

Texture Similarity Measure Using Kullback-Leibler Divergence between Gamma Distributions

John Reidar Mathiassen¹, Amund Skavhaug², and Ketil Bø^{3*}

¹ SINTEF Fisheries and Aquaculture, N-7465 Trondheim, Norway
`John.R.Mathiassen@fish.sintef.no`

² Department of Engineering Cybernetics (ITK), NTNU, N-7491 Trondheim, Norway
`Amund.Skavhaug@itk.ntnu.no`

³ Department of Computer and Information Science (IDI), NTNU, N-7491
Trondheim, Norway
`ketilb@idi.ntnu.no`

Abstract. We propose a texture similarity measure based on the Kullback-Leibler divergence between gamma distributions (KLGamma). We conjecture that the spatially smoothed Gabor filter magnitude responses of some classes of visually homogeneous stochastic textures are gamma distributed. Classification experiments with disjoint test and training images, show that the KLGamma measure performs better than other parametric measures. It approaches, and under some conditions exceeds, the classification performance of the best non-parametric measures based on binned marginal histograms, although it has a computational cost at least an order of magnitude less. Thus, the KLGamma measure is well suited for use in real-time image segmentation algorithms and time-critical texture classification and retrieval from large databases.

1 Introduction

Measuring the similarity between textures is an important task in low-level computer vision. A similarity measure for textures is essential for texture classification, supervised- and unsupervised image segmentation. The motivation of our research has been to find a similarity measure that is both low in computational cost and is statistically justified for classification purposes, and is therefore well suited for use in both texture classification/retrieval and real-time texture segmentation. Many similarity measures have been developed, based on various feature extraction schemes ([20],[21],[22],[23], [24]). In this paper we will focus on texture similarity between features extracted using Gabor filters. Observations of many textures, lead us to conjecture that some classes of visually homogeneous

* The authors would like to thank Henrik Schumann-Olsen for his useful comments and discussions. We also thank the anonymous reviewers for their feedback. This research was supported by The Research Council of Norway.

stochastic textures have gamma distributed spatially smoothed Gabor magnitude responses. We introduce the KLGamma texture similarity measure, which is based on the Kullback-Leibler divergence between gamma distributions.

We perform classification experiments on 35 texture images. These experiments are done to evaluate classification performance on image blocks of various sizes. Testing is done with several parametric and non-parametric similarity measures, with both disjoint and non-disjoint test and training sets. All the tested measures assume independence between features to avoid the computational cost of estimating and using the feature covariance matrix in the similarity measure. In addition to the proposed method, the following parametric methods are tested: weighted mean variance (WMV) [14], weighted nearest neighbour (WNN) and the Kullback-Leibler divergence between Gaussians (KLGauss). We also test several non-parametric measures applied to binned marginal probability distributions: Kullback-Leibler divergence (KL), χ^2 statistic (Chi) and Jeffrey divergence (JD) [15],[16].

In classification experiments with disjoint test and training sets, KLGamma performs better than the other tested parametric measures and KL. For block sizes smaller than 32x32 pixels KLGamma outperforms all other measures, and for larger block sizes it is slightly worse than JD and Chi.

We analyze the computational cost of all the tested similarity measures and find that KLGamma is somewhat slower than WNN, and faster than the other parametric measures. Compared to non-parametric measures, which have similar classification performance, KLGamma is at least an order of magnitude less computationally intensive. In tests with disjoint test and training sets, KLGamma achieves slightly better results than the second best parametric measure, KLGauss, while being computationally simpler.

Research by Puzicha et.al. ([15] and [16]) has shown that the non-parametric Chi and JD measures outperform other non-parametric measures and the parametric WMV measure in texture retrieval tests. In their experiments, the extracted texture features were Gabor filter magnitude responses. It has been shown [2] that spatially smoothing the Gabor filter magnitude responses can improve classification performance, and we will therefore do this kind of smoothing in our tests. Our main contributions are as follows:

- We introduce the KLGamma texture similarity measure based on the Kullback-Leibler divergence between gamma distributions.
- We conjecture that the spatially smoothed Gabor filter magnitude responses of some classes of visually homogeneous stochastic textures are gamma distributed.
- We compare the classification performance of several non-parametric and parametric similarity measures on spatially smoothed Gabor features.

In addition to being justified from a statistical point of view, the combination of speed and classification performance implies that the KLGamma measure is well suited to real-time texture segmentation and texture classification/retrieval.

In the next section we present texture feature extraction using spatially smoothed Gabor filter magnitude responses. We then introduce all the tested

similarity measures in the context of maximum-likelihood classification. Then we present results that show how well the gamma distribution fits the probability density function of the extracted KL features, and then present results from classification tests that compare the KLGamma similarity measure with other parametric and non-parametric measures. Moreover, we present a comparison of the computational cost of the various similarity measures. Finally we discuss the results of this paper and offer some conclusions and suggestions for future work.

