Lightweight Probabilistic Texture Retrieval

Roland Kwitt, *Student Member, IEEE,* Andreas Uhl

***Abstract*—This article contemplates the framework of probabilistic image retrieval in the wavelet domain from a computational point of view. We not only focus on achieving high retrieval rates, but also discuss possible performance bottlenecks which might prevent practical application. We propose a novel retrieval approach which is motivated by previous research work on modeling the marginal distributions of wavelet transform coefficients. The building blocks of our work are the Dual-Tree Complex Wavelet Transform and a number of statistical models for the coefficient magnitudes. Image similarity measurement is accomplished by using closed-form solutions for the Kullback-Leibler divergences between the statistical models. We provide an in-depth computational analysis regarding the number of arithmetic operations required for similarity measurement and model parameter estimation. The experimental retrieval results on a widely-used texture image database show that we achieve competitive retrieval results at low computational cost.**

***Index Terms*—Texture Image Retrieval, Wavelets, Kullback-Leibler Divergence**

I. Introduction

SINCE the amount of digital image data in multimedia databases is constantly growing, we face an increasing need for systems which allow content-based image retrieval (CBIR) through searching by example. The fields of application range from searching databases of natural images to images of textures or even medical content. The objective is to find the $K \ll L$ most similar images to a given query in a database of L candidate images, according to some similarity criterion. A typical image retrieval system consists of two elementary building blocks: the feature extraction (FE) block and the similarity measurement (SM) block, which are illustrated in Fig.1.

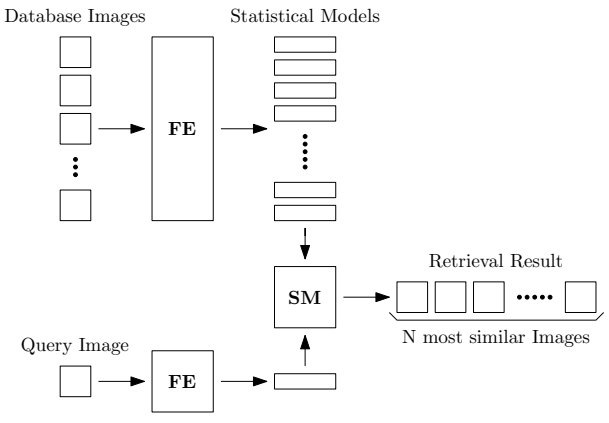


Fig. 1. Schematic diagram of an image retrieval system.

Roland Kwitt (corresponding author) and Andreas Uhl are with the Department of Computer Sciences, University of Salzburg, 5020 Salzburg, Austria (e-mail: {rkwitt,uhl}@cosy.sbg.ac.at). This work was funded by the Austrian Science Fund under Project no. L366-N15.

During feature extraction, a set of image features (signatures) is computed from the image representation in the pixel domain. This step usually comprises several intermediate steps such as preprocessing, image transformation or statistical model estimation. The ultimate aim is to design a set of image descriptors which can unambiguously characterize the content of an image. This fact implies an important chain of relations: on the one hand, the feature extraction part strongly depends on the type of image content, since buildings or objects for example will most likely require a different set of image features than textures or medical content. On the other hand, similarity measurement consequently depends on the type of image features which establishes the strong connection of both building blocks. Comparing image signatures obtained from an edge-detection system and texture features for example with the same similarity function could be suboptimal for many databases. In our work, we particularly focus on images which exhibit texture characteristics. Since most publications on (texture) image retrieval solely aim at an improvement in retrieval accuracy and often neglect computational issues, solutions which are both computationally inexpensive and minimize the retrieval error are rare. In a probabilistic framework, where each image is represented by some statistical model and image similarity is measured by a function of these models, we have to deal with the trade-off between model complexity and computational performance. Increasing the model complexity to better capture image characteristics might lead to higher retrieval rates on the one hand, but it is very likely that the computational demand for feature extraction and/or similarity measurement increases in a similar manner.

We consider two scenarios which impose computational constraints on particular parts of the retrieval framework. The scenarios differ in that possible performance bottlenecks arise at different locations. The first scenario is the classic retrieval scenario, where the model parameters of all database images are calculated off-line and new images are added to the database at a slow rate. Hence, overall runtime performance is predominantly limited by similarity measurement which inherently depends on the size of the image database. The runtime impact of model parameter estimation and image transformation is of secondary importance since both steps have to be performed only once (i.e. for each new query). The second retrieval scenario we discuss here has several facets and imposes additional requirements on the building blocks of the retrieval framework. First, we observe situations where new images arrive at a high rate and have to be stored in the database. At the same time, image queries are executed. The computational demand for similarity measurement is still the primary concern, however the complexity of parameter estimation becomes an important issue. If the images are represented in a domain other than spatial, the image trans-

formation step possibly contributes a significant amount of additional runtime as well. Other challenging situations occur when online texture similarity measurement is required, e.g. when then frames of an image stream have to be matched to a limited set of query templates. Real-world examples for that include video-controlled quality assurance in texture manufacturing, or the detection of cancerous tissue during video-colonoscopy. Computationally expensive parameter estimation or image transformation can scale up to the limiting factors for production throughput or complicate the diagnostic process. In order to cover both retrieval scenarios we need a low-complexity image transformation, a similarity measure which exclusively depends on the image model parameters and an efficient model parameter estimation procedure.

A. Related Work

Related work on the topic of texture image retrieval includes a huge variety of publications in the last decade. Due to space-limitations we focus on those approaches which are closely related to the framework of probabilistic image retrieval. To the best of our knowledge, the idea of a probabilistic approach was first introduced by Vasconcelos and Lippman in [1] and then taken up by several authors in the following years. Our approach is mainly motivated by their work presented in [2], [3] as well as the work of Do and Vetterli in [4]. Both approaches will be used as a reference for our retrieval experiments. In [3], multivariate Gaussian Mixture Models (MGMM) are used to model Discrete Cosine Transform (DCT) coefficients obtained from overlapping sliding windows and Maximum-Likelihood selection is employed for similarity measurement. In [4], the pyramidal Discrete Wavelet Transform (DWT) is used to decompose images and model the detail subband coefficients by Generalized Gaussian distributions (GGD). A similarity measure between two images is then computed using a closed-form solution to the KL-divergence between GGDs. Hence, the obtained retrieval framework allows efficient similarity measurement by using the Maximum-Likelihood estimates (MLEs) of the GGD distribution parameters of each wavelet detail subband. However, the final similarity measure depends on the assumption of subband independency which amounts for a crucial simplification and indicates the trade-off between model complexity and computational performance. The effect of assuming independency of the components in the feature space is covered by Vasconcelos [5] in detail, with the conclusion that independence has a negative impact on the retrieval accuracy. In [6], Do and Vetterli present an extension of the aforementioned probabilistic approach to achieve rotational invariance. The new statistical model is based on an overcomplete transform, known as the Steerable Pyramid [7] and two particular forms of Hidden Markov Trees (HMT) to capture coefficient dependencies across scales and orientations. Image similarities are then measured by an approximation of the KL-divergence between HMTs [8]. A similar approach which follows the idea of measuring KL-divergences is presented by Tzagkarakis et. al. [9] using Symmetric Alpha-Stable distributions (SαS) to model the wavelet detail subband coefficients. Since there exists no closed-form solution for the KL-divergence in case of general SαS distributions, the authors propose to use the characteristic functions instead of the probability density functions (PDF) to compute similarities. In [10], this approach is carried forward to achieve rotational invariance by employing a Steerable Pyramid approach together with alpha-stable modeling of the subband coefficients and a Gaussianization procedure to obtain a closed-form expression for the KL-divergence. Another interesting approach is presented by de Ves et. al [11], where the wavelet coefficients of the vertical and horizontal DWT detail subbands are considered as realizations of a bivariate random vector and the magnitude is modeled by a two-parameter Gamma distribution. The authors report good retrieval results using the Stationary Wavelet Transform (SWT, implemented by the à-trous algorithm) as a substitution for the pyramidal DWT to get rid of the shift-dependency problem (see Section II-A). Again, the KL-divergence is then used for similarity measurement.

B. Contribution

The contribution of this work is split into several parts: we introduce a novel, probabilistic texture image retrieval approach in the wavelet domain with the objective to strike a balance between model complexity and computational performance. In particular, we propose a set of statistical models for the magnitudes of complex wavelet transform coefficients which facilitate an easy derivation of closed-form expressions for the corresponding KL-divergences. By compiling some results form statistical literature on parameter estimation we show that model parameters can be estimated in an efficient manner which perfectly goes with the idea of lightweight retrieval. To quantify the computational complexity of our approach, we provide an in-depth analysis of the required arithmetic operations for the main building blocks as well as a comparative runtime study. This analysis confirms that our method facilitates application to scenarios where low computational cost is a crucial requirement. From the viewpoint of retrieval accuracy, the experimental results on a widely-used texture database show higher or at least competitive retrieval rates compared to previous research works. This is particularly interesting since our models are less complex than some of the presented reference approaches.

