5$  hypothesis tests as follows:

<sup>3</sup> G. Verdoolaege and P. Scheunders, On the Geometry of Multivariate Generalized Gaussian Models, submitted to Journal of Mathematical Imaging and Vision, 2010.

1. Test whether  $R^{2\beta}$  is distributed according to a gamma distribution  $\Gamma\left(\frac{m}{2\beta}, 2\right)$  (associated null hypothesis  $H_{0,1}$ ). This was done via Pearson's chi-square test.
2. Test whether  $\mathbf{V}$  is distributed uniformly on the unit sphere ( $H_{0,2}$ ). The points on the unit sphere can be seen as vectors pointing from the origin of the sphere. We chose the length  $T$  of the sample resultant  $\mathbf{T}$  of all vectors  $\mathbf{V}_i$  as a statistic. Indeed, if the points are truly distributed uniformly on the sphere, then ideally, as  $n \rightarrow \infty$ ,  $T$  should vanish. In addition, we used the fact that in the limiting case for large  $n$  (Stephens, 1964; Watson and Williams, 1956):

$$\frac{3}{n}T^2 \sim \chi_3^2.$$

Thus, again a chi-square test could be used.

3. Finally, perform  $m = 3$  tests of independence ( $H_{0,3}$ ,  $H_{0,4}$ ,  $H_{0,5}$ ) between  $R$  and each of the  $m$  components of  $\mathbf{V}$ . To do this we employed Kendall's  $\tau$ , with the normal distribution as an approximation to the sampling distribution of  $\tau$ .

This collection of significance tests was applied to a subset of the data employed in our classification experiments. The details of the image database used in this work are given in Section 4; suffice it to say for now that we checked the goodness of fit of an MGGD for each of the estimating algorithms on a data set consisting of the horizontal wavelet detail coefficients at scale 3 of 640 color ( $m = 3$ ) texture images. Wavelet histograms at higher scales are, compared to lower scales, typically harder to fit due to the lower number of samples available. This motivates our choice for conducting the goodness of fit study at scale 3. The corresponding  $p$ -values were derived and compared to a significance level of  $\alpha = 5\%$ . For a certain wavelet detail image, if none of the five tests leads to a significant deviation from the null hypothesis, the MGGD fit can be deemed good. Next, for each test the percentage was obtained, among the 640 wavelet images, of images for which the null hypothesis could be accepted. These percentages are displayed in Table 1 for the MM, MM-ML and MM-Fisher fit algorithms. Note that we are only concerned about relative differences in percentages, since the absolute percentage of accepted null hypotheses depends, amongst others, also on the sample size. On the whole set of 640 images, for which, naturally,  $\beta > 0$ , the MM-ML method clearly performs best, resulting in the highest percentage of accepted null hypotheses for each test. The difference with the pure moment method is, however, rather small. Nevertheless, the MM-ML method was used for routine estimation of MGGDs for the purpose of the retrieval experiments outlined in Section 4. We repeated the goodness of fit analysis for medium to large values of the shape parameter ( $\beta > 0.5$ , as estimated via ML), yielding a modestly enhanced performance of the MM and

MM-ML methods, but a considerable improvement for the MM-Fisher method. This trend is confirmed for even higher values of the shape parameter ( $\beta > 0.7$ ), although the effect on the performance of the MM-Fisher method is less clear. This result corresponds to the observation by Varanasi and Aazhang (1989) (univariate case) that the fit accuracy of the MM-Fisher method increases relative to the ML performance for successively larger values of  $\beta$ . In addition, as mentioned above, the Fisher scoring step may occasionally enter forbidden regions of parameter space. In order to judge the result of this effect on the  $H_0$  acceptance percentages, the entire analysis was rerun on the subset of images for which the Fisher step produced allowed estimates of  $\Sigma$  and  $\beta$ . As far as this resulted in a change of percentage accepted null hypotheses, the corresponding values are given between parentheses in Table 1. We may conclude that for  $\beta > 0$  a large proportion of MGGD parameters estimated by the MM-Fisher method lies in the forbidden region of parameter space. However, as we restrict our attention to larger values of  $\beta$ , this effect becomes negligible. Nevertheless, on the whole the Fisher step does not appear to succeed in augmenting the goodness of fit for the MGGDs in the data set under study, compared to the initial estimates obtained via the MM.

Furthermore, we should mention that also the moment method does not always lead to acceptable parameter estimates as the numerical algorithm for deriving  $\beta$  may fail to converge. The reason is the occurrence of outliers in the data that strongly affect the kurtosis and the tail behavior and that lead, again, to small values of  $\beta$  ( $\lesssim 0.5$ ). If the moment method is used to initialize the parameters for subsequent fine-tuning by ML or Fisher scoring, and if it fails, default initial parameter values have to be assumed (usually  $\beta = 0.5$ ,  $\Sigma = I_m$ , the  $m \times m$  unit matrix). These initial values in general are far away from the optimal parameters. Therefore, in such cases the linearity of the equations to be solved in the Fisher step, compared to the nonlinearity of the ML equations, results in a substantial difference in time it takes to estimate the parameters. Typically, in our experiments if the MM fails, the Fisher step is about an order of magnitude faster than the ML method. This effect, together with other factors, renders the MM-Fisher method on the entire data set about three times faster than the MM-ML method, both of which are substantially slower than the moment method in itself.

As an aside, we note that even in case of small  $p$ -values (this depends also on the sample size) the MGGD fit in practice still provides a not too bad description of the wavelet statistics. To see this, note that from Table 1 it can be concluded that, first there is a good correlation between the  $p$ -values for the five different tests. Second, the  $p$ -value for  $H_{0,1}$  (fit of the gamma distribution) appears to be a more sensitive measure of the overall MGGD goodness of fit, com-

| Method    | $\beta$ range | Percentage of $H_0$ accepted |           |           |           |           |
|-----------|---------------|------------------------------|-----------|-----------|-----------|-----------|
|           |               | $H_{0,1}$                    | $H_{0,2}$ | $H_{0,3}$ | $H_{0,4}$ | $H_{0,5}$ |
| MM        | $\beta > 0$   | 82 (87)                      | 90        | 95        | 94        | 95        |
|           | $\beta > 0.5$ | 84 (86)                      | 92        | 94        | 93        | 96        |
|           | $\beta > 0.7$ | 94                           | 75        | 94        | 94        | 94        |
| MM-ML     | $\beta > 0$   | 87 (89)                      | 91        | 95        | 94        | 95        |
|           | $\beta > 0.5$ | 89 (88)                      | 92        | 95        | 94        | 96        |
|           | $\beta > 0.7$ | 94                           | 75        | 94        | 94        | 94        |
| MM-Fisher | $\beta > 0$   | 44 (79)                      | 51 (91)   | 53 (95)   | 53 (95)   | 53 (95)   |
|           | $\beta > 0.5$ | 76 (83)                      | 83 (91)   | 86 (95)   | 85 (93)   | 87 (95)   |
|           | $\beta > 0.7$ | 69                           | 75        | 94        | 94        | 94        |

Table 1: Percentage, in a database of 640 color texture images, of accepted null hypotheses for the goodness of fit tests ( $\alpha = 5\%$ ) of a trivariate MGGD to horizontal wavelet detail coefficients at scale 3. The percentages were calculated for different MGGD fit methods: the Method of Moments (MM), MM followed by an optimization using maximum likelihood (MM-ML) and finally MM followed by a single Fisher scoring step (MM-Fisher). Different regions of the parameter space were considered, depending on the range of values of the shape parameter  $\beta$  (estimated using MM-ML). The analysis was repeated for all MGGDs where the MM-Fisher method led to estimates within the allowed parameter space. Numbers between parenthesis indicate any resulting changes in percentages compared to the full data set.

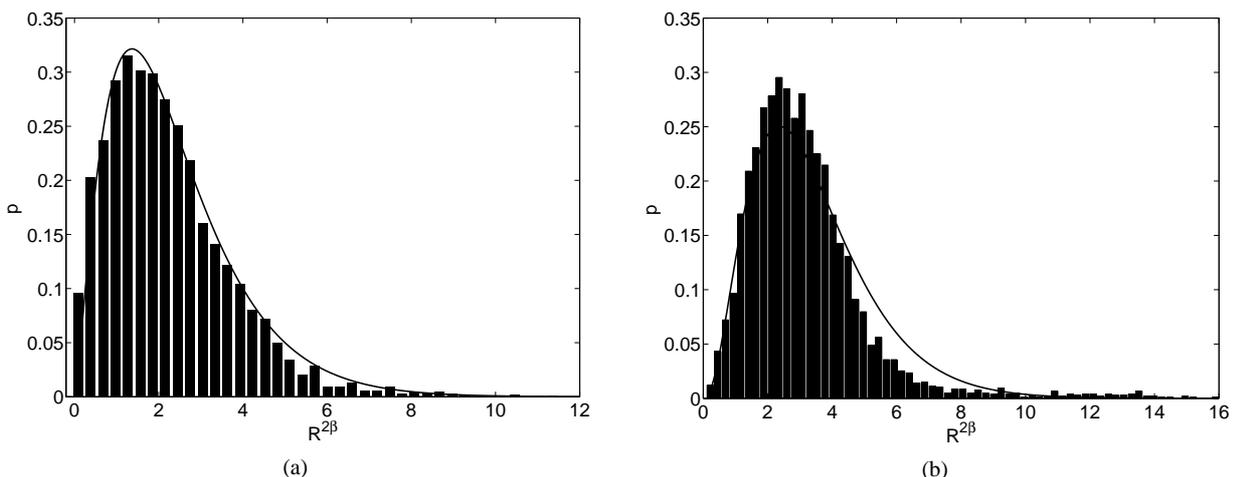


Fig. 2: Examples of a fit of a gamma distribution  $\Gamma\left(\frac{m}{2\beta}, 2\right)$  to the  $R^{2\beta}$  histogram, with  $\Sigma$  and  $\beta$  estimated via maximum likelihood. (a) Case where the  $p$ -value for  $H_{0,1}$  is 0.92. (b) Case with an  $H_{0,1}$   $p$ -value of 0.01.

pared to the  $p$ -values of the other tests. The consequence is that, in order to get a quick first idea of the MGGD goodness of fit, we may monitor only the  $p$ -value corresponding to  $H_{0,1}$ . Alternatively, we can judge the goodness of fit by visual inspection of the histogram for  $R^{2\beta}$ , together with the PDF of the gamma distribution  $\Gamma\left(\frac{m}{2\beta}, 2\right)$ . Two examples are shown in Figure 2, one where  $H_{0,1}$  was accepted ( $p = 0.92$ ) and one where  $H_{0,1}$  was rejected ( $p = 0.01$ ). Clearly, in the second case the fit of the gamma distribution is not optimal, although it still represents a fairly good summary of the histogram.