2 Texture Feature Extraction

2.1 Gabor Filtering

When extracting image features for use in texture analysis, it seems wise to look to the way texture is analyzed in the human visual cortex. Marcelja [4] and Daugman [5], [6] showed that the responses of simple cells in the visual cortex can be modeled by Gabor filter response. The 2D Gabor functions proposed by Daugman also achieve the theoretical resolution limit in the conjoint spatial-spatial frequency domain. Thus it is understandable that Gabor filters are commonly used in the feature extraction step of texture analysis. Based on the Gabor functions proposed by Daugman, Lee[1] derived a family of Gabor wavelets that satisfy the neurophysiological constraints for simple cells.

$$\psi(x, y, \omega_0, \theta) = \frac{\omega_0}{\sqrt{2\pi\kappa}} e^{-\frac{\omega_0^2}{8\kappa^2}(x'^2+y'^2)} \cdot \left[e^{i(\omega_0 x')} - e^{-\frac{\kappa^2}{2}} \right] \tag{1}$$

Here $x' = x \cos \theta + y \sin \theta$ and $y' = -x \sin \theta + y \cos \theta$. ω_0 is the radial frequency in radians per unit length and θ is the wavelet orientation in radians. The wavelet is centered at $(x = 0, y = 0)$ and $\kappa \approx \pi$ for a frequency bandwidth of one octave. The term $e^{-\frac{\kappa^2}{2}}$ in (1) is the d.c. response of the wavelet, and is subtracted to achieve invariance with respect to image brightness.

For simplicity, let $\psi_{\omega_0, \theta}$ denote the Gabor wavelet with parameters ω_0 and θ . Let $\psi_{\omega_0, \theta}^r$ be the real part of $\psi_{\omega_0, \theta}$ and $\psi_{\omega_0, \theta}^i$ be the imaginary part, and let I_{xy} be the image centered at (x, y) to be filtered by $\psi_{\omega_0, \theta}$. Then the magnitude response of I_{xy} filtered by $\psi_{\omega_0, \theta}$ is

$$|\psi_{\omega_0, \theta} * I_{xy}| = \sqrt{\left(\psi_{\omega_0, \theta}^r * I_{xy}\right)^2 + \left(\psi_{\omega_0, \theta}^i * I_{xy}\right)^2} \tag{2}$$

where $*$ is the convolution operator. Experiments have shown [2] that using this magnitude response achieves better texture separability than using full-wave rectification or the real component only. This is most likely due to the fact that optimal conjoint resolution in the spatial-spatial frequency domain is achieved with the complex form (quadrature-phase pair) of the Gabor filter [1]. Pollen and Ronner’s [7] discovery, that the simple cells in the visual cortex exist in quadrature-phase pairs, suggests that the design of these cells might indeed be optimal with respect to the conjoint resolution in the spatial-spatial frequency domains.

2.2 Spatial Smoothing

The performance of Gabor filters for texture analysis is known to improve by performing Gaussian smoothing of the Gabor filter magnitude response images [2]. If $g(x, y, \omega_0, \theta) = \frac{\omega_0}{\sqrt{2\pi\kappa}} e^{-\frac{\omega_0^2}{8\kappa^2}(x'^2+y'^2)}$ is the Gaussian term in (1), the magnitude response images are smoothed with oriented Gaussian filter $g(\gamma x, \gamma y, \omega_0, \theta)$. We denote the value at (x, y) of a smoothed magnitude image by

$$c_{xy, \omega_0, \theta, \gamma} = g(\gamma x, \gamma y, \omega_0, \theta) * |\psi_{\omega_0, \theta} * I_{xy}|. \quad (3)$$

3 Similarity Measures

In this section we will present the similarity measures. To understand the reason for choosing a given similarity measure, one must look at the underlying assumptions that are made in deducing the measure. Therefore we will describe the maximum-likelihood classification principle, and the similarity measures that result from making assumptions on the type of feature probability distributions. We then define the asymptotically (for large sample sizes) optimal similarity measures for Gaussian distributions, gamma distributions and binned marginal histogram distributions. For binned histograms we will also present some further measures used in [15] and [16]. Finally we will describe the parametric WMV similarity measure proposed by Manjunath and Ma [14] which was compared with several non-parametric methods in [15] and [16].

3.1 Maximum-Likelihood Texture Classification

Assume that a texture is described by a featurespace \mathcal{F} . Given this featurespace we define a mapping $g : \mathcal{F} \rightarrow C = \{1, \dots, K\}$ from the featurespace \mathcal{F} to the set C of possible texture class labels. The goal we set for a texture classification system is to *minimize the probability of classification error*. This can be formalized as the minimization of $p(g(\mathbf{x}) \neq y)$, with \mathbf{x} being a set of feature observations from a texture sample of class y . The optimal mapping g for this problem formulation is [13] the Bayes classifier

$$g^*(\mathbf{x}) = \arg \max_i p(\mathbf{x}|y=i) P(y=i) \quad (4)$$

where $p(\mathbf{x}|y=i)$ is the likelihood function for the i^{th} class, and $P(y=i)$ is its prior probability.

When all texture classes have the same prior probability $P(y=i) = 1/K$ (4) reduces to the maximum-likelihood (ML) classifier [12]

$$g(\mathbf{x}) = \arg \max_i \frac{1}{M} \sum_{j=1}^M \log p(\mathbf{x}_j|y=i). \quad (5)$$

It is in general computationally expensive to use (5) for large sample sizes M .