The remainder of the paper is organized as follows: in Section II we introduce image representation in the wavelet domain and the setting of probabilistic image retrieval. We further discuss statistical models for wavelet transform coefficients and present closed-form expressions for the KL-divergences. Model parameter estimation issues are investigated in Section III, followed by a computational analysis for the building blocks of our retrieval framework in Section IV. Experimental results are then shown in Section V and Section VI concludes the paper with a summary of the main points, open problems and an outlook on future research.

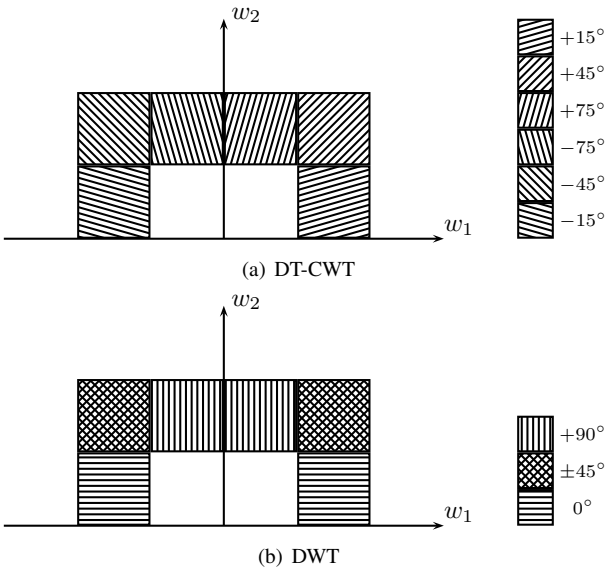
II. Statistical Models and Similarity Measurement

First, we introduce some notational conventions: if not stated otherwise, we use single indexing for the wavelet sub-

band coefficients $x_1, \ldots, x_N$ and go without a complete specification of the concrete subband position in the decomposition structure. Small boldface letters, such as $\mathbf{a}$ denote vectors, big boldface letters such as $\mathbf{A}$ denote matrices. Distribution parameters are denoted by Greek letters. In order to make a quantitative statement about the quality of the statistical models, we perform Chi-Square Goodness-of-Fit (GoF) tests at the 5% significance level. The number of bins to compute the Chi-Square test statistic is fixed to $0.3s$, where s denotes the sample standard deviation. This is the setting used in the software DATAPLOT [12].

A. Image Representation

In the feature extraction step of our work, we leave the spatial (pixel) domain and work in the wavelet domain instead. This is motivated by several reasons: first, wavelets are a convenient way to obtain a multiscale representation of an image which closely corresponds to the way the human visual system processes information [13]–[15]. Second, we assume that this multiscale representation allows to efficiently capture texture characteristics by computing wavelet coefficient statistics. Last, the characteristic shapes of the coefficient histograms lead to simple statistical models which in turn lead to efficient similarity measurement. We choose the Dual-Tree Complex Wavelet Transform (DT-CWT) [16], since it overcomes two shortcomings of the pyramidal DWT: lack of shift-invariance and lack of directional selectivity, as is vividly illustrated and explained in [17]. These shortcomings are particularly relevant for image analysis purposes. Lack of shift-invariance implies that singularities at different locations in an image lead to different representations in the wavelet domain (i.e. different coefficients). Hence, wavelet coefficients representing an edge along an object contour are not necessarily large across all scales which causes ringing artifacts when reconstruction is done using just the coefficients at a specific scale. Of course, the perfect reconstruction property guarantees that all artifacts are canceled when computing the inverse DT-CWT using all coefficients. The technical reason for the shift-dependency problem is that the wavelet and scaling filter used to implement the pyramidal DWT have finite support and the coefficients are downsampled by two after each decomposition stage. As a matter of fact, shift-dependency is a severe deficiency for image analysis. The second shortcoming – lack of directionally selectivity – is related to the fact that the filters of the DWT are real functions and are thus supported on both sides of the frequency axis. Since the 2-D DWT is usually implemented by separate row- and column filtering (which is equivalent to using tensor-product wavelets), this causes ambiguities in distinguishing features oriented along $\pm 45°$. All other features oriented mostly along the vertical or horizontal direction are lumped in the vertical and horizontal detail subbands. Since orientation information can be an important characteristic for some textures, better directional selectivity is desired. Both deficiencies are eliminated to a certain extent by using the DT-CWT at low computational overhead (see Section IV-A). The basic idea is to use complex wavelets which are composed of two real wavelets forming an approximate Hilbert transform pair. Since this construction ensures that negative frequencies are suppressed, aliasing effects are reduced and thus approximate shift-invariance is guaranteed. Further, a higher degree of directional selectivity is achieved with six complex detail subbands at each decomposition stage. The detail subbands are oriented along approximately $\pm 15°, \pm 45°$ and $\pm 75°$. A schematic frequency tiling of the DT-CWT and DWT is shown in Fig.2 for frequencies $w_2 > 0$.



(a) DT-CWT

(b) DWT

Fig. 2. Schematic frequency tiling of the DT-CWT and DWT ($w_2 > 0$).

B. Probabilistic Image Retrieval

We recapitulate some results from probabilistic, maximum a-posteriori probability (MAP) image retrieval which constitute the basis of our work. For a thorough treatment of the decision-theoretic details, we refer to [18] and [19]. Let $I_1, \ldots, I_L$ denote L database images. Each image is represented by some data vector $\mathbf{x}_i = [x_{i1}, \ldots, x_{id}]^T \in \mathcal{X} := \mathbb{R}^d$, where $\mathcal{X}$ denotes the feature space and d denotes the feature space dimensionality. The data vectors are obtained in the feature extraction block of the retrieval system. We assume that each image belongs to one of M image classes and that each class has some associated class PDF $p_i(\mathbf{x}), 1 \leq i \leq M$. Further, a variable $Y \in \{1, \ldots, M\}$ signifies the class membership and $P(Y = i)$ is the prior probability of class i. In addition, let $s : \mathcal{X} \to \{1, \ldots, M\}$ denote a similarity function which assigns a query feature vector to one of the M image classes. Minimization of the retrieval error is achieved using the MAP similarity function

$$s(\mathbf{x}_q) = \underset{i}{\operatorname{argmax}}\, P(Y = i|\mathbf{x}_q), \tag{1}$$

where $\mathbf{x}_q$ denotes the feature vector of an arbitrary query image I_q and $P(Y = i|\mathbf{x}_q)$ denotes the posterior probability of class i given $\mathbf{x}_q$. We assume that the elements of an arbitrary feature vector $\mathbf{x}$ are d realizations of i.i.d. random variables following some parametric PDF $p(x|\boldsymbol{\theta}_i), 1 \leq i \leq M$. Since, under the image retrieval formulation of [3], each image constitutes its own class (i.e. $M = L$), the parameter vector

TABLE I
PERCENTAGE OF REJECTED NULL-HYPOTHESIS FOR A SET OF STATISTICAL MODELS AND IMAGE REPRESENTATIONS

		Scale		
Transform	**Subband Model**	1	2	3
DWT	Laplace	91.98	68.36	48.13
DWT	GGD	38.30	17.51	22.70
DT-CWT	Rayleigh	96.99	85.05	54.87
DT-CWT	Gamma	62.64	38.26	13.02
DT-CWT	Weibull	75.41	40.84	11.93

$\boldsymbol{\theta}_i$ serves to identify the image (class). In case we apply the Bayes rule to (1) and assume equal priors $P(Y = i) = 1/M$, it can be shown that the resulting Maximum-Likelihood selection rule is asymptotically (i.e. $d \to \infty$) equivalent to searching the minimum of the KL-divergence between the PDF of the query image model and the PDFs of the candidate image models, given by

$$s(\mathbf{x}_q) = \underset{i}{\operatorname{argmin}} \int_D p(x|\boldsymbol{\theta}_q) \log\left(\frac{p(x|\boldsymbol{\theta}_q)}{p(x|\boldsymbol{\theta}_i)}\right) dx, \qquad (2)$$

with D denoting the domain of $p(x|\cdot)$ [20]. It is thus obvious that the setup of probabilistic image retrieval strongly connects the feature extraction and the similarity measurement block. For computational reasons we favor closed-form expressions of (2). Otherwise, we would have to resort to the discrete version of the KL-divergence which requires a reasonable histogram binning. Another possible way to compute similarities when there is no closed-form solution available, is to resort to Monte-Carlo (MC) simulation. This means that we exploit the fact that

$$\frac{1}{S}\sum_{k=1}^{S}\left[\log(p(x_{qk}|\boldsymbol{\theta}_q)) - \log(p(x_{qk}|\boldsymbol{\theta}_i))\right] \qquad (3)$$

converges to the KL-divergence between $p(x|\boldsymbol{\theta}_q)$ and $p(x|\boldsymbol{\theta}_i)$ as $S \to \infty$. For practical use we generate a random sample of size S from the query image model $p(\cdot|\boldsymbol{\theta}_q)$ and then compute (3). Obviously, this procedure is computationally expensive since S has to be chosen reasonable large to minimize the impact of the sample size. As a matter of fact, MC simulation is impractical in a lightweight approach. However, we mention it here since one reference approaches in our comparative study comprises a mixture-model with no closed-form KL-divergence.