Finally, we will want to use models for the wavelet statistics characterized by a fixed value of the shape parameter

as well. Specifically, we will use the Gaussian ( $\beta = 1$ ) and Laplacian ( $\beta = 1/2$ ) distributions. Focusing on the maximum likelihood method for estimating these models, we can directly calculate the dispersion matrix  $\Sigma$  from (4). Again, we checked the goodness of the Gaussian and Laplacian fits on the same data set introduced above. The percentages of accepted null hypotheses are given in Table 2, to be compared with the fits by the MM-ML method for the case  $\beta > 0$  in Table 1. Whereas the percentages for  $H_{0,2}, \dots, H_{0,5}$  are clearly similar to the case with variable  $\beta$ , by contrast the more sensitive  $H_{0,1}$  is substantially more rejected for the Laplacian and especially the Gaussian, relative to the MGGD with variable  $\beta$ . Thus, the multivariate Gaussian

| Model   | Percentage of $H_0$ accepted |           |           |           |           |
|---------|------------------------------|-----------|-----------|-----------|-----------|
|         | $H_{0,1}$                    | $H_{0,2}$ | $H_{0,3}$ | $H_{0,4}$ | $H_{0,5}$ |
| Gauss   | 19                           | 90        | 95        | 94        | 95        |
| Laplace | 75                           | 90        | 95        | 95        | 94        |

Table 2: Similar as in Table 1, but for a fit of a multivariate Gaussian and Laplacian using MM-ML on the complete data set of 640 wavelet images.

is in many cases an inferior model for multispectral image wavelet statistics, while the Laplacian, with its heavier tails, is much better suited. The MGGD model with variable shape parameter provides an even better description of the wavelet statistics, with a downside of a more complicated fitting procedure.

We may conclude that, in order to optimize both goodness of fit together with computational load, the moment method suffices as a good first approximation. A subsequent fine-tuning of the distribution parameters by recursive solution of the ML equations, may bring useful additional accuracy to the fit, although this adds to the computational complexity of the fitting procedure. On the contrary, in our experiments the Fisher algorithm did not appear to give an advantage in comparison with the moment method, at least as far as the database under study in the present work is concerned. Indeed, the Fisher step has proved to be counterproductive in many instances, not least because this technique might produce values outside the allowed parameter space. Also by means of significance testing, we have positively determined that the heavy-tailed MGGD model provides indeed a better description of the wavelet data than the Laplacian distribution, and even more so than the Gaussian.

Incidentally, the stochastic representation (7) also allows to sample from an MGGD by successively sampling a vector  $\mathbf{V}$  and a gamma variate  $R^{2\beta}$ . There are several methods to sample from a uniform distribution on the unit sphere (see Marsaglia, 1972), but one of the simplest is to sample independently from  $m$  univariate standard normal distributions, arrange the samples in a vector and normalize the result (Muller, 1959). E.g. for  $m = 3$  we get

$$\mathbf{V} = \frac{1}{\sqrt{x^2 + y^2 + z^2}} \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

with  $x$ ,  $y$  and  $z$  sampled from a standard normal distribution. Sampling from an MGGD can be useful in simulation studies, with another application being the calculation of the KLD between MGGDs through Monte Carlo integration, as outlined below.

### 3 MGGD similarity measures

We now consider two similarity measures between MGGDs, the geodesic distance and the Kullback-Leibler divergence, and we outline how to calculate them relatively fast. In both instances, we draw a distinction between the cases of a fixed and variable MGGD shape parameter.

#### 3.1 Geodesic distance

As mentioned above, within the framework of information geometry the Fisher information can be seen as a Riemannian metric tensor  $g_{\mu\nu}$  on an  $N$ -dimensional probabilistic manifold (Cramér, 1946; Rao, 1945). The manifold formed by zero-mean MGGDs ( $m$ -variate) is in principle parametrized by  $\beta$  and  $\Sigma_{ij}$ , the  $(i, j)$ -th element of the matrix  $\Sigma$ ,  $i \leq j$ , so that  $N = 1 + m(m+1)/2$ . However, it turns out that a more elegant parametrization can be found, with the geodesic in mind between a specific pair of MGGDs  $(\beta_1, \Sigma_1)$  and  $(\beta_2, \Sigma_2)$ <sup>4</sup>. In this parametrization the Fisher-Rao metric depends only on  $\beta$ , so it is easy to tabulate the metric for a range of  $\beta$  values, which needs to be done only once.

Given a metric one can calculate geodesics on the manifold. A geodesic path between two points (MGGDs) on the manifold is the curve lying within the manifold that connects the points and has minimum length. Geodesics can be found by solving the geodesic equations, whereupon the length of the geodesic, i.e. the geodesic distance between the two points, can be computed. In the case of a fixed shape parameter  $\beta$ , the geodesics take the form of straight lines in  $\mathbb{R}^m$  (Berkane et al, 1997). As a result, the geodesic distance between two MGGDs characterized by  $(\beta, \Sigma_1)$ , respectively  $(\beta, \Sigma_2)$ , exists in a closed form. Denoting this specific distance by  $\text{GD}(\beta, \Sigma_1 || \beta, \Sigma_2)$ , we have:

$$\text{GD}(\beta, \Sigma_1 || \beta, \Sigma_2) = \left[ \left( 3b_h - \frac{1}{4} \right) \sum_i (r_2^i)^2 + 2 \left( b_h - \frac{1}{4} \right) \sum_{i < j} r_2^i r_2^j \right]^{1/2}, \quad (8)$$

with  $r_2^i \equiv \ln \lambda_2^i$  and  $\lambda_2^i$ ,  $i = 1, \dots, m$ , the  $m$  eigenvalues of  $\Sigma_1^{-1} \Sigma_2$ . In addition,  $b_h$  is defined by

$$b_h \equiv \frac{1}{4} \frac{m + 2\beta}{m + 2}.$$

If on the other hand  $\beta$  is allowed to vary, the geodesic equations are more difficult to solve and we did not obtain a closed-form solution. Instead, through numerical optimization, polynomial solutions can be obtained for the coordinates as a function of the geodesic's parameter  $t$  (i.e. the

<sup>4</sup> G. Verdoolaege and P. Scheunders, On the Geometry of Multivariate Generalized Gaussian Models, submitted to Journal of Mathematical Imaging and Vision, 2010.

coordinate functions)<sup>4</sup>. In practice only a few iterations of a BFGS Quasi-Newton method usually suffice to obtain substantial information on the geodesics. However, this scheme still is computationally too intensive to be practical for e.g. real-time image retrieval. Therefore, in this paper we use a linear approximation to the geodesic coordinate functions<sup>5</sup>. The calculation of the geodesic distances via integration along the geodesic was then carried out using a simple (and fast) trapezium rule.

### 3.2 Kullback-Leibler divergence

The KLD between two distributions  $p_1(\mathbf{x}|\theta_1)$  and  $p_2(\mathbf{x}|\theta_2)$ , denoted by  $\text{KLD}(p_1||p_2)$ , is defined by

$$\text{KLD}(p_1||p_2) = \int p_1(\mathbf{x}|\theta_1) \ln \frac{p_1(\mathbf{x}|\theta_1)}{p_2(\mathbf{x}|\theta_2)} d\mathbf{x}, \quad (9)$$

where the integral is over the whole data space. A closed-form expression for the KLD between two univariate zero-mean generalized Gaussians was obtained by Do and Vetterli (2002). In the terminology that we use, and introducing the dispersions  $\sigma_i$ ,  $i = 1, 2$  ( $\Sigma_i$  reduces to  $\sigma_i^2$  in the univariate case), the KLD between two GGDs characterized by  $(\beta_1, \sigma_1)$  and  $(\beta_2, \sigma_2)$  (shortly  $\text{KLD}(\beta_1, \sigma_1||\beta_2, \sigma_2)$ ) is given by

$$\begin{aligned} \text{KLD}(\beta_1, \sigma_1||\beta_2, \sigma_2) = & \ln \left[ \frac{\beta_1 2^{\frac{1}{2\beta_2}} \sigma_2 \Gamma\left(\frac{1}{2\beta_2}\right)}{\beta_2 2^{\frac{1}{2\beta_1}} \sigma_1 \Gamma\left(\frac{1}{2\beta_1}\right)} \right] \\ & + \left( \frac{2^{\frac{1}{2\beta_1}} \sigma_1}{2^{\frac{1}{2\beta_2}} \sigma_2} \right)^{2\beta_2} \frac{\Gamma\left(\frac{2\beta_2+1}{2\beta_1}\right)}{\Gamma\left(\frac{1}{2\beta_1}\right)} - \frac{1}{2\beta_1}. \end{aligned}$$

An analytic expression for the KLD between multivariate zero-mean Gaussian distributions ( $\beta = 1$  in (2)) is also known since long (Kullback, 1968). It is given by

$$\text{KLD}(\Sigma_1||\Sigma_2) = \frac{1}{2} \left[ \ln \frac{|\Sigma_2|}{|\Sigma_1|} + \text{tr}(\Sigma_2^{-1} \Sigma_1) - m \right], \quad (10)$$

with  $\Sigma_1$  and  $\Sigma_2$  the respective covariance matrices.