Kullback-Leibler Divergence. In [12] it is shown that for large values of M we can apply the law of large numbers to (5), resulting in

$$g(\mathbf{x}) \xrightarrow{M \rightarrow \infty} \arg \max_i E_q [\log p(\mathbf{x}|y = i)] = \arg \min_i KL(q; p^{(i)}) \quad (6)$$

where $KL(q; p_i)$ is the Kullback-Leibler divergence (KL) between the query density q and the density $p^{(i)}$ associated with texture class i . This shows that KL is the asymptotic limit of the ML classification criteria. If \mathcal{F} has N dimensions, and these dimensions are independent of each other, the Kullback-Leibler divergence is $KL(q; p^{(i)}) = \sum_{j=1}^N KL(q_j; p_j^{(i)})$. This follows from (6) assuming $p(\mathbf{x}) = \prod_{j=1}^N p_j(x_j)$. Here j denotes the feature dimension.

3.2 Gaussian Distribution

The univariate Gaussian distribution is given by $p_{Gauss}(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-0.5[(x-\mu)/\sigma]^2}$. If the N dimensions of \mathcal{F} are independent and Gaussian distributed we have from [3] the following Kullback-Leibler divergence

$$KL_{Gauss}(q; p) = \frac{1}{2} \sum_{j=1}^N \left(\log \left(\frac{\sigma_j^{(p)}}{\sigma_j^{(q)}} \right)^2 + \left(\frac{\mu_j^{(q)} - \mu_j^{(p)}}{\sigma_j^{(p)}} \right)^2 + \left(\frac{\sigma_j^{(q)}}{\sigma_j^{(p)}} \right)^2 - 1 \right) \quad (7)$$

A simplification of (7) gives us the weighted nearest neighbour classifier (WNN)

$$D_{WNN}(q; p) = \sum_{j=1}^N \left(\frac{\mu_j^{(q)} - \mu_j^{(p)}}{\sigma_j^{(p)}} \right)^2 \quad (8)$$

The condition under which this simplification is justified is $\sigma_j^{(q)} \approx \sigma_j^{(p)}$. We observe that the WNN measure is equivalent to the Mahalanobis measure [12] if the feature covariance matrix is diagonal.

3.3 Gamma Distribution

The gamma distribution is given by $p_{Gamma}(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$ where $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt, z > 0$.

Parameter Estimation. There are two main methods of calculating estimates of the parameters α and β of the gamma distribution: the moment method and maximum-likelihood (ML) method. Using the moment method [9] we obtain the parameters by $\hat{\beta} = \frac{\hat{\sigma}^2}{\hat{\mu}}$ and $\hat{\alpha} = \frac{\hat{\mu}}{\hat{\beta}}$, where $\hat{\mu}$ and $\hat{\sigma}$ denote the estimated mean and standard deviation of the distribution. Using the ML estimation method [10] we obtain parameter estimates using

$$\hat{\alpha}\hat{\beta} = \langle X \rangle = \frac{1}{n} \sum_{i=1}^n X_i \quad (9)$$

$$\log \hat{\alpha} - \Psi(\hat{\alpha}) = \langle \log X \rangle - \log \langle X \rangle = \frac{1}{n} \sum_{k=1}^n \log X_k - \log \langle X \rangle \quad (10)$$

where $\langle X \rangle$ and $\langle \log X \rangle$ are the mean and mean logarithm of the distribution given by a set of independent observations $X = \{X_1, \dots, X_n\}$, and $\Psi(\cdot)$ is the digamma function, i.e. $\Psi(z) = \Gamma'(z) / \Gamma(z)$.

Kullback-Leibler Divergence between Gamma Distributions. A closed-form expression for the KL-divergence between two Gamma distributions is given by [3] and results in the following expression

$$KL_{Gamma}(q; p) = \sum_{j=1}^N \left[(\alpha_j^{(q)} - 1) \Psi(\alpha_j^{(q)}) - \log \beta_j^{(q)} - \alpha_j^{(q)} - \log \Gamma(\alpha_j^{(q)}) + \log \Gamma(\alpha_j^{(p)}) + \alpha_j^{(p)} \log \beta_j^{(p)} - (\alpha_j^{(p)} - 1) \left(\Psi(\alpha_j^{(q)}) + \log \beta_j^{(q)} \right) + \frac{\alpha_j^{(q)} \beta_j^{(q)}}{\beta_j^{(p)}} \right] \quad (11)$$

Without any loss of classification performance we can simplify (11) for use in (6), leading to the following similarity measure

$$D_{KLGamma}(q; p) = \sum_{j=1}^N \left[\log \Gamma(\alpha_j^{(p)}) + \alpha_j^{(p)} \log \beta_j^{(p)} - \alpha_j^{(p)} \left(\Psi(\alpha_j^{(q)}) + \log \beta_j^{(q)} \right) + \frac{\alpha_j^{(q)} \beta_j^{(q)}}{\beta_j^{(p)}} \right] \quad (12)$$

which we introduce in this paper for use as a texture similarity measure.