C. DWT Detail Subband Models

Several statistical distributions have been proposed in literature to model the marginal detail subband coefficient distributions of the pyramidal DWT. It is commonly accepted that the coefficients are highly non-Gaussian, exhibit heavy-tails and can be accurately modeled by a Generalized Gaussian distribution [21]. In the remainder of this work, we use the parametrization of [22], where the PDF of the GGD is given by

$$p(x|\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-\left|\frac{x}{\alpha}\right|\right)^{\beta}, \qquad (4)$$

with $-\infty < x < \infty$, $\alpha > 0$ (scale) and $\beta > 0$ (shape). For $\beta = 2$, the GGD is the Gaussian distribution, whereas for $\beta = 1$ we obtain the Laplace distribution [23]. Parameter estimation will be briefly discussed in Section III. Table I lists the percentage of rejected null-hypothesis of the Chi-Square GoF tests when using either Laplace or GGD to model the marginal coefficient distributions of our database images (see Section V). The GoF tests confirm, that the GGD should be the model of choice due to the low percentage of rejected null-hypothesis. A closed-form expression of the KL-divergence between two GGDs $p_i := p(x|\alpha_i,\beta_i)$ and $p_j := p(x|\alpha_j,\beta_j)$ is provided in [4] as

$$\begin{aligned}\underset{\text{GGD}}{\text{KL}}(p_i||p_j) = \log\left(\frac{\beta_i\alpha_j\Gamma(1/\beta_j)}{\beta_j\alpha_i\Gamma(1/\beta_i)}\right) + \\ \left(\frac{\alpha_i}{\alpha_j}\right)^{\beta_j}\frac{\Gamma((\beta_j+1)/\beta_i)}{\Gamma(1/\beta_i)} - \frac{1}{\beta_i},\end{aligned} \qquad (5)$$

where $\Gamma(\cdot)$ denotes the Gamma function [24]. Another way to model the non-Gaussian nature of the coefficients is to use zero-mean Gaussian mixture models (GMM) [25]. Although there exists no closed-form expression for the KL-divergence between two GMMs we mention the GMM model at this point since we will need it to explain a model for the DT-CWT detail subbands later. The reasoning for a GMM is built upon the compression property (i.e. two-Population property [25]) of the DWT. By transforming a signal with the DWT we obtain a sparse representation with a small number of large coefficients and a large number of small coefficients. This gives rise to a two component GMM: one Gaussian component to model the large coefficients and one Gaussian component to model the small coefficients. Given that the PDF for the small coefficients is denoted by $p(x|0,\sigma_S^2)$ and the PDF for the large coefficients is denoted by $p(x|0,\sigma_L^2)$, the PDF of the two-component GMM can be written as

$$p(x) = \sum_{i\in\{S,L\}} w_i p(x|0,\sigma_i^2), \text{ with } \sum_{i\in\{S,L\}} w_i = 1, \qquad (6)$$

where w_i denote the prior probabilities for the corresponding components. Usually, the model parameters for the GMM are estimated by an Expectation-Maximization (EM) algorithm (see [18]). Both, the GGD and the GMM represent accurate models for marginal distributions of the subband coefficients.

D. DT-CWT Detail Subband Models

In case of the DT-CWT, each detail subband coefficient x_i is complex-valued. We start with a very simple model for the complex detail coefficient magnitudes and refine it to enhance the goodness-of-fit. An exemplary histogram of coefficient magnitudes $|x_i|$ is shown in Fig.3. Apparently, candidate models are positively skewed distributions (i.e. skewed to the right) which often arise in reliability and life-span modeling [26]. Similar distributions can be also observed in modeling the statistics (especially the amplitude) of Synthetic Aperture Radar (SAR) data (see [27], [28]). In order to derive a model for the coefficient magnitudes, we first follow the naive approach of assuming that both the real and imaginary component of the complex-valued signal can be modeled by zero-mean Gaussian distributions with equal variance. Under this assumption, it can easily be verified by a

variable transformation to polar coordinates that the magnitude follows a Rayleigh distribution with PDF

$$p(x|\alpha) = \frac{x}{\alpha^2}\exp\left(-\frac{x^2}{2\alpha^2}\right), \tag{7}$$

$0 < x < \infty$ and $\alpha > 0$ (shape). The ML parameter estimates have an explicit expression and are given in [23]. The KL-divergence between two Rayleigh distributions $p_i := p(x|\alpha_i)$ and $p_j := p(x|\alpha_j)$ is

$$\underset{\text{Rayleigh}}{\text{KL}}\,(p_i||p_j) = \left(\frac{\alpha_i}{\alpha_j}\right)^2 + 2\log\left(\frac{a_j}{a_i}\right) - 1. \tag{8}$$

However, due to the high percentage of rejected null hypothesis (see Table I), this first statistical model is probably not flexible enough to describe the underlying data.

The next model we consider is the two-parameter Weibull distribution which includes the Rayleigh distribution as a special case. This is a reasonable choice since there are more degrees of freedom to adapt to the data. In [29] we exploited the Weibull distribution parameters for the purpose of medical image classification and in [30] this model was first successfully employed in texture image retrieval. The PDF of a Weibull distribution is

$$p(x|\alpha,\beta) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left\{-\left(\frac{x}{\beta}\right)^{\alpha}\right\}, \tag{9}$$

with $0 < x < \infty$, $\alpha > 0$ (shape) and $\beta > 0$ (scale). For $\alpha = 2$ and $\beta = \sqrt{2\beta}$ (9) reduces to the Rayleigh distribution. Solutions for the MLEs of β and α are again given in [23] and will be discussed in Section III, together with an alternative estimation method. Inserting the PDF of the Weibull distribution into (2) leads to the following closed-form expression for the KL-divergence between $p_i := p(x|\alpha_i,\beta_i)$ and $p_j := p(x|\alpha_j,\beta_j)$,

$$\underset{\text{Weibull}}{\text{KL}}\,(p_i||p_j) = \Gamma\left(\frac{\alpha_j}{\alpha_i}+1\right)\left(\frac{\beta_i}{\beta_j}\right)^{\alpha_j} + \log\left(\beta_i^{-\alpha_i}\alpha_i\right) - \log\left(\beta_j^{-\alpha_j}\alpha_j\right) + \log\left(\beta_i\right)\alpha_i - \log\left(\beta_i\right)\alpha_j + \frac{\gamma\alpha_j}{\alpha_i} - \gamma - 1, \tag{10}$$

where $\gamma = 0.577216$ denotes the Euler-Mascheroni constant. Table I shows, that the Weibull distribution is a good statistical model, especially for decomposition depths ≥ 2.

The third model we consider is the Gamma distribution, which has previously been proposed as an alternative to the Rayleigh distribution for modeling the magnitudes of Gabor filter outputs [31]. The Gamma PDF is

$$p(x|\alpha,\beta) = \frac{\beta^{-\alpha}x^{\alpha-1}}{\Gamma(\alpha)}\exp\left(-\frac{x}{\beta}\right), \tag{11}$$

with $x < 0 < \infty$, $\alpha > 0$ (shape) and $\beta > 0$ (scale). Again, there exists a closed-form expression for the KL-divergence between the PDFs $p_i := p(x|\alpha_i,\beta_i)$ and $p_j := p(x|\alpha_j,\beta_j)$,

$$\underset{\text{Gamma}}{\text{KL}}\,(p_i||p_j) = \psi(\alpha_i)(\alpha_i-\alpha_j) - \alpha_i + \log\left(\frac{\Gamma(\alpha_j)}{\Gamma(\alpha_i)}\right) + \alpha_j\log\left(\frac{\beta_j}{\beta_i}\right) + \frac{\alpha_i\beta_i}{\beta_j}, \tag{12}$$

where $\psi(\cdot)$ denotes the Digamma function [24]. The numbers in Table I indicate that the Gamma distribution seems to be a good model as well, with almost equal percentages of rejected null-hypothesis compared to the Weibull distribution.