The obvious generalization of these results to an analytic expression for the KLD between two multivariate zero-mean GGDs is not straightforward. Recently, we obtained a closed form for the KLD between two bivariate zero-mean GGDs parametrized by  $(\beta_1, \Sigma_1)$  and  $(\beta_2, \Sigma_2)$  (Verdoolaege

et al, 2009), given by

$$\begin{aligned} \text{KLD}(\beta_1, \Sigma_1||\beta_2, \Sigma_2) = & \ln \left[ \frac{\Gamma\left(\frac{1}{\beta_2}\right)}{\Gamma\left(\frac{1}{\beta_1}\right)} 2^{\left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right)} \left( \frac{|\Sigma_2|}{|\Sigma_1|} \right)^{\frac{1}{2}} \frac{\beta_1}{\beta_2} \right] \\ & - \frac{1}{\beta_1} + \left[ 2^{\left(\frac{\beta_2}{\beta_1} - 1\right)} \frac{\Gamma\left(\frac{\beta_2+1}{\beta_1}\right)}{\Gamma\left(\frac{1}{\beta_1}\right)} \right. \\ & \left. \times \left( \frac{\gamma_1 + \gamma_2}{2} \right)^{\beta_2} {}_2F_1\left(\frac{1-\beta_2}{2}, -\frac{\beta_2}{2}; 1; A^2\right) \right]. \quad (11) \end{aligned}$$

Here,  $\gamma_i \equiv (\lambda_i^j)^{-1}$ ,  $i = 1, 2$ , with again  $\lambda_i^j$  the eigenvalues of  $\Sigma_1^{-1} \Sigma_2$ , while  $A \equiv \frac{\gamma_1 - \gamma_2}{\gamma_1 + \gamma_2}$ .  ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$  represents the Gauss hypergeometric function (Abramowitz and Stegun, 1965), which may be tabulated for  $-1 < A < 1$  and for realistic values of  $\beta$ . In the case of two Gaussians,  $\beta_1 = \beta_2 = 1$ , the hypergeometric function in (11) becomes identically 1 and it can easily be verified that (11) reduces to the familiar expression (10).

However, we were unable to find an analytic expression for the KLD between two multivariate zero-mean GGDs in general, i.e. for dimension greater than 2. Therefore, in the general case the KLDs have to be estimated through numerical evaluation of the integral in (9). This is usually carried out via Monte Carlo integration, by viewing (9) as an expectation over  $p_1$  of the logarithm in the integrand. This can be done reliably and reproducibly only if a sufficient amount of samples from  $p_1$  is available. These samples can be obtained from the data (images) itself or, alternatively, the samples may be drawn directly from  $p_1$  via the procedure for sampling from an MGGD outlined in Section 2.3. Nevertheless, we must point out that Monte Carlo integration in general is a too slow process to be useful in many applications, such as image retrieval. Moreover, the calculation of the KLD in itself yields little geometric insight concerning the MGGD space. As a result, the accuracy of approximations of the KLD via Monte Carlo integration can only be adjusted in a less controlled way, compared to the approximation of a geodesic path. For these reasons we did not carry out the Monte Carlo integration to estimate the KLD in the cases where no closed expression was available.

As an aside, we note that for distributions that lie infinitesimally close on the probabilistic manifold, it can be proved that the Kullback-Leibler divergence equals half of the squared geodesic distance between the distributions (see e.g. Kullback, 1968; Lenglet et al, 2006a). Hence, in such a case the KLD and GD yield similar results, but in general they are quite different measures of similarity between distributions.

<sup>5</sup> Note that the geodesic itself still lives on the curved MGGD manifold.

### 3.3 Multiple wavelet subbands

So far we have considered the calculation of the GD and KLD between single MGGDs. However, with a view to our application of texture discrimination, we will want to decompose the images in a multiscale representation, namely through the wavelet transform. After the example of Do and Vetterli (2002), we will assume independence among the  $B$  wavelet subbands, modeling the spectral components belonging to each subband via an independent MGGD. More precisely, we will neglect the dependence of wavelet subbands compared to the correlation structure associated to the spectral bands. To check whether this approximation is indeed valid, we calculated the correlation coefficient between wavelet data (resulting from the stationary wavelet transform) belonging to different directions (horizontal, vertical, diagonal) for every level. In addition, for every direction we computed the correlation between wavelet data associated to different levels. We compared this with the correlation between spectral bands, for every wavelet subband. We performed this check for wavelet data obtained from every image in the database of 640 images described below. Overall, the median of the absolute value of the correlation coefficient between directions was 0.037 (interquartile interval  $[0.013, 0.075]$ ) and between levels this was 0.022 ( $[0.004, 0.062]$ ). However, the correlation between spectral bands was substantially stronger, resulting in a median absolute value of 0.986 ( $[0.968, 0.994]$ ) between red and green, 0.965 ( $[0.925, 0.985]$ ) between red and blue and 0.987 ( $[0.963, 0.995]$ ) between green and blue. Thus, we may safely neglect the correlation between wavelet directions and levels, compared to the correlation between spectral bands and we will also neglect any higher-order dependence between directions and levels.

If independence among wavelet subbands is assumed, the joint distribution of the image over all  $B$  subbands is just the product of the MGGDs corresponding to the individual subbands. It is not difficult to see that the corresponding Fisher matrix is block diagonal and that the geodesic equations become  $B$  mutually independent systems of  $m + 1$  equations each. The  $B$  systems can be solved along the same lines as described above. For the case of a fixed shape parameter, this yields a sum of terms of the form under the square root in (8), corresponding to the respective subbands. For a variable shape parameter, we make use of the fact that a geodesic is a curve on the manifold not only with minimum length, but also with minimum squared length (energy). Therefore, one may just sum the *squared* geodesic distances for the respective subbands. Regarding the KLD, when dealing with multiple independent wavelet subbands, the respective KLDs may just be summed according to the chain rule (Kullback, 1968; Do and Vetterli, 2002).

Naturally, in the case where the independence assumption of the wavelet subbands is not justified, a better model that takes into account the dependence of both the spectral bands and the wavelet subbands may well yield still enhanced results in a texture classification experiment.

## 4 Classification and retrieval experiments

Equipped with techniques to calculate the GD and KLD between MGGDs, we proceed with the discussion of a series of experiments on the classification and retrieval of textured color images based on texture discrimination. We start with an overview of our experimental setup.

### 4.1 Experimental setup

We first conducted a series of classification experiments on gray-level and color textures. We next repeated the experiments in an image retrieval application. We started the experiments on a small database that is meant to provide a bench-mark for the comparison of different statistical models as well as comparison of the GD with the KLD. This database was the same as the one used by Do and Vetterli (2002), which allows easy comparison to the retrieval results on gray-level images obtained by Do and Vetterli (2002). The database comprised a set of 40 images from the Vis-Tex database. These are real world  $512 \times 512$  color images from different natural scenes (textures), displayed in Fig. 3, selected because of their sufficient homogeneity. The images were expressed in the RGB color space. Every image was divided in sixteen nonoverlapping  $128 \times 128$  subimages, constituting a database of 640 subimages. Gray-level images were obtained from the original color images by calculating their luminance. In order to render the retrieval task sufficiently challenging, every color (or gray-level) component of each subimage was individually normalized to zero mean and unit variance. As a result, the gray scales of subimages from the same original image were generally not in the same range. Then, on every component individually a discrete wavelet transform was applied with three levels using the Daubechies filters of length eight. The wavelet detail coefficients of every subband over the three color components (or the gray-level) were modeled by an (M)GGD using the MM-ML approach. The parameters of the (M)GGD models for all subbands comprise the feature set for a single subimage. The experiments were carried out once using only one wavelet scale and once using all three wavelet scales. In order to get an idea of the range of values for the shape parameter  $\beta$  in our data set, a histogram of  $\beta$ , over all nine wavelet subbands, is given in Figure 5a for the gray-level images. Figure 5b shows a similar histogram for the case of RGB images modeled by trivariate MGGDs. For gray-scale

images the mean  $\beta$  is about 0.58, while for correlated color images it is slightly lower: 0.48 (about Laplacian). Again, this is an indication that the Laplacian is much more suitable as a model than the Gaussian.

The classification experiments were carried out by means of a  $k$ -nearest neighbor classifier validated by the leave-one-out method. In practice, we considered one of the 640 subimages (the test image), to be assigned to one of the 40 original texture classes. The class labels of the other subimages were assumed to be known, which constitutes the training phase of the classifier. Then, the similarity of the test image to each of the remaining images was determined and the test image was assigned to the class most common among the  $k = 15$  nearest neighbors of the test image. We chose  $k = 15$  since ideally the 15 nearest neighbors of the test image should be the 15 subimages originating from the same original texture class to which the test image belonged in reality. We next compared the assigned class label with the true class label of the test image. We repeated the procedure, successively using every subimage once as a test image (hence every time leaving out one of the subimages in the training phase of the classifier). We then determined the rate of correct classification and used this as a performance measure for the classifier.