3.4 Binned Histograms

Marginal probability density functions can be modeled as binned histograms. For a given feature dimension j we then have the following empirical distribution $p_j(x_j) = \left(\left\lfloor \frac{x_j}{\Delta l_j} \right\rfloor \right)$, where $\lfloor \cdot \rfloor$ denotes the floor operator, and $p_j(z) = 0$ if $z \notin \{0, \dots, L-1\}$. L is the number of bins and Δl_j is the bin size.

Kullback-Leibler Divergence. The Kullback-Leibler divergence between two empirical distributions is

$$D_{KL}(q, p) = \sum_{j=1}^N \sum_{l=0}^{L-1} q_j(l) \log \frac{q_j(l)}{p_j(l)}. \quad (13)$$

Appendix A shows that when the feature probabilities are modelled with binned histograms, $g(\mathbf{x}) = \arg \min_i D_{KL}(q, p^{(i)})$ is exactly the maximum-likelihood classifier.

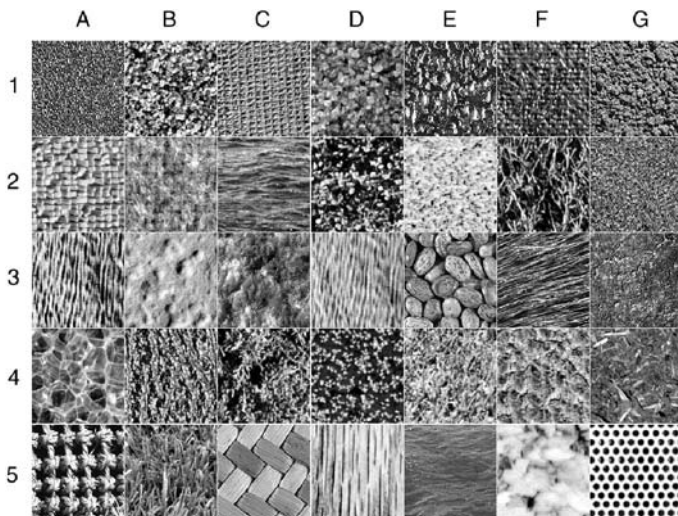


Fig. 1. 128x128 pixel patches of the textures used in the experiments. The textures are displayed in order of goodness-of-fit with the gamma distribution model, with the fit decreasing from left to right, and top to bottom

Jeffrey Divergence. A symmetrized version of the Kullback-Leibler divergence is the Jeffrey divergence [15] which is more numerically stable for empirical distributions. For binned histograms it is

$$D_{JD}(q, p) = \sum_{j=1}^N \sum_{l=0}^{L-1} \left[q_j(l) \log \frac{q_j(l)}{q_j(l) + p_j(l)} + p_j(l) \log \frac{p_j(l)}{q_j(l) + p_j(l)} \right]. \quad (14)$$

χ^2 Statistic. An approximation to the Kullback-Leibler divergence is the χ^2 statistic [12]. A symmetrized version of this similarity measure is used in [15],[16] and is defined for binned histograms as follows

$$D_{\chi^2}(q, p) = \sum_{j=1}^N \sum_{l=0}^{L-1} \frac{[q_j(l) - p_j(l)]^2}{q_j(l) + p_j(l)}. \quad (15)$$

3.5 Weighted Mean Variance (WMV)

A parametric similarity measure used for comparing textures is the weighted mean variance (WMV) measure proposed by Manjunath and Ma [14]. Let $\mu_j^{(i)}$ and $\sigma_j^{(i)}$ be the mean and standard deviation of feature dimension j . The superscript (i) denotes the texture class. The WMV measure between two textures q and p is then defined as

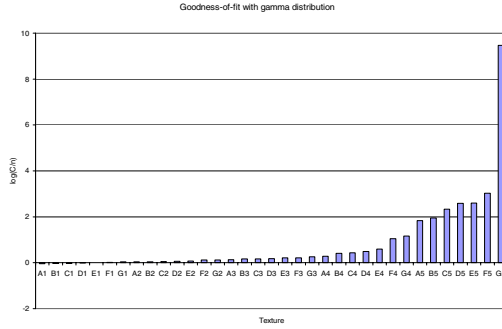


Fig. 2. Goodness-of-fit with gamma distribution

$$D_{W MV}(q; p) = \sum_{j=1}^N \left(\left| \frac{\mu_j^{(q)} - \mu_j^{(p)}}{\alpha(\mu_j)} \right| + \left| \frac{\sigma_j^{(q)} - \sigma_j^{(p)}}{\alpha(\sigma_j)} \right| \right) \tag{16}$$

where $\alpha(\mu_j)$ and $\alpha(\sigma_j)$ are the standard deviations of the respective features over the entire database, and are used to normalize the individual feature components.