The last model we discuss here is a direct result of modeling the DWT coefficients by GMMs (see Section II-C). Since the real and imaginary components of the complex coefficients can be fairly well modeled by a GGD, it seems reasonable to derive the corresponding distribution for the absolute values. This has already been done in [27], however with the result that the PDF (termed Generalized Gaussian Rayleigh) has no analytic expression and requires numerical integration. Fortunately, the real and imaginary component can also be modeled by separate GMMs. Relying on the same arguments that are used to derive the Rayleigh distribution leads to a two-component mixture of Rayleigh distributions (RMM) as a statistical model for the absolute values. The parameters of the RMM are estimated using an EM algorithm which is given in Appendix A. As with GMMs, there exists no closed-form expression for the KL-divergence and we have to resort the MC simulation. Although the RMM is impractical for lightweight texture retrieval we use it as a reference model for our comparative study. To visualize the PDF shape of the discussed statistical models, Fig.3 shows a histogram of DT-CWT coefficient magnitudes together with the fitted PDFs.

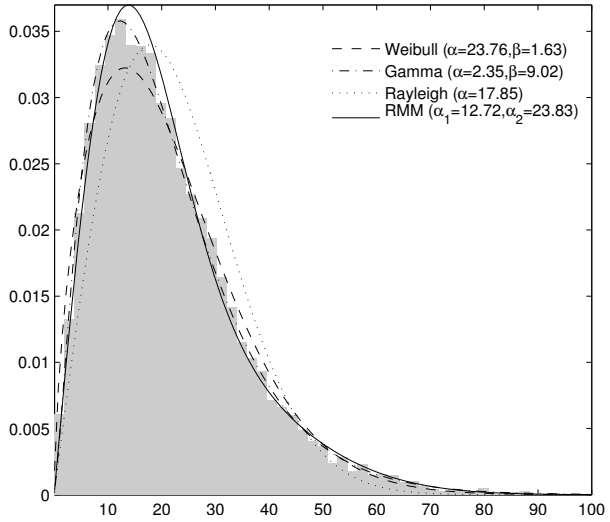


Fig. 3. Histogram (shaded) of DT-CWT wavelet coefficients magnitudes $|x_i|, x_i \in \mathbb{C}$ together with fitted PDFs of the discussed statistical models (second-level detail subband of image Fabric.0000, see Fig.5(a)).

E. A Final Similarity Measure

In the previous sections we discussed how to measure similarity between the statistical wavelet coefficient models of one subband. Yet, we do not have a final similarity measure between two images. For that reason we follow the same approach as in [4] and assume that the detail subbands of the DT-CWT are independent. This allows application of the chain-rule of entropy which leads to a simple final similarity measure where we can sum up the KL-divergences over all subbands [32]. To formalize this, let J denote the maximum

decomposition depth, k denotes the subband index and K denotes the number of subbands per scale. The final similarity measure between two images I_q and I_i is defined as

$$S(I_q, I_i) := \sum_{j=1}^{J} \sum_{k=1}^{K} \mathrm{KL}(p_q^{sk} || p_i^{sk}), \tag{13}$$

where the superscripts s and k identify the PDF of the statistical model for the k-th subband at decomposition level j. Although, the independency assumption is likely to be violated for any overcomplete transform, we neglect this fact for the sake of computational simplicity. We further note, that the KL-divergence is not a metric, since it violates the symmetry property and the triangle inequality. To remedy the lack of symmetry, we follow the approach of [33] and artificially symmetrize the KL-divergence. Let p_i and p_q denote two arbitrary PDFs, then the symmetrized KL-divergence is defined as

$$\overset{*}{\mathrm{KL}}(p_q||p_i) := \frac{1}{2}\left(\mathrm{KL}(p_i||p_q) + \mathrm{KL}(p_q||p_i)\right). \tag{14}$$

We use this symmetric version of the KL-divergence for all experiments.

III. Parameter Estimation

In this section we cover the issue of model parameter estimation which is crucial to achieve good retrieval performance. To the best of our knowledge, the issue of both precise and computationally inexpensive estimation is often neglected in research works on probabilistic image retrieval. We discuss the two most prominent estimation methods in the context of distribution parameter estimation: Maximum-Likelihood estimation (MLE) and the method of Moment Matching (MM).

A. GGD Parameters

Since estimation of the GGD parameters has already been extensively covered in literature, we only provide a brief overview of the main results. Maximum Likelihood estimation is studied in the work of Varanasi et al. [34] including both joint parameter estimation and situations where one parameter is already known. In image retrieval both parameters are unknown and estimation of the shape parameter β requires to find the root of a transcendental equation. Do and Vetterli provide a Newton-Raphson algorithm in [4] which involves computation of the Digamma and Trigamma function. For the computational analysis in Section IV-C we refer to their implementation. The starting value for the Newton-Raphson iteration is usually obtained using the moment estimate of β, presented by Birney et al. [35] and Mallat [36]. However, even moment matching is challenging since it requires a numerical solution to a function inversion problem. This problem can either be solved using the combination of a lookup-table and some sort of interpolation method, or by employing the approximation of Krupinski [37]. The author proposes to define an invertible approximation to the aforementioned function and solves a non-linear curve fitting problem for certain ranges of β. Moment matching then reduces to a simple function evaluation. In a very recent publication, Song [38] proposes another alternative method to ML estimation of β. The idea is based on exploiting a convex shape equation, leading to a globally convergent and computationally appealing Newton-Raphson algorithm which is free of any kind of Gamma function. Furthermore, with probability tending to one, the author shows that if we choose $\hat{\beta}_1 \in [\beta, \infty]$, where β denotes the true shape parameter, the Newton-Raphson iteration converges to β for $N \to \infty$. Hence, a good choice for $\hat{\beta}_1$ is 3, since most detail subband coefficient distributions of natural images exhibit $0 < \beta < 3$. This remedies at least the starting value problem. In our retrieval experiments, we assess the interesting question whether it is actually necessary to compute ML estimates, or if it is enough to use moment estimates with low computational requirements.

B. Weibull Parameters

We discuss two ML estimation methods for the parameters α and β of the Weibull distribution. Our experiments show, that by using a theoretical result from statistics, we can significantly decrease the computational effort required to determine $\hat{\alpha}$ and $\hat{\beta}$. First, we present the direct approach of using the log-likelihood function of the Weibull distribution. For that reason, let $x_1, \ldots, x_N$ be a random sample drawn from a Weibull distribution with parameters α, β. According to [23], the MLE of α is the solution to

$$g(\alpha) := \sum_{i=1}^{N} x_i^{\alpha} \log(x_i) - K \sum_{i=1}^{N} x_i^{\alpha} - \frac{1}{\alpha} \sum_{i=1}^{N} x_i^{\alpha} = 0, \tag{15}$$

with $K := \frac{1}{N}\sum_{i=1}^{N} \log(x_i)$. In order to solve (15) using Newton-Raphson, we determine the first derivative $\partial g / \partial \alpha$ as

$$\begin{aligned} g'(\alpha) := \frac{\partial g}{\partial \alpha} = \sum_{i=1}^{N} x_i^{\alpha} \log(x_i)^2 - \\ K\left(\sum_{i=1}^{N} x_i^{\alpha} \log(x_i)\right) + \frac{1}{\alpha^2}\sum_{i=1}^{N} x_i^{\alpha} - \frac{1}{\alpha}\sum_{i=1}^{N} x_i^{\alpha} \log(x_i). \end{aligned} \tag{16}$$

The MLE is then obtained by using the update step $\hat{\alpha}_n = \hat{\alpha}_{n-1} - g(\hat{\alpha}_{n-1})/g'(\hat{\alpha}_{n-1})$ for $n \geq 2$. Subsequently, the MLE of β has the explicit expression:

$$\hat{\beta} = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^{\hat{\alpha}}\right)^{1/\hat{\alpha}} \tag{17}$$

The starting value $\hat{\alpha}_1$ is usually computed by moment matching. Unfortunately, even that requires a numerical procedure, since the parameter estimate $\hat{\alpha}$ is the solution to

$$\frac{\Gamma_3 - 3\Gamma_2\Gamma_1 + 2\Gamma_1^3}{(\Gamma_2 - \Gamma_1^2)^{3/2}} - a_3 = 0, \tag{18}$$

where $\Gamma_k := \Gamma(1 + k/\alpha)$ and

$$a_3 := \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^3}{\left[\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2\right]^{3/2}} \tag{19}$$

denotes the sample skewness. A first approximation of $\hat{\alpha}$ to solve (19) can be obtained from a α-versus-a_3 table and linear interpolation. However, as noted in [26], computational difficulties can arise for ML estimation in cases where $\alpha < 2.2$.