The classification experiment can easily be applied in a retrieval context, again by sequentially presenting every subimage as a query image. The retrieval effectiveness is then measured by calculating the average *retrieval rate*, i.e. the ratio of relevant images in the top 15 images (excluding the query image). Here, a subimage is considered relevant if it is part of the same original  $512 \times 512$  image as the query subimage.

We conducted both the classification and retrieval experiments for various choices of the statistical model and using the GD and KLD as a similarity measure. We started the experiments on the gray-level equivalent of the 640 color images and we next treated the corresponding full RGB color images assuming no interband correlation (hence univariate modeling, referred to as ‘UV’). In this case, the GD and KLD for the joint (product) distribution over the color bands can be calculated remembering the comments above on the calculation for independent wavelet subbands. Then, color images were considered where the (bivariate) correlation was modeled pairwise between the color bands (bivariate modeling, referred to as ‘MV-2’). Since a closed expression for the KLD between bivariate GGDs is available, this procedure allows us to explicitly compare the GD with the KLD in the bivariate case. Finally, color images considering the full (trivariate) correlation structure between the spectral bands (referred to as ‘MV-3’) were used (recall that in every case the wavelet subbands were considered to be mutually independent). For each of these instances, we employed several MGGD models, with fixed and with variable shape

parameter. In the case of fixed shape parameter, we chose two models, viz  $\beta = 1$ , i.e. the (multivariate) Gaussian and the multivariate Laplacian, characterized by  $\beta = 1/2$ . For a variable shape parameter, the GD was evaluated using a linear approximation for the geodesic coordinate functions, as mentioned above. In the cases where a closed expression is available for the KLD, we compared its performance as a similarity measure with the results using the GD.

We then compared our retrieval results for gray-level images to a recent state of the art method that employs image features based on image patch exemplars (Varma and Zisserman, 2003, 2009), questioning the advantage of filter banks. We applied this method to retrieval on our database of 640 subimages as follows. First, for every texture each pixel in the corresponding 16 subimages was represented by a vector consisting of the normalized gray-level values in the pixel’s surrounding  $3 \times 3$  region. A  $k$ -nearest neighbor clustering algorithm was then applied in this vector space, resulting in 10 cluster centers called *textons*. The total collection of 400 *textons* (40 textures) was then used to build a *texton dictionary*. In the learning stage, each pixel in the 640 database images was then converted to its closest *texton* and a *texton histogram* was generated for each subimage. In the classification stage, the *texton histogram* of each query image was compared to all database histograms (neglecting the small influence of the query image itself on the dictionary that was used for translation) using the  $\chi^2$  statistic (the GD could be applied here as well). Finally, again the images were ordered according to their distance to the query image and average retrieval rates were calculated. We refer to this method as the ‘*texton method*’. We did not extend this method to color texture discrimination.

We next examined the influence of Gaussian noise added to the images and studied the impact of image conversion to a less correlated color space, namely HSV space.

Finally, in order to demonstrate the texture discrimination capabilities of our techniques in a realistic retrieval setting, we conducted several texture retrieval experiments in a database of 100,000 images of size  $256 \times 256$ . Every database image was made up of about four to five irregularly shaped patches of different colored textures. These textures were chosen randomly from a set of 198 relatively homogeneous base textures, obtained from the VisTex database and from an online image repository (Image★After, 2010). Some example images are displayed in Figure 4. During construction of the database images, for every base texture  $T_i$  ( $i = 1, \dots, 198$ ) the total number of times  $N_i$  was counted that this specific base texture was used in a patch in a database image. The image features were defined as follows. Every image was divided in 16 nonoverlapping  $64 \times 64$  subimages and MGGD features were calculated for each of the subimages, just as in the previous experiment, except that the color components were not normalized. In addition, we used only

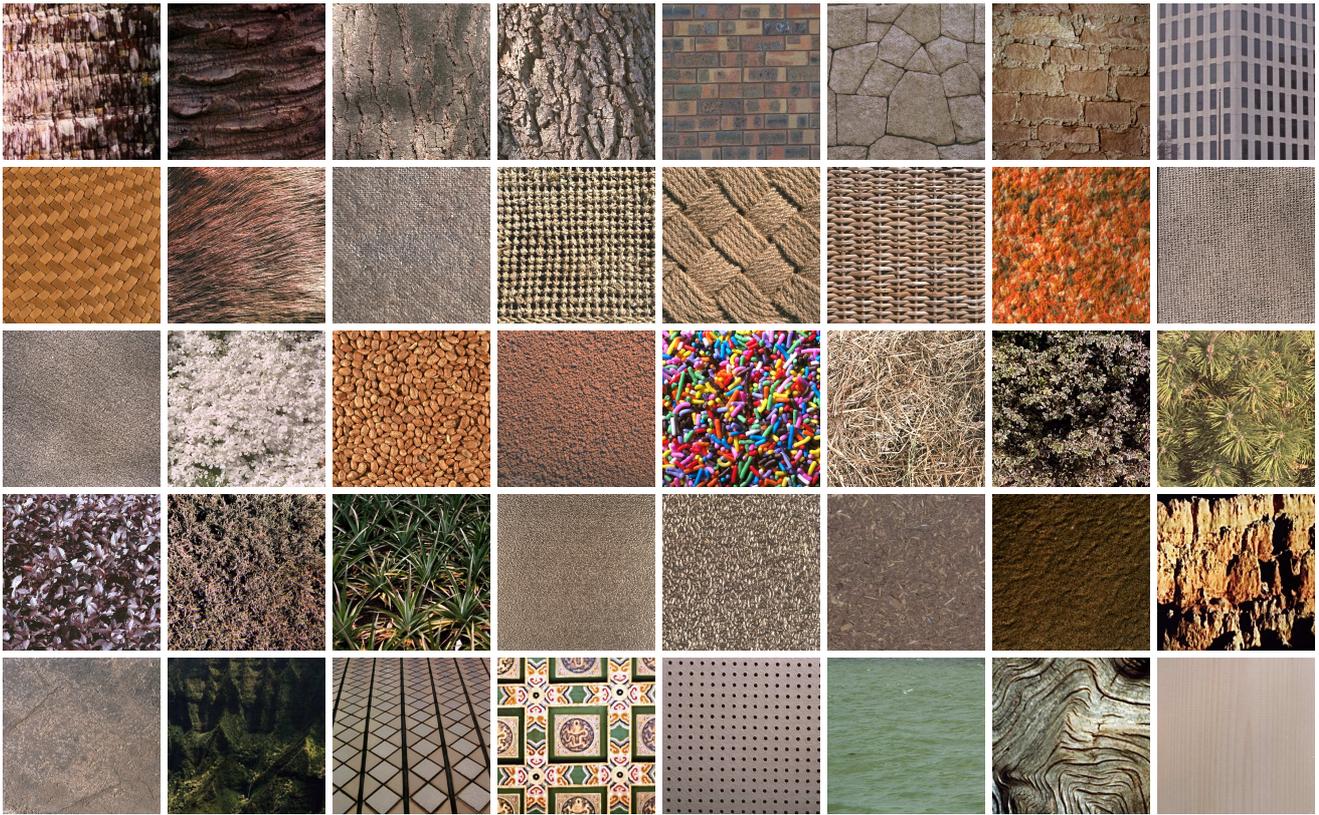


Fig. 3: The  $512 \times 512$  texture images from the VisTex database used in our experiments. From left to right and top to bottom the images are: Bark0, Bark6, Bark8, Bark9, Brick1, Brick4, Brick5, Buildings9, Fabric0, Fabric4, Fabric7, Fabric9, Fabric11, Fabric14, Fabric15, Fabric17, Fabric18, Flowers5, Food0, Food5, Food8, Grass1, Leaves8, Leaves10, Leaves11, Leaves12, Leaves16, Metal0, Metal2, Misc2, Sand0, Stone1, Stone4, Terrain10, Tile1, Tile4, Tile7, Water5, Wood1 and Wood2.

one wavelet scale in order to reduce the computational load. The set of MGGD features for all 16 subimages, constituted the features for a single database image. A query consisted of the presentation of a random  $64 \times 64$  patch chosen from one of the base textures  $T_i$ . For every database image, the similarity of the MGGD features of this query image to the features of each of the 16 subimages was calculated. The similarity of the query image to the database image was then defined as the smallest of the 16 similarities. Next, the  $N_i$  closest database images were considered. Ideally, each of these  $N_i$  images should contain a patch filled with texture  $T_i$ . In reality, our success rate for a single query was the ratio of database images in the top  $N_i$  that contained texture  $T_i$ . We conducted 1000 queries and finally computed the average success rate. This experiment was carried out for gray-scale images (univariate GGD) and RGB color images (with correlated spectral bands, trivariate MGGD), using a fixed shape parameter (Gaussian, Laplacian) or a variable shape parameter and applying the GD or KLD as a similarity measure.

#### 4.2 Computational demands

The performance of a classifier or a retrieval system is not only expressed in terms of its actual classification efficiency, but is also a function of its speed. Depending on the application, a trade-off needs to be considered between classification effectiveness and computational demands. In order to estimate the computational resources required by any combination of models and similarity functions introduced above, we measured the amount of time necessary for the evaluation of the similarity measure between two images, characterized by in total nine wavelet subbands at three scales. This duration was obtained on the machine on which all calculations in this work were performed, namely a Dell Optiplex 755 equipped with an Intel Core Duo Quad CPU at 2.7 GHz and 8 GB of RAM, running the 64-bit version of the Windows Vista operating system. The codes for the classifier and the retrieval system were implemented and run in MATLAB (version 7.6, R2008a, 64 bit (The Mathworks, 2008)). The durations are summarized in Table 3 (two significant digits were considered in every case).