4 Results

4.1 Validity of Gamma Distribution Model

In this section we illustrate how well the gamma distribution model fits the distribution of the extracted texture features. As a goodness-of-fit measure we use the χ^2 statistic [18]. With N feature dimensions it is $C = n \sum_{j=1}^N \sum_{l=1}^L \frac{(p_j(l) - \pi_j(l))^2}{\pi_j(l)}$, where n is the number of samples, L is the number of bins, $p_j(l)$ the observed probability of bin l and $\pi_j(l)$ the expected probability of bin l assuming the feature is gamma distributed. $\pi_j(l) = F_j(l\Delta l_j) - F_j((l-1)\Delta l_j)$, where Δl_j is the bin size and $F_j(z)$ is the cumulative gamma distribution. Given a gamma distribution with parameters α_j and β_j , $F_j(z) = \frac{\Gamma_x(\alpha_j, z/\beta_j)}{\Gamma(\alpha_j)}$, $\Gamma_x(\alpha, t) = \int_0^x e^{-t} t^{\alpha-1} dt$ [19]. The goodness-of-fit with the gamma distribution for a given texture is calculated for feature distributions extracted from the entire 512x512 texture image.

From figures 1 and 2 we see that the textures that best fit the gamma distribution model are visually homogeneous and appear to be stochastic in nature. We also see that the textures which are either structured, such as G5, or non-homogeneous such as C5 and F5, do not fit the said model very well. We would like to point out that the texture patches in figure 1 are 128x128 blocks taken from larger 512x512 images. A consequence of this, is that some 128x128 patches

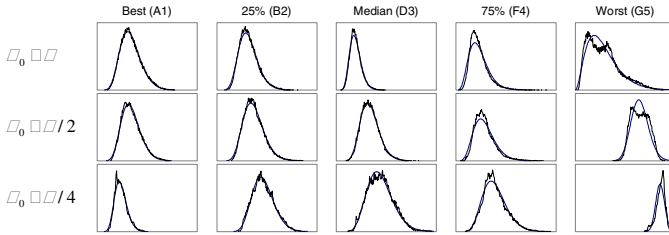


Fig. 3. An illustration of how well the gamma distribution model fits the extracted features. We show the histogram (jagged line) and fitted gamma model (smooth line) for a representative set of textures. In these figures we have $\theta = \pi/4$

may look homogeneous, while the entire 512x512 is not. This is for example the case for textures D4 and E4, and explains why they look very homogeneous in figure 1, but are not among the textures with the best goodness-of-fit wrp the gamma distribution. The degree to which the features histograms are gamma distributed can be observed in figure 3. It is also worth noting that in preliminary experiments, we find that adding a small amount of Gaussian noise to the texture images increased the goodness-of-fit with the gamma model. From these observations we make the following conjecture:

Conjecture 1. Spatially smoothed Gabor filter magnitude responses of some classes of visually homogeneous stochastic textures are gamma distributed.

4.2 Classification Performance

We perform two different sets of experiments to validate the classification performance when using the different similarity measures:

- Experiment A: Training and testing on the same data
- Experiment B: Training and testing disjoint data sets

The first set of experiments will indicate how well the feature set and similarity measure represent the textures. The second set of experiments will shed light on such issues as overtraining and generalization.

Gabor Filter Parameters. For our classification experiments we filter the images with Gabor filters of 4 different orientations, $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$, and 5 different scales, $\omega_0 = \{\pi, \frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{8}, \frac{\pi}{16}\}$ for a total of 20 filters. In the experiments with spatially smoothed magnitude images, a smoothing factor of $\gamma = 2/3$ was used for smoothing all Gabor filter magnitude response images. This is just a small portion of the possible parameter range of $\{\theta, \omega_0, \gamma\}$, and we see it as being similar to what is commonly used in the texture analysis community [11],[2].

A vector x in the featurespace \mathcal{F} is defined as $x = [x_1, \dots, x_N]$ where N is the number of dimensions in the featurespace. The parameters used to extract the feature x_j using (3) are $\{\theta_j, \omega_{0j}\}$.

Table 1. Misclassified blocks in experiment A (in percent)

Measure	Block size (in pixels)					
	<i>128x128</i>	<i>64x64</i>	<i>32x32</i>	<i>16x16</i>	<i>8x8</i>	<i>4x4</i>
<i>D_{KLGamma}</i>	0.9	4.8	10.6	15.9	19.2	21.1
<i>D_{KL}</i>	0.4	4.3	7.9	13.5	17.4	19.4
<i>D_{JD}</i>	0.9	4.2	10.3	16.8	22.3	26.5
<i>D_{χ²}</i>	0.9	3.9	10.0	17.3	23.5	28.6
<i>KL_{Gauss}</i>	0.7	4.4	10.8	16.8	20.5	22.8
<i>D_{WNN}</i>	1.6	6.0	14.0	20.5	24.8	27.3
<i>D_{WMV}</i>	6.1	16.5	34.6	53.4	62.5	65.5

Experimental Setup. In our experiments we used a 35 image subset of the 40 images used in [16]. The image size is 512x512 pixels. For experiment A, training was performed on the entire images, and testing was performed on all non-overlapping blocks of size 128x128, 64x64, 32x32,...,4x4. One might question whether blocks as small as 4x4 pixels can represent the variance of the texture they are extracted from. Recall that the features we are analyzing are spatially smoothed Gabor filter responses. Thus the Gabor features in a 4x4 block contain information extracted from a 'receptive field' slightly larger than the (largest) Gabor filter.