The second, alternative estimation method we discuss is based on the theoretical result, that if a random variable X follows a Weibull distribution, then the random variable $Y = \log(X)$ follows an Extreme Value distribution of type I (Gumbel distribution). Now, let $y_i := \log(x_i)$ using the random sample $x_1, \ldots, x_N$ of the Weibull distribution from above. The PDF of the Gumbel distribution is

$$p(y|\mu,\sigma) = \frac{1}{\sigma}\exp\left(\frac{y-\mu}{\sigma}\right)\exp\left\{-\exp\left(\frac{y-\mu}{\sigma}\right)\right\}, \quad (20)$$

with $-\infty < y < \infty$, $0 < \mu < \infty$ (location) and $\sigma > 0$ (scale). This Extreme Value distribution might be though of as a Log-Weibull distribution [26]. The MLE of σ requires a numerical solution to

$$f(\sigma) := \overline{y} - \sigma - \frac{\sum_{i=1}^{N} y_i \exp\left(-\frac{y_i}{\sigma}\right)}{\sum_{i=1}^{N}\exp\left(-\frac{y_i}{\sigma}\right)} = 0, \quad (21)$$

where $\overline{y}$ denotes the sample mean of the observations. We can determine the first derivative of $f(\sigma)$ w.r.t. σ as

$$f'(\sigma) := \frac{\partial f}{\partial \sigma} = \frac{1}{\sigma^2}\sum_{i=1}^{N} y_i^2 \exp\left(-\frac{y_i}{\sigma}\right) - \sum_{i=1}^{N}\exp\left(-\frac{y_i}{\sigma}\right) - \frac{1}{\sigma}\sum_{i=1}^{N} y_i \exp\left(-\frac{y_i}{\sigma}\right) \quad (22)$$

which now allows to employ the Newton-Raphson algorithm. In contrast to the problematic computation of the starting value $\hat{\alpha}_1$ in case of the direct MLE approach from above, the starting value $\hat{\sigma}_1$ can be easily obtained from the explicit expressions of the moment estimates [26]

$$\hat{\sigma}_1 := \hat{\sigma} = \frac{1}{\pi}\sqrt{6}s \approx 0.779697s \;\text{ and }\; \hat{\mu} = \overline{y} - \gamma\hat{\sigma}, \quad (23)$$

where s denotes the sample standard deviation. Given that we have determined a ML solution $\hat{\sigma}$, the MLE $\hat{\mu}$ then directly follows as

$$\hat{\mu} = \hat{\sigma}\log\left(\frac{1}{N}\sum_{i=1}^{N}\exp\left(\frac{y_i}{\hat{\sigma}}\right)\right). \quad (24)$$

The only thing left to do is to transform the estimates (either MLE or moment estimates) $\hat{\mu}$ and $\hat{\sigma}$ to estimates $\hat{\alpha}$ and $\hat{\beta}$ of the Weibull parameters by using

$$\hat{\alpha} = \exp(\hat{\sigma}) \quad \text{and} \quad \hat{\beta} = \frac{1}{\hat{\mu}}. \quad (25)$$

As we have noted in case of the GGD parameters, we evaluate whether the moment estimates lead to different retrieval results when they are used instead of the MLEs for similarity measurement. In what follows, we use the abbreviation *Weibull-G* in order to refer to the Weibull parameter estimation method via the Gumbel distribution.

C. Gamma Parameters

In order to estimate the parameters α and β of the Gamma distribution, we follow the approach presented in [39]. Given that $x_1, \ldots, x_N$ denotes a random sample drawn from a Gamma distribution with parameters α and β, then the explicit expression for the Newton-Raphson update step is

$$\hat{\alpha}_n = \hat{\alpha}_{n-1} - \frac{\log(\hat{\alpha}_{n-1}) - \psi(\hat{\alpha}_{n-1}) - M}{1/\hat{\alpha}_{n-1} - \psi'(\hat{\alpha}_{n-1})}, \quad (26)$$

for $n \geq 2$, where $\psi(\cdot)$ denotes the Digamma function, $\psi'(\cdot)$ denotes the Trigamma function and

$$M := \log(\overline{x}) - \frac{1}{N}\sum_{i=1}^{N}\log(x_i). \quad (27)$$

The MLE of β then follows as

$$\hat{\beta} = \frac{\hat{\mu}}{\overline{x}}. \quad (28)$$

We can significantly reduce the computational overhead to evaluate the Digamma and Trigamma function by using a lookup-table and linear interpolation for example. A starting value $\hat{\alpha}_1$ is obtained from the moment estimates [40]

$$\hat{\alpha}_1 := \hat{\alpha} = \left(\frac{\overline{x}}{s}\right)^2 \;\text{ and }\; \hat{\beta} = \frac{s^2}{\overline{x}}. \quad (29)$$

Note, that no computationally expensive operations have to be performed to estimate these starting values.

IV. Computational Analysis

In this section we present an in-depth computational analysis in terms of arithmetic operations for the main building blocks of our image retrieval framework: image representation, similarity measurement and parameter estimation. This analysis is a necessary step, since it allows to quantify the term *lightweight*. We further present a comparative runtime analysis using MATLAB implementations of the different estimation methods and similarity measurement parts.

A. Image Representation

Besides its advantages for image analysis, the DT-CWT is appealing from a computational point of view, since it can be implemented very efficiently by four parallel pyramidal DWTs using appropriate filter sets. Regarding space requirements, the DT-CWT is an overcomplete transform with a redundancy factor of four in case of 2-D images. In contrast, the DCT (used in [5]) is non-redundant, the Steerable Pyramid [7] is overcomplete by a factor of $4k/3$ (k denotes the number of orientation subbands) and the Stationary Wavelet Transform (SWT) [41] is overcomplete by a factor of $3J$, where J denotes the maximum decomposition depth. The computational complexity of the DT-CWT is linear $\mathcal{O}(N)$ in the number of input pixels N, since it requires computation of four parallel DWT decompositions which are of linear complexity. Hence, both DWT and DT-CWT differ only by a constant factor. For comparison, the DCT, SWT, Steerable Pyramid and Gabor wavelets (when implemented in the frequency domain) have complexity $\mathcal{O}(N\log N)$. However, in case of a block-based DCT with 8×8 blocks, the $\log N$ term carries no weight compared to a full-frame DCT.

B. Similarity Measurement

In the classic retrieval scenario, the similarity measurement part is most critical for runtime performance since each new query image requires computation of the similarity measure for all candidate images in the database. In case the statistical model parameters of the database images are estimated at the time of storage, the runtime performance of the retrieval task is completely determined by the performance of the similarity measurement process. Although all presented KL-divergences can be computed with constant complexity (except for the RMM model, where we use MC simulation), it is interesting to take a closer look at the required arithmetic operations. By *arithmetic operations* we understand the number of additions/subtractions and multiplications/divisions (basic arithmetic operations) as well as the computationally expensive $\log$, e^x and x^r operations with $x, r \in \mathbb{R}$. We further take into account any non-trivial operation, such as the evaluation of the Gamma $\Gamma(\cdot)$ or the Digamma $\psi(\cdot)$ function. To avoid numerical difficulties, we compute $\log\Gamma(\cdot)$ instead of $\Gamma(\cdot)$ at the cost of perhaps one additional exponentiation. The function values of $\log\Gamma(\cdot)$ and $\psi(\cdot)$ are obtained by employing lookup-tables and the method of linear interpolation. Both operations can be performed with constant complexity and only require basic arithmetic (e.g. 5 additions/subtractions, 4 multiplications/divisions and 2 table-lookups in our implementation). To conduct a relative runtime measurement, all KL-divergences are implemented in MATLAB. The runtime is measured on a Intel Core2 Duo 2.66Ghz system with 2GB of memory running MATLAB 7.6. We emphasize that the focus is on relative runtime differences not on absolute values. Given that the statistical model parameters of an arbitrary wavelet subband are available (i.e. pre-computed) for the query and all L database images, we simulate a search for $L = 10^4$. Table II lists the number of arithmetic operations for each KL-divergence as well as the runtime relative to the longest runtime (marked bold). As we can see, the KL-divergence for the GGD has the worst performance, due to the computations of $\log\Gamma(\cdot)$. The KL-divergence of the Gamma model shows slightly worse runtime performance than the KL-divergence for the Weibull model which can be attributed to computation of $\psi(\cdot)$ and the additional $\log\Gamma(\cdot)$. The Rayleigh KL-divergence performs best but unfortunately the model as such is too inflexible, as we have seen in Section II-D. Last, we note that since all KL-divergences have a closed-form expression, no histogram computation and discrete version of the KL-divergence is required. In practice, this is a huge advantage since we only have to store the model parameters per image and further avoid the search for a reasonable histogram binning.