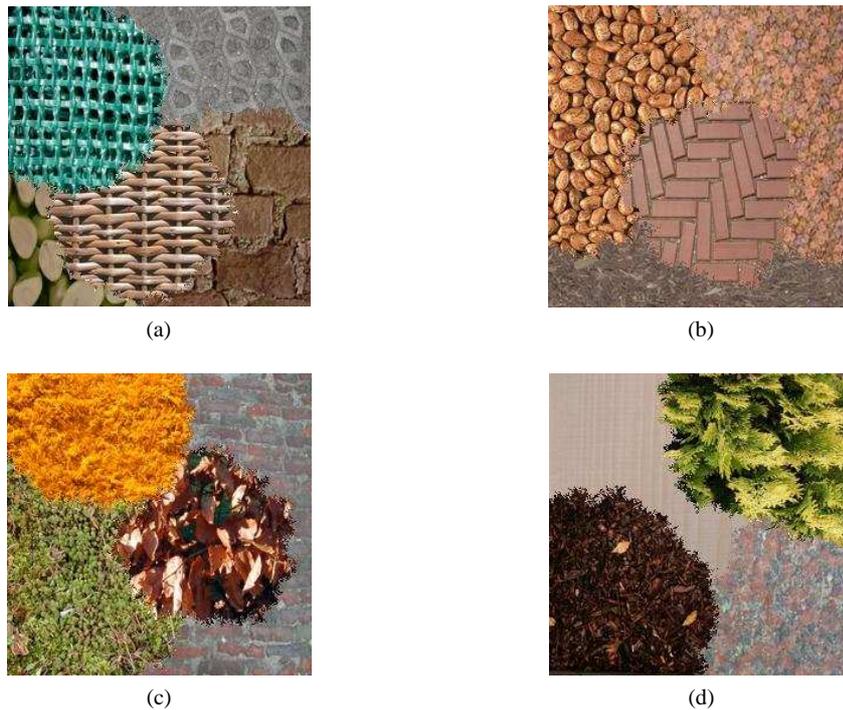


Fig. 4: Example images from a database of size 100,000. Every image ( $256 \times 256$ ) consists of arbitrarily shaped patches of random colored textures.

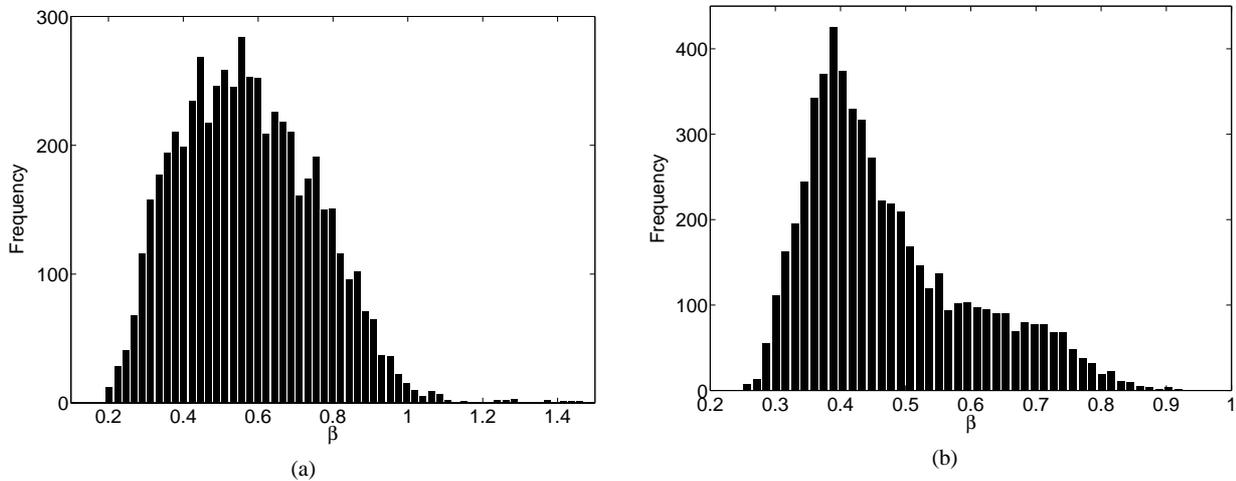


Fig. 5: Histograms of the (M)GGD shape parameters  $\beta$  obtained via MM-ML over the nine wavelet subbands (three scales) in an image database of 640 color texture images. (a) For gray-scale images and (b) for three correlated RGB color bands.

Several points in this table are noteworthy. First, clearly the calculation of similarities between gray-level images is the fastest, followed by color images without correlation between the spectral bands. Evaluating distances between color images with the correlation between the spectral bands modeled pairwise (MV-2) takes considerably more time. The reason is that for each comparison in effect 27 ( $3 \times 9$ ) dis-

tances between bivariate GGDs have to be computed. However, the durations obtained by computing the distances between trivariate MGGDs (MV-3), characterizing the full correlation structure between the three color bands, are again somewhat lower because here effectively only nine KLDs or GDs have to be evaluated. Further tests on synthetic data have pointed out a quadratic asymptotic scaling of the com-

| Measure | Model   |       |       |      |      |
|---------|---------|-------|-------|------|------|
|         |         | Gray  | UV    | MV-2 | MV-3 |
| KLD     | Gauss   | 0.021 | 0.023 | 0.96 | 0.37 |
|         | Laplace | 0.018 | 0.020 | 1.7  | –    |
|         | GGD     | 0.066 | 0.089 | 2.2  | –    |
| GD      | Gauss   | 0.053 | 0.074 | 0.99 | 0.57 |
|         | Laplace | 0.053 | 0.074 | 0.99 | 0.57 |
|         | GGD     | 1.9   | 5.5   | 7.3  | 2.8  |

Table 3: Time, in ms, necessary for the calculation of the similarity between two textured images (nine wavelet sub-bands) characterized by different models (Gray = gray-level, UV = univariate color, MV-2 = multivariate color with pairwise correlation between the spectral bands, MV-3 = multivariate color with full correlation structure) using the KLD and GD.

putational time as a function of the number of spectral bands, both for the fixed and variable  $\beta$  cases.

A second observation is that it takes consistently more time (except for the MV-2 Laplace) to evaluate the GD, compared to the corresponding KLD. However, in most cases the durations are of the same order of magnitude. An exception is formed by some of the GD calculations for variable shape parameter, which take several milliseconds, in general much longer than is necessary to evaluate the KLD for corresponding models and for variable shape parameter as well. Hence, in order to render the effort of computing the GD for a variable  $\beta$  worthwhile, the associated correct classification rates and retrieval rates will need to be significantly higher than any of the rates obtained using other models and/or the KLD. In the next subsection, we will see that for classification and retrieval on the present database, this is convincingly the case for univariate models. In the multivariate case it will become clear that the application of the Laplace distribution as a model and the GD as a similarity measure, yields the best classification results with the additional benefit of an acceptable computational load.

## 4.3 Results and discussion

### 4.3.1 Main experiments

Our experiments enable a study of the influence on the correct classification rate (CCR) and the average retrieval rate (RR) of the image model as well as the similarity measure used. The results of all classification and retrieval experiments on the database of 640 images are discussed below and summarized in Tables 4 and 5, respectively. The main conclusions that can be drawn from these tables are the following.

- When the three colors in an image are described individually by a univariate distribution, instead of using only the gray-level information, both the CCR and RR increase significantly. However, the most substantial performance improvement compared to gray-level modeling is obtained by taking into account the correlation structure between the spectral bands. The downside is a substantially increased demand for computational resources, as discussed in Section 4.2. Note that the retrieval result on gray-scale textures using the KLD and GGD was already obtained by Do and Vetterli (2002).

In order to assess the associated specific gain in retrieval performance, a comparison was made, as a function of each texture class separately, of the cases with and without correlation. To do this, within every texture class the mean was taken of the average retrieval rate obtained by successively presenting each of the subimages in the respective class to the retrieval system. Since the Laplace distribution gives the highest retrieval rates in three dimensions (see also below), the comparison was made by means of the Laplacian model, using the GD. The results are displayed in Figure 6a (three wavelet scales) for the first 20 texture images shown in Figure 3, while the other images are analyzed in Figure 6b. The differences are in some cases spectacular, implying strong correlation between the color bands. These figures may also be compared to the corresponding results by Do and Vetterli (2002) on gray-level textures.

The same situations were compared in terms of their corresponding average retrieval rate as a function of the number of retrieved images considered. This is shown in Figure 7. At 16 retrieved images considered, naturally we obtain the respective retrieval rates that are mentioned in Table 5. In order to reach the same average retrieval rate, typically more than 2.5 times the number of images need to be considered when the correlation is not modeled, compared to the case where the correlation information is taken into account. Again these findings can be compared to the results by Do and Vetterli (2002).