In experiment B we look at the ability to generalize. Training is performed on one quadrant of the image and testing on all non-overlapping blocks of size 128x128, 64x64, 32x32,...,4x4 in the remaining three quadrants. This procedure is repeated for all four quadrants, and the results presented are the average performance over all combinations of training and test sets. In both experiment A and B, the classification performance was measured by the classification precision, i.e. the percentage of correctly classified image blocks.

The histograms were binned by setting $\Delta l_j = \frac{1}{L} \max_{x,y,i} c_{xy,\omega_{0j},\theta_j,\gamma}^{(i)}$, where (i) denotes the texture and j the feature dimension. To avoid numerical instability when using the non-parametric methods we initialized the histogram bins to a small value $\delta = 10^{-5}$ before estimating the histograms. We experimented with values of L between 32 and 256, but found that classification errors did not decrease noticeably when increasing L from 64 to 256. Decreasing L from 64 to 32 gave marginally better classification results for small block sizes, and worse for large block sizes. Tables 1 and 2 therefore display the results of the non-parametric measures with 64 bins.

Our experiments showed that there was no noticeable difference in classification error between KLGamma with gamma distribution parameters estimated using the moment-method and KLGamma with parameters estimated using the ML method. We therefore display results from the former, due to its computational simplicity.

Interpretation. In experiment A we see that the best measure is KL. This is expected since it is equivalent to using the maximum-likelihood classifier, and the

Table 2. Misclassified blocks in experiment B (in percent)

Measure	Block size (in pixels)					
	<i>128x128</i>	<i>64x64</i>	<i>32x32</i>	<i>16x16</i>	<i>8x8</i>	<i>4x4</i>
$D_{KLGamma}$	8.0	12.7	18.7	23.7	26.9	28.8
D_{KL}	13.9	17.7	22.8	27.1	30.2	31.2
D_{JD}	7.6	11.4	18.2	24.5	29.6	33.6
D_{χ^2}	7.4	11.3	17.8	24.7	30.7	35.5
KL_{Gauss}	9.7	13.8	20.2	25.4	28.8	30.8
D_{WNN}	9.9	15.1	22.9	28.9	32.9	35.0
D_{WMV}	11.1	22.7	39.2	57.3	68.3	73.3

training data are representative of the test data. We also observe that KLGamma is the best parametric method for block sizes smaller than 64x64 pixels, which is also understandable since the assumption that the data are gamma distributed is a valid one for most of the textures. For block sizes of 128x128 and 64x64 pixels, the KL Gauss measure is slightly better than KLGamma. The difference between KL and KLGamma can be attributed to the fact the conjecture 1 does not hold for all of the textures in the dataset. JD and Chi are marginally better than KLGamma for block sizes larger than 16x16, and worse for smaller block sizes.

Experiment B reveals that KL does not do well when the test and training sets are disjoint. This is perhaps due to problems in estimating accurate histograms from the training data, either because the sample size is too small (256x256 pixels) or the training data are not representative of the test data. Chi is the best method for block sizes larger than 16x16 with KLGamma being only slightly worse. KLGamma is the best parametric measure, and for smaller block sizes it is the overall best similarity measure. We attribute the difference in classification error between experiment A and B to the fact that quite a few of the test images had internal textural variations, and were not entirely homogeneous across the whole image.

4.3 Computational Cost

In this section we will look at how computationally intensive the similarity measures tested in the previous section are. We will analyze them in terms of how many operations are required to classify a query texture as being one of M textures.

Parametric Measures. To give a fair speed comparison we will for each similarity measure define a set of parameters Q that must be calculated for each feature dimension of the query texture, and parameters P that must be calculated for each feature dimension of the database textures, in order for the similarity measure calculation to be as simple as possible. The equations in the second column of table 3 show the simplified calculations of each similarity measure for one feature dimension.

Table 3. Computational cost of parametric similarity measures

Measure	Expression	Q	P	No. of ops
D_{WNN}	$\frac{(Q_1 - P_1)^2}{P_2}$	$[\mu]$	$[\mu, \sigma^2]$	$4NM$
D_{WMV}	$ Q_1 - P_1 + Q_2 - P_2 $	$\left[\frac{\mu}{\alpha(\mu)}, \frac{\sqrt{\sigma^2}}{\alpha(\sigma)}\right]$	$\left[\frac{\mu}{\alpha(\mu)}, \frac{\sqrt{\sigma^2}}{\alpha(\sigma)}\right]$	$6NM + 4N$
KL_{Gauss}	$P_3 - Q_3 + \frac{(Q_1 - P_1)^2 + Q_2}{P_2}$	$[\mu, \sigma^2, \log \sigma^2]$	$[\mu, \sigma^2, \log \sigma^2]$	$7NM + 2N$
$D_{KLGamma}$	$P_3 - P_1 Q_1 + \frac{Q_2}{P_2}$	$[\Psi(\alpha) + \log \beta, \alpha \beta]$	$[\alpha, \beta, \log \Gamma(\alpha) + \alpha \log \beta]$	$5NM + 7N$