TABLE II
NUMBER OF ARITHMETIC OPERATIONS FOR SIMILARITY MEASUREMENT BETWEEN TWO ARBITRARY SUBBANDS

Model	$\pm$	$\times, \div$	e^x, x^r $\log$	$\log\Gamma$	ψ	Relative Runtime
GGD, see (5)	6	10	3	4	0	**1.00**
Gamma, see (12)	6	5	1	2	1	0.56
Weibull, see (10)	8	9	8	1	0	0.31
Rayleigh, see (8)	2	4	1	0	0	0.01

TABLE III
NUMBER OF ARITHMETIC OPERATIONS FOR NEWTON-RAPHSON AS A FUNCTION OF THE SIGNAL LENGTH N

Model	$\pm$	$\times, \div$	$\lvert\cdot\rvert$	e^x, x^r $\log$	ψ, ψ'	Relative Runtime
GGD, MLE [4]	$3N$	$2N$	N	$2N$	2	0.76
GGD, Song [38]	$4N$	$3N$	N	$2N$		**1.00**
Weibull-G	$4N$	$3N$		N		0.21
Weibull	$4N$	$2N$		$2N$		0.62
Gamma	$2N$	4		N	2	0.21

C. Parameter Estimation

Except for the Rayleigh model, all ML parameter estimation procedures require numerical root-finding to obtain estimates. Since we can determine the derivatives of the log-likelihood functions w.r.t. the relevant parameters in all cases, it is reasonable to use the Newton-Raphson algorithm due to its good convergence properties. However, optimal (i.e. quadratic) convergence is only possible if the first estimate (starting value) is close to the actual root. To fulfill this requirement, we use moment estimates for the Gamma, Weibull and GGD model. By employing the invertible approximation proposed in [37] in case of the GGD and the Weibull-G moment matching method in case of the Weibull distribution, we can at least eliminate the issue of computationally intensive starting value calculations. The exact computational requirements for moment matching will be discussed later. To get an impression of the computational demand in each iteration step of the Newton-Raphson algorithm, we determine the number of required arithmetic operations. For the exact expressions of the update steps in case of the GGD MLE approach and Song's method, we refer to [4] and [38]. We optimize computation in such a way, that terms (e.g. summations, logarithms, etc.) which occur repeatedly in an iteration step are temporary stored for further use. Since many operations depend on the signal length N, we omit any additional constants for readability reasons in these cases. The number of arithmetic operations per iteration and the runtime performance of the ML estimation procedures relative to the longest runtime (marked bold) are listed Table III. Further, Fig.4 shows a boxplot of the mean estimation times over a set of parameter values for all ML estimation approaches. For each parameter value, estimation is repeated 100 times on 10^5 random numbers drawn from the corresponding statistical model. ML estimation using the Weibull-G approach shows the best performance, with only one iteration on average to reach convergence (i.e. absolute difference of two successive estimates is less than 10^{-6}). In contrast, direct estimation of the Weibull parameters is less competitive, mainly due to the impact of starting value computation which takes about 50% of the total runtime. The Gamma MLE procedure performs good as well, although the number of iterations is the limiting factor here, since one Newton-Raphson update step requires fewer arithmetic operations compared to the Weibull-G approach. As expected, the complex update step of the GGD model with more $\log, x^r, e^x$ operations leads to an increase in computation time compared to the Weibull-G or Gamma model. Regarding the number of iterations, we confirm the results of [4] with

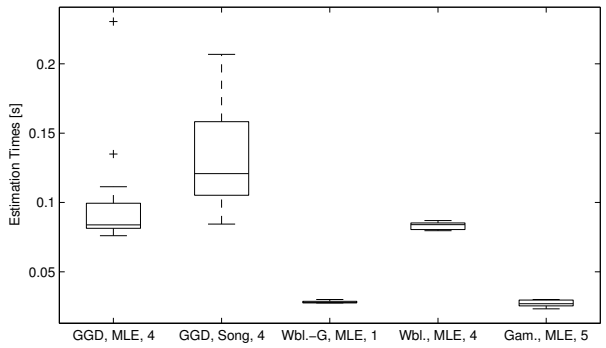


Fig. 4. Boxplot of the model parameter estimation times, together with the average number of Newton-Raphson iterations to reach the defined convergence bound.

TABLE IV
NUMBER OF ARITHMETIC OPERATIONS FOR MOMENT MATCHING AS A FUNCTION OF THE SIGNAL LENGTH N

Model	$\pm$	$\times, \div$	$\lvert\cdot\rvert$	e^x, x^r log	$\log\Gamma$	Relative Runtime
GGD [37]	$2N$	N	N	3	2	0.15
Weibull-G, see (23)	$3N$	N		N		**1.00**
Gamma, see (29)	$3N$	N				0.32

four iterations on average to reach convergence. The estimation approach proposed by Song [38] shows the worst runtime performance and a quite strong dispersion as well. A closer look at the number of iterations for each shape parameter reveals an average of 10 iterations for $\beta < 1.0$ which slightly distorts the average. This seems reasonable, since the starting value of $\hat{\beta}_1 = 3$ is actually far-off the true value in these situations. We omit the estimation effort for the RMM, since comparing an EM algorithm to MLE procedures is not fair. The estimation time for a RMM subband model with 65536 coefficients is in the order of seconds, i.e. by a factor of at least one magnitude slower than the method of Song for example.

As a last point, we assess the number of arithmetic operations to compute moment estimates of the GGD, Gamma and Weibull distribution. A careful analysis of moment estimation is reasonable, since we use these estimates as a fast alternative to the MLEs in our experiments. The corresponding numbers are listed in Table IV. We emphasize, that this is the total effort to compute a first parameter estimate. No iterative procedures are necessary and mostly basic arithmetic operations are performed. Only in case of Weibull-G moment estimation, the logarithm operation is dependent on the signal length N. This fact is reflected in the relative runtime differences because the logarithm is an expensive operation compared to addition/subtraction or multiplication/division. The fast approximative GGD parameter estimation of [37] shows the best performance since the expensive computations like $\log\Gamma(\cdot)$, e^x or $\log$ do not depend on the signal length N. Further, this approach apparently benefits from our lookup-table implementation of $\log\Gamma(\cdot)$. Regarding moment estimation of the Gamma parameters, we note that this approach basically requires to compute the sample mean and sample standard deviation and hence performs at a competitive level compared to [37] as well.

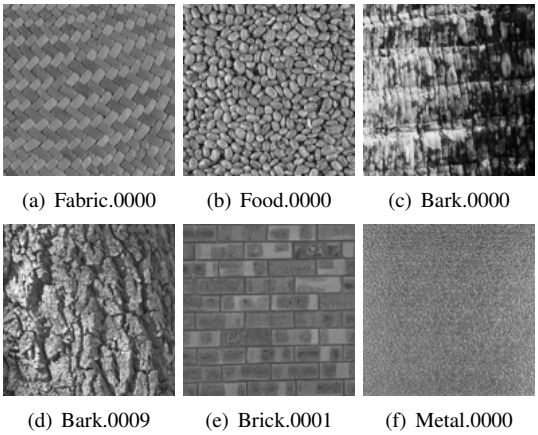
V. EXPERIMENTAL RESULTS

The objective of the experimental section is to cover three important issues: first, we address the impact of either using moment estimates or ML estimates on the retrieval results. Second, we compare the retrieval performance of the DT-CWT and the proposed statistical models to previous research work, including the approaches presented in [3], [4] and the classic mean/standard deviation features of [42] (abbreviated by *Classic*). Third, based on the computational analysis of the previous section and the achieved retrieval rates, we intend to give a guideline for lightweight retrieval w.r.t. the scenarios of Section I.