- The Laplace distribution yields in all cases (both KLD and GD) a higher CCR and RR than the Gaussian. This corroborates our findings in Section 2.3, where it was concluded that the Laplace distribution describes image wavelet statistics more effectively than the Gaussian. Furthermore, the highest rates in our series of experiments are obtained by means of a trivariate Laplace distribution, employing the GD as a distance measure. According to Section 4.2 the computational cost is still reasonable in this situation. Also note that in this case there is no closed expression available for the KLD, nor for a trivariate GGD with variable shape parameter.
- The (M)GGD, characterized by a variable shape parameter in turn may lead to superior classification and re-

trieval performance compared to the Laplacian (both KLD and GD). However, this effect is most clearly noticed when using univariate distributions, but diminishes, or even marginally reverses, as the correlation between color bands is taken into account and as we include more wavelet scales. This is partly understandable since, first, enhancing the complexity of the model, hence increasing the classification performance, it becomes more and more difficult to still improve on the results. Second, we have to keep in mind that the accuracy of the calculation of the GD between variable  $\beta$  MGGDs depends on the degree to which the linear approximation of the geodesic coordinate functions remains valid. As a third possible cause we may point out that the estimation of the MGGD shape parameter is a difficult task, particularly for low values of the shape parameter, as is evident from the discussion in Section 2.3. As a result, the estimated shape parameter might yield a suboptimal MGGD description of the wavelet statistics. On the other hand, fitting a Laplace distribution (hence fixed  $\beta$ ) to wavelet data is easier and it may well be that in the end in some cases the estimated Laplace distributions provide, on the average over the entire database, a better description of the data than the hard to fit MGGD models with variable  $\beta$ , thus explaining the better performance of the Laplace distribution in a classification experiment. As a result, although the reason is not entirely clear as yet, we have to conclude that evidently, on the present image database, for multivariate distributions a variable  $\beta$  model provides no significant advantage over a Laplace distribution. This is especially apparent considering the substantially higher computational load both during estimation and classification using the variable  $\beta$  distributions in comparison with fixed  $\beta$  models.

Incidentally, we also carried out a classification experiment using the GD on trivariate GGDs with variable shape parameter, this time calculating a polynomial approximation of higher degree (degree 10) to the geodesic coordinate functions, as we have described elsewhere<sup>6</sup>. However, we did not observe any resulting significant increase in classification performance.

- The GD is in all cases more effective as a similarity measure than the KLD. The largest disparity is obtained when the colors of the image are taken into consideration as opposed to using only the gray levels. However, as noted in Section 4.2, the superior classification performance of the GD comes at the cost of higher computational demands.
- The modeling of the full correlation structure using trivariate distributions yields slightly better results than through

pairwise bivariate modeling of the color bands. Note, however, that since at this point no analytic expression is available for the KLD between trivariate GGDs (except for the Gaussian), leaving only the possibility of computationally intensive multidimensional integration, the KLD between such distributions was not computed in this work. This is one of the advantages of the GD over the KLD.

#### 4.3.2 Comparison to state of the art

Using the texton method, again average retrieval rates for gray-level images were considered as a function of the number of retrieved images considered. At 16 images retrieved, this yielded a rate of 78.7%, somewhat more than using the GD on three wavelet scales of the gray-level images. However, at only a slightly higher number of retrieved images, the texton method performs inferior to the wavelet methods used in this paper, as can be seen in Figure 7.

#### 4.3.3 Influence of noise

We next examined the influence of noise on the classification results. For that purpose we added Gaussian noise to all images in the database of size 640, before carrying out the wavelet transform (three scales). We chose a signal-to-noise ratio (ratio of the standard deviations) of 1 and we repeated the classification experiments using the gray-scale and MV-3 features. The corresponding CCRs are given in Table 6. Overall the classification rates are lower than those for the original images, but there is no dramatic drop in classification efficiency, despite the relatively high noise level. Second, the KLD appears to be slightly more robust against the noise compared to the GD, except in the case of the univariate Laplace distribution. Third, in this case there is a merit of the trivariate MGGD with variable shape parameter, compared to the trivariate Laplace distribution. The other trends are similar as in the unperturbed case.

#### 4.3.4 Influence of color space

It is interesting to study the effect of the particular color space that is used to compute the wavelet features from. For instance, the dimensions in HSV space (with cylindrical coordinates hue, saturation and value) are much less correlated than in RGB space. Indeed, as an example we converted the images in the database of size 640 into HSV space and then performed a wavelet transform independently on each coordinate (three scales). In this space, the correlation between wavelet directions and levels (for the stationary wavelet transform) turned out to be comparable to that obtained in RGB space. However, for the correlation between the hue and saturation we computed a median absolute value

<sup>6</sup> G. Verdoolaege and P. Scheunders, On the Geometry of Multivariate Generalized Gaussian Models, submitted to Journal of Mathematical Imaging and Vision, 2010.

| Measure |         | Model     |      |      |      |              |      |      |      |
|---------|---------|-----------|------|------|------|--------------|------|------|------|
|         |         | One scale |      |      |      | Three scales |      |      |      |
|         |         | Gray      | UV   | MV-2 | MV-3 | Gray         | UV   | MV-2 | MV-3 |
| KLD     | Gauss   | 68.0      | 78.1 | 89.2 | 90.5 | 77.8         | 83.8 | 94.4 | 95.9 |
|         | Laplace | 69.4      | 79.2 | 91.3 | –    | 81.4         | 85.9 | 95.2 | –    |
|         | GGD     | 78.4      | 82.2 | 91.9 | –    | 87.5         | 89.2 | 94.5 | –    |
| GD      | Gauss   | 68.4      | 78.9 | 89.5 | 92.7 | 79.5         | 85.8 | 96.4 | 97.0 |
|         | Laplace | 70.3      | 83.9 | 92.0 | 93.4 | 83.6         | 90.0 | 96.7 | 97.5 |
|         | GGD     | 80.0      | 84.5 | 93.4 | 93.0 | 88.9         | 91.3 | 96.1 | 97.3 |

Table 4: Correct classification rates (%) using different models (Gray = gray-level, UV = univariate color, MV-2 = multivariate color with pairwise correlation between the spectral bands, MV-3 = multivariate color with full correlation structure) for a single and for three wavelet scales, using the KLD and GD as similarity measures in a database of 640 color texture images.

| Measure |         | Model     |      |      |      |              |      |      |      |
|---------|---------|-----------|------|------|------|--------------|------|------|------|
|         |         | One scale |      |      |      | Three scales |      |      |      |
|         |         | Gray      | UV   | MV-2 | MV-3 | Gray         | UV   | MV-2 | MV-3 |
| KLD     | Gauss   | 53.7      | 62.6 | 79.2 | 80.2 | 63.6         | 70.2 | 85.8 | 87.4 |
|         | Laplace | 54.5      | 64.3 | 81.5 | –    | 65.5         | 71.2 | 87.8 | –    |
|         | GGD     | 66.4      | 71.2 | 82.3 | –    | 76.6         | 77.1 | 87.6 | –    |
| GD      | Gauss   | 53.9      | 63.4 | 80.2 | 81.2 | 64.4         | 71.5 | 87.7 | 89.6 |
|         | Laplace | 56.8      | 67.6 | 83.3 | 84.1 | 68.3         | 74.9 | 89.9 | 91.7 |
|         | GGD     | 66.7      | 73.4 | 84.1 | 84.7 | 77.8         | 79.9 | 89.5 | 91.3 |

Table 5: Same as Table 4, but with application in a retrieval experiment. The average retrieval rates (%) are mentioned.

of only 0.119 (interquartile interval [0.045, 0.274]), between hue and value this was 0.110 ([0.038, 0.260]) and between saturation and value we obtained 0.70 ([0.45, 0.82]). Thus, only the saturation and value coordinates display an appreciable correlation, while the correlation between hue and the other coordinates is much weaker than between the RGB bands. In fact, we would expect a substantially enhanced classification efficiency primarily by modeling the correlation between saturation and value. Also, since obviously there lies less information in the HSV correlation structure compared to the case of the RGB space, there should be more information in the individual hue, saturation and value dimensions, without considering the correlation. Indeed, in Table 7 the CCRs and RRs for different GGDs modeling the HSV coordinates independently (UV) are compared to the case where correlation between all coordinates is taken into account (MV-3). As is to be expected, the classification and retrieval performance of the UV case is better compared to UV modeling in RGB space. On the contrary, the MV-3 results are worse for HSV relative to RGB space. Nevertheless, the best results are still obtained in the multivariate context. We also noted that in HSV space it is more difficult to fit a model with variable shape parameter, due to the extreme leptokurtic behavior of particularly the hue coordinate. This is the cause of the reduced classification perfor-

| Measure |         | Model |      |
|---------|---------|-------|------|
|         |         | Gray  | MV-3 |
| KLD     | Gauss   | 60.5  | 81.1 |
|         | Laplace | 61.1  | –    |
|         | GGD     | 68.8  | –    |
| GD      | Gauss   | 59.8  | 80.8 |
|         | Laplace | 62.5  | 81.3 |
|         | GGD     | 67.8  | 83.0 |

Table 6: Correct classification rates (%) using different models (Gray and MV-3) for three wavelet scales, using the KLD and GD as similarity measures in a database of 640 color texture images, with added Gaussian noise (SNR = 1).

mance of the models with variable shape parameter. This effect is even more apparent with the GD than using the KLD, since the GD with variable shape parameter is an approximation. Comparing with the results obtained in the RGB space, we conclude that for multivariate modeling using an MGGD, the RGB space provides a better representation than a less correlated space such as the HSV space.