Table 4. Computational cost of non-parametric similarity measures

Measure	Expression	Q	P	No. of ops
D_{KL}	$Q_1 P_1$	$[-q]$	$[\log p]$	$2LNM$
D_{JD}	$(-Q_1 - P_1) \log(Q_1 + P_1) + P_2$	$[q]$	$[p, p \log p]$	$8LNM$
D_{χ^2}	$\frac{(Q_1 - P_1)^2}{Q_1 + P_1}$	$[q]$	$[p]$	$5LNM$

We assume that multiplications/divisions are as fast as additions/subtractions, one table lookup takes two operations and the $|\cdot|$ operator takes one operation. A further assumption is that the quantities μ and σ^2 are already calculated and the moment method is used to calculate the parameters of the gamma distribution.

Also assume that the texture featurespace has N dimensions and the database texture parameters P are calculated in advance. In the rightmost column of table 3 we display the number of operations required to classify a query texture as being one of M possible database textures using the different similarity measures. From these results we see that WNN is the fastest method. If $M > 3$ the proposed KLGamma measure is the second fastest.

Non-parametric Measures. We will do a similar analysis for the nonparametric measures. Simplifications will be done where possible. Here Q and P will be calculated for each of the L histogram bins in all the N feature dimensions. q and p denote the bin values for the query texture and database texture respectively. These results show that the KL measure is the easiest to compute, and the Chi measure is significantly faster than the JD measure.

If we compare the computational cost of the parametric measures with the non-parametric measures we see that they are approximately L times faster.

5 Discussion and Future Work

A question that naturally arises is whether the classification results we have obtained represent what would typically be found in texture classification applications. The first issue that should be addressed in order to answer this question is whether the test data are representative of "typical" textures. To say they are would be jumping to conclusions without much evidence. However, one can not avoid noticing how well the gamma distribution model fits the features

extracted from the visually homogeneous and stochastic textures. We suggest that further research should be done to determine for which classes of homogeneous stochastic textures our conjecture holds, and if it holds for homogeneous deterministic (e.g. polkadot) textures when there is noise present. Preliminary experiments we have done, suggest that this might be the case. In a later paper we will investigate more thoroughly the validity of our conjecture, give a rationale for why the extracted Gabor features appear to be gamma distributed, and determine how to extend the use of the KLGamma similarity measure to scale and rotation invariant texture classification/retrieval.

The second issue we wish to touch upon, is whether our computational cost comparison gives results that are valid in a realistic implementation scenarios. If the implementation scenario follows the assumptions we have made, the results should be valid. The plethora of different specialized hardware and software platforms in existence makes it almost impossible to give a computational cost comparison that is valid in general. However, we can determine that the relative difference in computational cost between our proposed KLGamma measure and non-parametric measures is approximately an order of magnitude.

6 Conclusion

Based on observations of 35 texture images, we have conjectured that the spatially smoothed Gabor filter magnitude responses of some classes of visually homogeneous stochastic textures are gamma distributed. For the class of textures under which this conjecture holds, we have derived a computationally simple parametric similarity measure, the KLGamma measure, which is asymptotically optimal when used for texture classification/retrieval. This texture similarity measure is based on the Kullback-Leibler divergence between gamma distributions.

We compared the texture classification performance of KLGamma with several parametric and non-parametric measures. Classification experiments confirmed that KLGamma is close to optimal. When training and testing was performed on the same dataset, the only method that noticeably outperformed our method was the Kullback-Leibler divergence between binned histograms. We attribute this to the fact that some of the textures in our test suite were non-homogeneous. Our conjecture does not hold for these textures. In tests with disjoint test and training sets, KLGamma was only slightly outperformed by the non-parametric Jeffrey divergence and symmetric χ^2 statistic for block sizes larger than 16x16 pixels. For smaller blocks, KLGamma achieved the best classification results of all the similarity measures we have tested.

An analysis of the computational cost of all the tested measures, indicate that KLGamma is at least an order of magnitude faster than non-parametric measures, while having comparable, and under some conditions better, classification performance. Only the simplest parametric method, WNN, is faster.

Judging from the results we have obtained, we can conclude that, in scenarios under which our conjecture holds, our proposed KLGamma measure is the

method of choice in e.g. real-time block-based segmentation algorithms such as that proposed by Hofman et.al. in [17]. KLGamma is also well suited for texture classification applications where speed is of the essence and the number of texture classes is large.