We work with a selection of images from the popular MIT Vision Texture Database (VisTex) [43], consisting of 40 textures which have already been extensively used in texture image retrieval literature. All images are first converted to the LUV color model and only the luminance (L) channel information is retained. The 512×512 pixel versions of the textures are split into 16 non-overlapping subimages (128×128 pixel). The only preprocessing step is to normalize the subimages by subtracting the pixel mean and dividing by the standard deviation. A selection of example textures is shown in Fig.5. Regarding the filter sets for the wavelet transforms, we use the CDF 9/7 filter [44] for the pyramidal DWT and Kingsbury's Q-Shift (14,14)-tap filters (levels ≥ 2) in combination with (13,19)-tap near-orthogonal filters (level 1) for the DT-CWT (see [45]).

The experimental setup for the reference approach of [3] is as follows: we extract the first 16 DCT coefficients (in zigzag scan order) of a sliding 8×8 window (step size of two pixel in vertical and horizontal direction) as features and then fit a multivariate GMM with eight components using the classic EM algorithm. The EM algorithm terminates if either 200 iterations are reached or the log-likelihood difference between two successive iterations is less than 10^{-6}. Covariance matrices are restricted to diagonal matrices and are regularized by a small $\epsilon > 0$ to ensure positive definiteness. The starting parameters of the EM algorithm are initialized according to [20], where the author uses the Linde-Buzo-Gray (LBG, aka generalized Lloyd) algorithm [46] with the codeword splitting procedure proposed by Gray [47]. LBG terminates when the difference between the average distortions of two successive iterations is less than 10^{-3}. For similarity measurement, we compute the feature-likelihood as proposed in [3]. Features are extracted from the first 16 DCT coefficients of all non–overlapping 8×8 blocks. Hence, we obtain 256 feature vectors for each query image. We further evaluate the impact of using only every second or every fourth block. This gives 128 or 64 query feature vectors (QV), resp. Other possible choices for similarity measurement are the Asymptotic-Likelihood approximation of [20] or the approximations proposed by Goldberger [48]. In what follows, we abbreviate the approach by DCT (MGMM, EM) and additionally add the number of extracted QVs when necessary.

To evaluate the performance of the retrieval system we determine the number of correctly retrieved images among the top K matches. A retrieval result is *correct* if an image of



(a) Fabric.0000 (b) Food.0000 (c) Bark.0000

(d) Bark.0009 (e) Brick.0001 (f) Metal.0000

Fig. 5. Six example textures of the VisTex [43] database.

the same parent is retrieved. Formally, let $Q := \{r_1, \ldots, r_B\}$ denote the set of correct membership indices to a given query, where B denotes the number of subimages. Further, let $\{q_1, \ldots, q_K\}$ denote the index set of the top K matches. The percentage of correctly retrieved images is then calculated as follows:

$$s_K = \frac{1}{B} \sum_{i=1}^{K} \mathbf{1}_Q(q_i), \text{ with } \mathbf{1}_Q(x) := \begin{cases} 1, & \text{if } x \in Q, \\ 0, & \text{else} \end{cases} \tag{30}$$

Since each image is split into 16 subimages, we set $K = B = 16$. Table V lists the percentage of correctly retrieved images among the top $K = 16$ matches with decomposition depths less than three. Table VI provides detailed texture-specific results, where the highest retrieval rates per texture are marked bold (in case of no ties).

In case of the pyramidal DWT, we can almost reproduce the results of [4] and further observe a slight trade-off between estimation accuracy and retrieval performance when comparing the rates for ML estimation and moment matching. Regarding the DT-CWT results, the situation is somewhat different. The estimation accuracy vs. retrieval rate trade-off is only evident in case of the Weibull-G model, since moment matching for the Gamma parameters leads to the second-best overall retrieval results with 81.73%. This is 3.21 percentage points less than the DCT (MGMM, EM, 256) rate and 0.94 percentage points above the rate of the DT-CWT (RMM, EM) reference model. Figures 6(a) and 6(b) show a receiver operating characteristic (ROC) curve comparison when relying on either moment matching or ML estimation. The figures are obtained by plotting the retrieval rate $100 \cdot s_K$ as a function of the number of retrieved images K. We conclude, that the differences in retrieval accuracy are rather small, no matter which estimation method we choose. Nevertheless, our observations lead to the presumption that ML estimation does not necessarily lead to the best retrieval results.

With respect to the quality of the proposed statistical models for the DT-CWT coefficient magnitudes, we see that the DT-CWT (Weibull-G, MLE) as well as the DT-CWT (Gamma, MM) approach clearly improve retrieval performance by 5.43 and 3.91 percentage points, resp., compared to the DWT (GGD, MLE) results of [4]. Even the combinations DT-CWT (Rayleigh, MLE) and DT-CWT (Classic) show acceptable performance, although the Rayleigh model is a very coarse approximation to the distribution of the complex coefficient magnitudes (see Table I) and the *Classic* features do not assume any statistical model at all. We thus conjecture that the increase in retrieval performance cannot be exclusively attributed to the statistical models, but in a large part to the choice of image representation. The advantages of the DT-CWT over the DWT (see Section II-A) are definitely reflected in the experimental results. To visualize the retrieval performance of the top approaches, Fig.6(c) shows a ROC curve comparison between the DWT (GGD, MLE) [4], DT-CWT (Gamma, MM), DT-CWT (RMM, EM) and the DCT (MGMM, EM, 256) [3] approach. It is evident that the slope of the curves becomes more shallow as the number of retrieved images K increases. This implies that the difference in retrieval accuracy becomes increasingly relevant as K grows, since significantly more retrieved images are required to achieve equal rates.

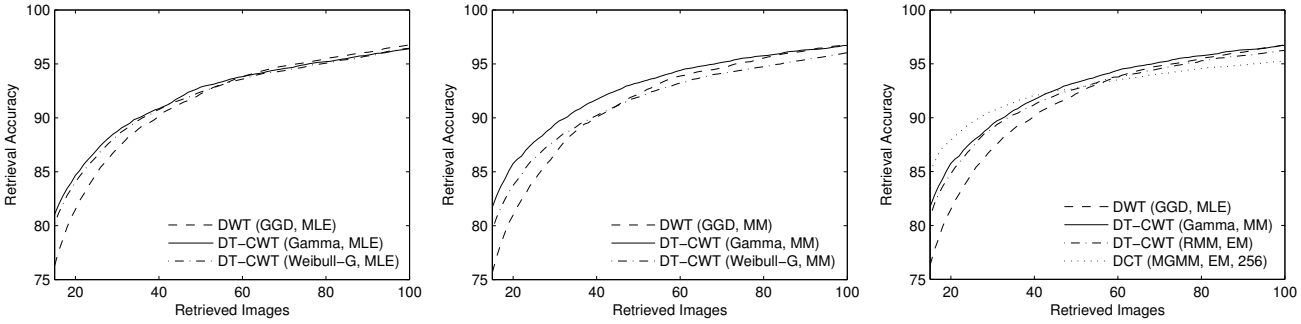
With respect to the two retrieval scenarios we mentioned in Section I, it seems plausible that not all presented approaches are applicable in both situations. Considering the first scenario from a purely computational viewpoint, all models except the DT-CWT (RMM, EM) are practicable since the complexity of similarity measurement is the primary concern. However, we emphasize that the numbers in Table II indicate considerable runtime differences for the KL-divergences which can be a relevant issue in cases of large databases. In case we accept to to trade computational performance for retrieval accuracy, the DCT (MGMM, EM, 256) approach of [3] is the first choice. Although the computation of the feature likelihood is expensive compared to constant complexity similarity measurement with closed-form expressions for the KL-divergences, we get a remarkable increase of 3.21 percentage points. Decreasing the computational cost by reducing the number of QVs to 64 still leads to the best results with 82.28%, however, the margin to the DT-CWT (Gamma, MM) approach shrinks to only 0.55%. Taking the retrieval results into account, we recommend the the DT-CWT (Gamma, MM) approach for very large databases where no indexing is in place and favor the DCT (MGMM, EM) approach otherwise. With respect to the second retrieval scenario, the situation is different. Virtually all parts of the retrieval system are constrained by computational limitations, hence only a few approaches remain practicable. As a general guideline, we propose to switch from ML estimates to moment estimates since we cannot report any significant decrease in retrieval performance. With the results of our computational study in mind, we favor the DT-CWT (Gamma, MM) approach which achieves constant complexity similarity measurement, linear-complexity parameter estimation and linear-complexity image transformation at the second-highest retrieval rate of our experiments.

VI. Conclusion

In this work, we analyzed the framework of probabilistic texture retrieval in the wavelet domain from the viewpoints of retrieval accuracy and computational performance.