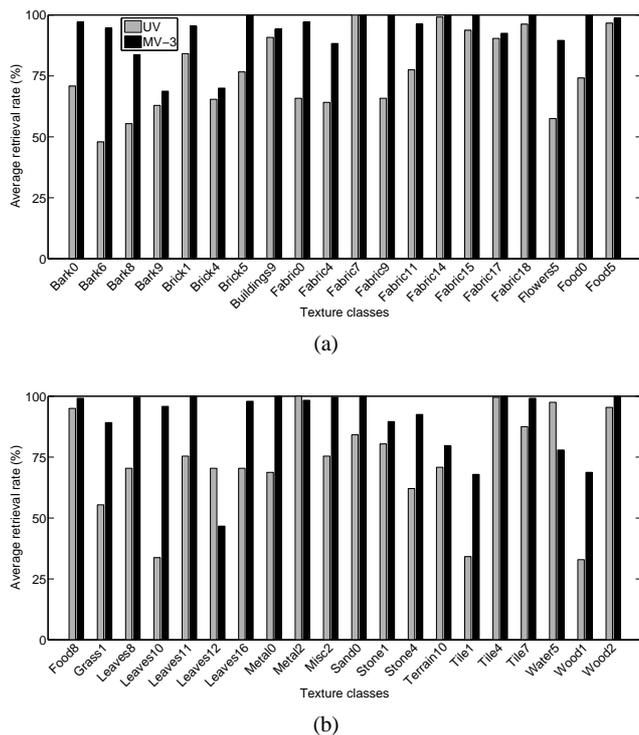


Fig. 6: Comparison of the average retrieval rates over each texture class in the database, between univariate modeling (UV) of the colors individually on the one hand and modeling of the three correlated color bands (MV-3) on the other hand. Three wavelet scales were used, modeled by Laplace distributions, and the GD was applied as a similarity measure.

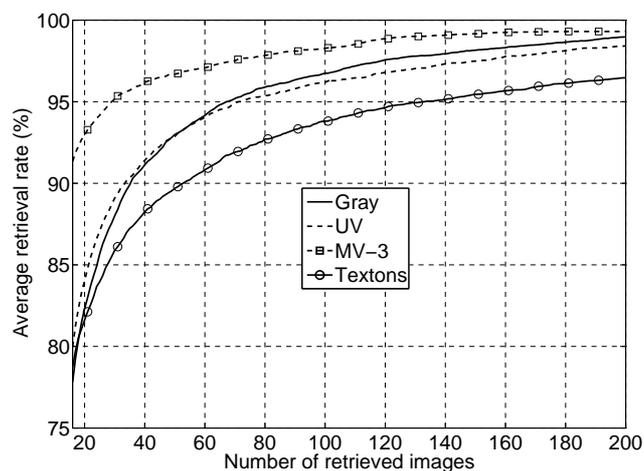


Fig. 7: Average retrieval rate as a function of the number of top retrieved images (excluding the query image) for gray-level modeling (Gray, variable  $\beta$ ), univariate modeling of the colors individually (UV, variable  $\beta$ ) and modeling of the three correlated color bands (MV-3, Laplace model) using the GD on three wavelet scales. The rates obtained via the texton method are also shown.

| Measure |         | Classification |      | Retrieval |      |
|---------|---------|----------------|------|-----------|------|
|         |         | Model          |      |           |      |
|         |         | UV             | MV-3 | UV        | MV-3 |
| KLD     | Gauss   | 85.8           | 91.6 | 70.7      | 79.4 |
|         | Laplace | 88.1           | –    | 75.1      | –    |
|         | GGD     | 90.5           | –    | 73.5      | –    |
| GD      | Gauss   | 87.3           | 93.9 | 72.6      | 81.3 |
|         | Laplace | 90.6           | 94.5 | 77.2      | 83.9 |
|         | GGD     | 82.5           | 72.5 | 64.7      | 59.5 |

Table 7: Correct classification rates and average retrieval rates (%) using different models (UV and MV-3) for three wavelet scales, using the KLD and GD as similarity measures in a database of 640 color texture images in HSV space.

| Measure |         | Model |      |
|---------|---------|-------|------|
|         |         | Gray  | MV-3 |
| KLD     | Gauss   | 28.0  | 60.1 |
|         | Laplace | 32.5  | –    |
|         | GGD     | 35.8  | –    |
| GD      | Gauss   | 30.1  | 62.4 |
|         | Laplace | 33.9  | 68.8 |
|         | GGD     | 37.4  | 64.2 |

Table 8: Average retrieval rates (%) using different models (Gray and MV-3) for a single wavelet scale, using the KLD and GD as similarity measures in a texture retrieval experiment. The database consists of 100,000 color images, every image containing multiple textures.

#### 4.3.5 Large database of multiple texture images

Finally, Table 8 contains the success rates, as defined above, for the retrieval experiment consisting of 1000 queries of random texture patches in a database of 100,000 multiple texture images. Only gray-level textures and the corresponding color images were treated and one wavelet scale was used. Again, similar trends can be noticed as in the experiments on 640 images. This shows that our methods can be applied successfully in realistic experimental settings involving large databases.

#### 4.3.6 Practicalities and future research

We next would like to formulate several considerations that are relevant to the implementation in practical applications of the techniques introduced in this paper. At the same time we suggest various possible improvements to the methods as well as some topics deserving further study. First, we have stressed upon the additional computational demands associated with some of the improvements in classification rate observed in our study. Although an elevated computational

complexity of the similarity measurement can be prohibitive in real-time applications of e.g. a retrieval system, this does not exclude the usefulness in other or related applications of even the most computationally demanding methods applied in the present work. For instance, a practical retrieval mechanism might start querying the database using a computationally lightweight method (e.g. by means of the KLD on the gray-level component modeled by a GGD), and subsequently, in a reduced set of retained images, refine the search by means of a more advanced technique (e.g. using the GD and a trivariate Laplace model). The GD in conjunction with complicated multivariate data models can be fruitful as well in a database search using key images. Moreover, the higher classification rates obtained with the GD and an advanced model, may also be an indication of good, and perhaps more distinctive performance in other applications, where computational load is less an issue. For example, it was noted above that for two distributions that are close on the probabilistic manifold, the GD and KLD are comparable similarity measures. However, the more distant two distributions, the larger the disparity between the GD and the KLD. Therefore, it would be worth studying an application where mainly large distance are involved (e.g. a classification problem with extensive but mutually close clusters in feature space), and check the relative performance of the GD and KLD, compared to their performance observed in the present application or database.

Another possible topic for future research concerns the classification results using a distribution with variable shape parameter, relative to fixed shape parameter models. Indeed, the variable shape parameter induces a significant additional model flexibility and this is in fact reflected by the substantial increase of the classification rate for univariate models. However, the reason why a similar large improvement is not observed in the multivariate case (except in the experiment with added Gaussian noise), at least on our data set, is not entirely clear, although we have identified several possibilities above. Similarly, the lack of an increased performance when using a more accurate higher-degree polynomial approximation to the geodesic coordinate functions, is definitely a matter for further investigation. First, this behavior could be different in other applications. Second, the effect of a linear geodesic approximation on classification effectiveness should be studied in more detail. Third, the hypothesis, which we have stated, should be examined that the fact that multivariate models with variable shape parameter do not result in a significantly enhanced classification performance, is related to a suboptimal estimation of the distribution. This, in turn, could trigger the search for improved algorithms for parameter estimation, or even other distribution families that provide a still better description of multivariate wavelet data, and to which the methods outlined in this work can be applied.

As a final recommendation to practitioners as far as classification performance is concerned, we would like to propose to use the GD (with linear approximation for the coordinate functions) together with the GGD with variable shape parameter in the case of gray-level texture discrimination. For the trivariate case where the correlation between the spectral bands is modeled, we recommend the GD together with the Laplacian distribution.

As an aside we still mention that for the calculation of the GD between the distributions modeling the images in the present databases, it was verified that the resulting geodesic distances are symmetric in the order of the arguments. In addition, for the case of a Gaussian and Laplacian, the numerically computed GDs using linear geodesic coordinate functions correspond very well to the exact results obtained via the analytic expressions for the GD. The reason is that for the distributions with fixed shape parameter, in the coordinate system used in this work, the linear approximation is exact, see Section 3.1.

## 5 Conclusion

In this paper, we have studied the measurement of multi-component image texture similarity through a statistical approach in the wavelet domain. We applied this in several experiments for the classification of color textures in a database, relying solely on the actual image content. We used a  $k$ -nearest neighbor classifier evaluated via the leave-one-out method. We chose image retrieval as a specific application of the classifier. We proposed to profit by the information residing in the rich spectral band correlation structure, by jointly modeling, through a heavy-tailed multivariate generalized Gaussian distribution (MGGD), the wavelet detail coefficients corresponding to the respective spectral bands in an image. We tested several methods for fitting an MGGD to multiband wavelet data, concluding that the moment method provides a good approximation to the wavelet histograms, with the possibility of subsequent fine-tuning through recursive solution of the ML equations.

A second accomplishment of the present work is the use in a series of classification and retrieval experiments of the geodesic distance (GD) as a similarity measure between MGGDs (hence between the corresponding images). The GD has several interesting advantages compared to the Kullback-Leibler divergence (KLD), a traditional measure of similarity between probability distributions. An initial set of experiments was conducted using a small image database serving as a bench-mark system. Models of varying complexity were employed for characterizing the wavelet data. The classification effectiveness for all models, using both the GD and KLD, was evaluated by on the one hand the actual correct classification rates, and on the other hand by

the computational complexity. A substantial gain in classification efficiency was obtained when considering the inter-band correlation (both pairwise and trivariate) as opposed to univariate modeling. In the univariate case, the use of a GGD with variable shape parameter, compared to a Gaussian or a Laplacian, produced a distinct benefit in terms of correct classification rate. Another important observation is the clearly superior classification performance of the GD in comparison with the KLD, particularly for multivariate distributions. The highest correct classification rates, in combination with a manageable computational complexity, were obtained using the GD on trivariate Laplace distributions. For gray-scale images, our method performs better than a state of the art technique employing image patch exemplars.

Next, the impact of perturbation of the images through additive Gaussian noise was examined and the KLD was found to be slightly more robust against the noise than the GD. Nevertheless, the best classification results were still obtained using the multivariate Laplace distribution or the MGGD with variable shape parameter, in conjunction with the GD.

We also repeated the experiments in the HSV color space, which is a less correlated representation compared to the RGB space. Although the classification results via univariate modeling improve in a less correlated space such as HSV, the efficiency of multivariate models deteriorates considerably. Hence for multivariate modeling the RGB space is a better alternative.