Appendix A

Let X be a sample of S feature vectors from a texture q . X_{jk} denotes sample k from feature dimension j . Let N be the number of dimensions in the feature vectors. If a texture i is described by a set of binned histograms $p_j^{(i)}$, the likelihood of the samples X being from texture i is $\lambda_i = \prod_{j=1}^N \prod_{k=1}^S p_j^{(i)} \left(\left\lfloor \frac{X_{jk}}{\Delta l_j} \right\rfloor \right)$. The maximum-likelihood classifier is $g(X) = \arg \max_i \lambda_i$. If we let $b_{jk} = \left\lfloor \frac{X_{jk}}{\Delta l_j} \right\rfloor$ and take the log of λ_i we have $\log \lambda_i = \sum_{j=1}^N \sum_{k=1}^S \log p_j^{(i)}(b_{jk})$. Now let n_{jl} denote the number of samples X_{jk} for which $b_{jk} = l$. With this notation we obtain $\log \lambda_i = \sum_{j=1}^N \sum_{l=0}^{L-1} n_{jl} \log p_j^{(i)}(l)$. Furthermore, dividing by S on each side implies that

$$g(X) = \arg \max_i \lambda_i = \arg \max_i \frac{1}{S} \log \lambda_i = \arg \max_i \lambda_i \sum_{j=1}^N \sum_{l=0}^{L-1} p_j^{(q)}(l) \log p_j^{(i)}(l) \quad (17)$$

Where $p_j^{(q)}$ are the binned histograms of the query texture.

$$g(X) = \arg \max_i \lambda_i = \arg \min_i D_{KL} \left(q, p^{(i)} \right) = \arg \min_i \sum_{j=1}^N \sum_{l=0}^{L-1} p_j^{(q)}(l) \log \frac{p_j^{(q)}(l)}{p_j^{(i)}(l)} \quad (18)$$

and thus minimizing the KL divergence is identical to the maximum-likelihood classifier.

References

1. Tai Sing Lee, Image Representation Using 2D Gabor Wavelets, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 18, No. 10, October 1996
2. David A. Clausi, M. Ed Jernigan, Designing Gabor filters for optimal texture separability, Pattern Recognition 33 (2000) 1835-1849
3. W.D. Penny, KL-Divergences of Normal, Gamma, Dirichlet and Wishart densities, Wellcome Department of Cognitive Neurology, University College London, March 30, 2001
4. S. Marcelja, Mathematical Description of the Responses of Simple Cortical Cells, J. Optical Soc. Am., vol. 70, pp. 1297-1300, 1980
5. J.G. Daugman, Two-dimensional Analysis of Cortical Receptive Field Profile, Vision Research, vol. 20, pp. 847-856, 1980

6. J.G. Daugman, Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters, *J. Optical Soc. Am.*, vol. 2, no. 7, pp. 1160-1169, 1985
7. D.A. Pollen and S.F. Ronner, Phase Relationship Between Adjacent Simple Cells in the Visual Cortex, *Science*, vol. 212, pp. 1409-1411, 1981
8. J.H. van Hateren and A. van der Schaf, Independent Component Filters of Natural Images Compared with Simple Cells in Primary Visual Cortex, *Proc.R.Soc.Lond.B* 265:359-366, 1998
9. Ronald E. Walpole and Raymond H. Myers, *Probability and Statistics for Scientists and Engineers* 5th Edition, Prentice Hall, 1993
10. C.T.J. Dodson and J. Scharcanski, *Information Geometry For Stochastic Refinement of Database Search and Retrieval Algorithms*, 2001
11. A.C. Bovik, M. Clark, W.S. Geisler, Multichannel Texture Analysis Using Localized Spatial Filters, *IEEE Trans. Pattern Anal. Machine Intell.* 12 (1) (1990) 55-73
12. Nuno Vasconcelos and Andrew Lippman, *A Unifying View of Image Similarity*, MIT Media Lab
13. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, 1990
14. B.S. Manjunath and W.Y. Ma, Texture Features for Browsing and Retrieval of Image Data, *IEEE Trans. Pattern Anal. Machine Intell.* 18 (8) (1996) 837-842
15. Jan Puzicha, Joachim M. Buhmann, Yossi Rubner and Carlo Tomasi, Empirical Evaluation of Dissimilarity Measures for Color and Texture, *Proceedings of the International Conference on Computer Vision (ICCV'99)*, 1165-1173, 1999
16. Jan Puzicha, Thomas Hofmann and Joachim M. Buhmann, Non-parametric Similarity Measures for Unsupervised Texture Segmentation and Image Retrieval, *Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, San Juan, 1997
17. Thomas Hofmann, Jan Puzicha and Joachim M. Buhmann, Deterministic Annealing for Unsupervised Texture Segmentation, *Proc. of the International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR'97)*
18. Richard J. Larsen and Morris L. Marx, *An Introduction to Mathematical Statistics and Its Applications* 3rd Edition, Prentice Hall, 2001
19. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.V.P. Flannery, *Numerical Recipes in C*, Cambridge, 1992
20. G. R. Cross and A. K. Jain. Markov Random Field Texture Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 5/1, 1983
21. Chun-Shien Lu and Pau-Choo Chung. Wold Features for Unsupervised Texture Segmentation. Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan, ROC, 1998
22. K. I. Laws. Rapid texture identification. In *Proc. of the SPIE Conference on Image Processing for Missile Guidance*, pages 376-380, 1980
23. Timo Ojala and Matti Pietikäinen, Unsupervised texture segmentation using feature distributions. *Pattern Recognition*, Vol. 32, pp. 477-486, 1999
24. Trygve Randen and John Håkon Husøy. Filtering for Texture Classification: A Comparative Study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, Apr 1999. Web: <http://www.ux.his.no/~tranden/>