We then performed a more realistic set of texture retrieval experiments in a significantly larger database with images containing multiple colored texture patches. Similar tendencies were observed as in the initial series of experiments on the smaller database, indicating that our methods are scalable and suitable in realistic situations as well.

Finally, we identified some topics for future research, including the relative performance of the GD and KLD, as well as the impact of variable versus fixed MGGD shape parameter models. To this end it would be advantageous to investigate the behavior of the techniques developed in this work on other data sets and in other applications.

**Acknowledgements** This work was partially funded by the Fund for Scientific Research–Flanders (FWO-Vlaanderen).

## References

- Abramowitz M, Stegun I (1965) *Handbook of Mathematical Functions*. Dover Publications, New York
- Amari S, Nagaoka H (2000) *Methods of information geometry*, Transactions of mathematical monographs, vol 191. American Mathematical Society, New York
- Atkinson C, Mitchell A (1981) Rao's distance measure. *Sankyā: The Indian Journal of Statistics* 48:345–365
- Benazza-Benyahia A, Pesquet JC (2005) Building robust wavelet estimators for multicomponent images using Steins' principle. *IEEE Transactions on Image Processing* 14(11):1814–1830
- Berkane M, Oden K, Bentler P (1997) Geodesic estimation in elliptical distributions. *J Multivariate Anal* 63(1):35–46
- Berman A, Shapiro L (1999) A flexible image database system for content-based retrieval. *Comput Vis Image Underst* 75(1–2):175–195
- Bonet J, Viola P (1998) Texture recognition using a non-parametric multi-scale statistical model. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, pp 641–647
- Boubchir L, Fadili J (2005) Multivariate statistical modeling of images with the curvelet transform. In: *Proceedings of the 8<sup>th</sup> International Symposium on Signal Processing and its Applications*, pp 747–750
- Burbea J (1986) Informative geometry of probability spaces. *Expo Math* 4:347–378
- Burbea J, Rao C (1982) Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. *J Multivariate Anal* 12:575–596
- Burkhard W, Keller R (1973) Some approaches to best-match file searching. *Commun ACM* 16(4):230–236
- Calvo M, Oller J (2002) A distance between elliptical distributions based in an embedding into the Siegel group. *J Comput Appl Math* 145(2):319–334
- Castano-Moraga C, Lenglet C, Deriche R, Ruiz-Alzola J (2007) A Riemannian approach to anisotropic filtering of tensor fields. *Signal Processing* 87(2):263–276
- Čenkov N (1982) *Statistical decision rules and optimal inference*, Translations of Mathematical Monographs, vol 53. American Mathematical Society, Providence, RI
- Chang CI (2007) *Hyperspectral Data Exploitation: Theory and Applications*. Wiley-Interscience, New York
- Cho D, Bui T (2005) Multivariate statistical modeling for image denoising using wavelet transforms. *Signal Process* 20(1):77–89
- Cramér H (1946) A contribution to the theory of statistical estimation. *Skandinavisk Aktuarietidskrift* 29:85–94
- Datta R, Joshi D, Li J, Wang J (2008) *Image retrieval: Ideas, influences, and trends of the new age*. *ACM Computing Surveys* 40(2)
- Daugman J (1980) Two-dimensional spectral analysis of cortical receptive field profile. *Vis Res* 20(10):847–856
- De Backer S, Pizurica A, Huysman B, Philips W, Scheunders P (2008) Denoising of multispectral images using wavelet least-squares estimators. *Image and Vision Computing* 26(7):1038–1051
- Do M, Vetterli M (2002) Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Trans Image Process* 11(2):146–158

- do Carmo M, Flaherty F (1992) *Riemannian Geometry*. Birkhäuser, Boston, MA
- Fang KT, Zhang YT (1990) *Generalized Multivariate Analysis*. Springer-Verlag, Berlin
- Fang KT, Kotz S, Ng KW (1990) *Symmetric multivariate and related distributions*, Monographs on Statistics and Applied Probability, vol 36. Chapman and Hall, New York
- Gómez E, Gómez-Villegas M, Marín J (1998) A multivariate generalization of the power exponential family of distributions. *Commun Statist—Theory Meth* 27(3):589–600
- Image★After (2010) online at <http://www.imageafter.com/>
- James A (1973) The variance information manifold and the functions on it. In: Krishnaiah P (ed) *Multivariate Analysis III*, Academic Press, New York, pp 157–169
- Kass R, Vos P (1997) *Geometrical foundations of asymptotic inference*. Wiley Series in Probability and Statistics, Wiley-Interscience, New York
- Kullback S (1968) *Information theory and statistics*. Dover Publications, New York
- Lehmann E, Casella G (2003) *Theory of Point Estimation*, 2nd edn. Springer Texts in Statistics, Springer-Verlag, New York
- Lenglet C, Rousson M, Deriche R (2006a) DTI segmentation by statistical surface evolution. *IEEE Trans Med Imaging* 25(6):685–700
- Lenglet C, Rousson M, Deriche R, Faugeras O (2006b) Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. *J Math Imaging Vis* 25(3):423–444
- Mallat S (1989) A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 11(7):674–692
- Mallat S (1999) *A wavelet tour of signal processing*, 2nd edn. Academic Press, New York
- Manjunath B, Ma W (1996) Texture features for browsing and retrieval of image data. *IEEE Trans Pattern Anal Mach Intell* 18(8):837–842
- Manjunath B, Ohm JR, Vasudevan V, Yamada A (2001) Color and texture descriptors. *IEEE Trans Circ Syst Video Technol* 11(6):703–715
- Mardia K, Kent J, Bibby J (1982) *Multivariate Analysis*. Academic Press, London
- Marsaglia G (1972) Choosing a point from the surface of a sphere. *Ann Math Stat* 43(2):645–646
- Mathiassen J, Skavhaug A, Bo K (2002) Texture similarity measure using Kullback-Leibler divergence between gamma distributions. In: *Proceedings of the European Conference on Computer Vision*, Copenhagen, vol 2352, pp 19–49
- Mercier G, Lennon M (2002) On the characterization of hyperspectral texture. In: *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Toronto, vol 5, pp 2584–2586
- Micak M, Kozintsev I, Ramchandran K, Moulin P (1999) Low-complexity image denoising based on statistical modeling of wavelet coefficients. *IEEE Signal Process Lett* 6(12):300–303
- MIT Vision and Modeling Group (2010) *Vision texture*. online at <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/>
- Mitchell A (1989) The information matrix, skewness tensor and  $\alpha$ -connections for the general multivariate elliptic distribution. *Ann Inst Statist Math* 41(2):289–304
- Muller M (1959) A note on a method for generating points uniformly on  $N$ -dimensional spheres. *Commun ACM* 2(4):19–20
- Murray M, Rice J (1993) *Differential geometry and statistics*. Monographs on Statistics and Applied Probability, 48, Chapman and Hall, New York
- O’Neill B (1982) *Elementary Differential Geometry*, second revised edn. Academic Press, New York
- Pižurica A, Philips W (2006) Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising. *IEEE Transactions on Image Processing* 15(3):654–665
- Rao C (1945) Information and accuracy attainable in the estimation of statistical parameters. *Bull Calcutta Math Soc* 37:81–89
- Scheunders P, De Backer S (2007) Wavelet denoising of multicomponent images, using Gaussian Scale Mixture models and a noise-free image as priors. *IEEE Trans Image Process* 16(7):1865–1872
- Skovgaard L (1981) A Riemannian geometry of the multivariate normal model. Tech. Rep. 81/3, Statistical Research Unit, Danish Medical Research Council, Danish Social Science Research Council
- Skovgaard L (1984) A Riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics* 11(4):211–223
- Stephens M (1964) The testing of unit vectors for randomness. *J Am Stat Assoc* 59(305):160–167
- Synyavskyy A, Voloshynovskiy S, Prudyus I (2001) Wavelet-based map image denoising using provably better class of stochastic I.I.D. image models. *Facta Universitatis (Series: Electronics and Energetics)* 14(3):375–385
- The Mathworks (2008) Natick, MA, [www.mathworks.com](http://www.mathworks.com)
- Tzagkarakis G, Beferull-Lozano B, Tsakalides P (2006) Rotation-invariant texture retrieval with gaussianized steerable pyramids. *IEEE Trans Image Process* 15(9):2702–2718
- Van de Wouwer G, Scheunders P, Van Dyck D (1999) Statistical texture characterization from discrete wavelet representations. *IEEE Trans Image Process* 8(4):592–598
- Varanasi MK, Aazhang B (1989) Parametric generalized Gaussian density estimation. *J Acoust Soc Am*

---

86(4):1404–1415

- Varma M, Zisserman A (2003) Texture classification: are filter banks necessary? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol 2, pp 691–698
- Varma M, Zisserman A (2009) A statistical approach to material classification using image patch exemplars. *IEEE Trans Pattern Anal Mach Intell* 31(11):2032–2047
- Verdoolaege G, De Backer S, Scheunders P (2008) Multi-scale colour texture retrieval using the geodesic distance between multivariate generalized Gaussian models. In: Proceedings of the IEEE International Conference on Image Processing, San Diego, CA, pp 169–172
- Verdoolaege G, Rosseel Y, Lambrechts M, Scheunders P (2009) Wavelet-based colour texture retrieval using the Kullback-Leibler divergence between bivariate generalized Gaussian models. In: Proceedings of the IEEE International Conference on Image Processing, Cairo, pp 265–268
- Watson G, Williams E (1956) On the construction of significance tests on the circle and the sphere. *Biometrika* 43(3–4):344–352