TABLE V
TOP $K = 16$ RETRIEVAL RATES [%]

Scales	DT-CWT							DWT			DCT & MGMM [3]
	Rayleigh	Gamma		Weibull-G		Classic	RMM	GGD [4]		Classic	
		MLE	MM	MLE	MM			MLE	MM		
1	62.96	72.83	73.07	71.80	71.94	67.49	74.25	67.53	67.27	59.77	84.94 (256 QV)
1,2	68.44	77.27	77.74	76.11	76.04	70.91	77.93	74.43	73.83	64.55	84.53 (128 QV)
1,2,3	73.96	81.06	81.73	80.21	79.61	75.99	80.79	76.30	75.65	65.50	82.28 (64 QV)



(a) Maximum-Likelihood (b) Moment Matching (c) Top Results (incl. [3], [4])

Fig. 6. ROC curves of the retrieval rate as a function of the number of retrieved images K.

TABLE VI
PER-TEXTURE RETRIEVAL RATES ($K = 16$) [%]

Texture	DWT		DT-CWT						DCT & MGMM (256 QV) [3]
	GGD [4]		Rayleigh	Weibull		Gamma		RMM	
	MLE	MM		MLE	MM	MLE	MM		
Bark.0000	60.16	55.86	54.69	**64.45**	63.67	62.50	55.47	60.94	53.12
Bark.0006	52.34	50.78	49.22	57.81	56.25	60.16	63.67	58.20	**71.48**
Bark.0008	72.66	69.92	74.22	82.81	78.12	81.64	86.72	88.28	**92.58**
Bark.0009	58.20	57.81	58.98	64.84	63.28	63.67	63.67	58.98	**70.70**
Brick.0001	71.09	72.27	76.95	80.08	78.12	79.69	83.98	79.30	**96.09**
Brick.0004	63.67	61.33	50.78	70.70	73.44	**75.39**	74.61	71.09	67.19
Brick.0005	85.16	77.73	82.03	92.19	92.58	92.58	90.23	**93.36**	76.95
Buildings.0009	93.36	92.97	91.41	94.92	96.09	95.70	96.09	94.53	93.36
Fabric.0000	85.55	85.16	80.08	91.41	91.02	94.14	96.48	95.31	**99.61**
Fabric.0004	64.45	65.62	63.28	66.41	66.41	66.41	**67.97**	64.45	59.77
Fabric.0007	96.88	96.88	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Fabric.0009	82.81	80.86	89.06	99.61	99.61	99.61	99.61	94.53	95.31
Fabric.0011	82.81	83.59	71.09	78.52	75.78	78.91	83.98	85.55	**91.80**
Fabric.0014	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Fabric.0015	99.61	99.61	94.92	98.44	98.05	99.22	99.22	99.61	99.61
Fabric.0017	86.33	86.33	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Fabric.0018	89.84	90.23	99.22	100.00	100.00	100.00	100.00	100.00	99.61
Flowers.0005	58.20	55.86	53.12	65.23	65.62	69.53	66.41	69.53	**94.14**
Food.0000	78.52	75.00	82.42	92.19	92.58	94.53	96.09	96.88	**100.00**
Food.0005	86.72	86.33	93.36	97.27	97.66	98.05	**99.61**	99.22	98.83
Food.0008	96.48	97.27	90.62	93.75	89.06	99.61	99.61	98.83	**100.00**
Grass.0001	64.06	64.84	57.42	71.48	69.14	69.14	67.97	69.53	**87.11**
Leaves.0008	64.45	66.80	**79.69**	72.66	71.88	66.80	69.14	75.39	59.77
Leaves.0010	34.38	32.03	21.48	32.42	32.81	36.33	37.89	36.72	**66.02**
Leaves.0011	82.42	78.52	76.56	84.77	82.42	82.81	78.52	80.08	**94.92**
Leaves.0012	73.83	76.95	69.14	75.78	75.78	78.52	86.33	82.81	**94.14**
Leaves.0016	83.98	81.25	55.08	80.08	75.39	80.08	83.20	67.97	**89.45**
Metal.0000	68.75	69.14	66.80	66.80	67.97	71.48	73.83	66.80	**82.81**
Metal.0002	99.61	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Misc.0002	76.56	76.17	66.02	76.95	80.08	81.64	85.16	83.98	**96.09**
Sand.0000	76.95	76.95	81.25	91.02	88.67	92.58	92.58	94.14	**94.53**
Stone.0001	48.83	59.77	**69.53**	60.94	59.38	57.81	58.59	58.20	32.42
Stone.0004	81.25	80.86	73.83	78.91	78.52	80.86	85.16	80.86	**85.55**
Terrain.0010	56.64	52.73	50.78	57.81	57.81	60.55	60.16	48.05	**64.45**
Tile.0001	51.56	51.95	38.28	51.56	49.61	54.30	56.25	54.69	**76.17**
Tile.0004	**99.61**	97.66	82.03	98.83	99.22	99.22	98.44	96.48	98.44
Tile.0007	99.61	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Water.0005	96.48	96.09	92.97	94.14	95.70	**96.88**	96.09	92.19	96.48
Wood.0001	**47.66**	45.70	26.56	29.30	30.08	32.42	30.47	37.50	38.28
Wood.0002	80.47	77.34	95.31	94.53	92.58	89.84	85.94	**97.66**	80.86
Avg.	76.30	75.65	73.96	80.21	79.61	81.06	81.73	80.79	**84.94**

We introduced a novel retrieval approach based on image representation in the complex wavelet domain and several statistical models for the magnitude of the complex transform coefficients. We further presented closed-form expressions for the KL-divergences between the proposed statistical models, thus allowing constant complexity similarity measurement. By assessing the impact of using moment estimates instead of ML estimates for computing the KL-divergence, we showed that the difference between estimation and retrieval accuracy is negligible and we can achieve a remarkable improvement w.r.t. computational performance. Unexpectedly, the DT-CWT (Gamma, MM) model even lead to the second-best retrieval rate. In general, the experimental results indicate superior performance of our approach compared to [4] and competitive performance to a significantly more complex method [3] (in terms of computational cost). In addition, we observed that all DT-CWT based approaches apparently benefit from the transform-specific advantages of the DT-CWT over the pyramidal DWT. Future research includes three particular points: first, we plan to evaluate the quality of our models for natural images. Second, we focus on a rotational-invariant extension while keeping the computational cost as low as possible. Third, the reasonable incorporation of color information is still an open issue and poses a challenging task.

APPENDIX A
EM ALGORITHM FOR RAYLEIGH MIXTURES

The following steps provide an EM algorithm for estimating the parameters of a K-component Rayleigh Mixture Model (RMM). Let $x_1, \ldots, x_N$ denote a sample drawn from the RMM model and let π_k denote the prior probability of selecting the k-th component. Further, let $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_K]$ and $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]$. The EM steps are as follows:

1) **Initialize:** $\boldsymbol{\theta}^{(0)} = \left[\boldsymbol{\alpha}^{(0)}\ \boldsymbol{\pi}^{(0)}\right]$ and $m \leftarrow 0$
2) **E-Step:** Evaluate the conditional probability of selecting component k given datapoint x_n and current model parameters $\boldsymbol{\theta}^{(m)}$ as follows:

$$p^{(m)}(k|n) = \frac{\pi_k^{(m)} p\left(x_n|\alpha_k^{(m)}\right)}{\sum_j \pi_j^{(m)} p\left(x_n|\alpha_j^{(m)}\right)}, \tag{31}$$

3) **M-Step:** Determine $\pi_k^{(m+1)}$, $\alpha_k^{(m+1)}$ $\forall k$ as follows:

$$\pi_k^{(m+1)} = \frac{1}{N} \sum_{n=1}^{N} p^{(m)}(k|n) \tag{32}$$

$$\alpha_k^{(m+1)} = \frac{\sum_{n=1}^{N} p^{(m)}(k|n) x_n^2}{2 \sum_{n=1}^{N} p^{(m)}(k|n)} \tag{33}$$

4) Evaluate the log-likelihood $\log p\left(\mathbf{x}|\boldsymbol{\theta}^{(m+1)}\right)$ as follows:

$$\sum_{n=1}^{N} \log \left\{ \sum_{k=1}^{K} \pi_k^{(m+1)} p\left(x_n|\alpha_k^{(m+1)}\right) \right\} \tag{34}$$

and check for convergence of the log-likelihood, i.e. $|\log p\left(\boldsymbol{x}|\boldsymbol{\theta}^{(m+1)}\right) - \log p\left(\boldsymbol{x}|\boldsymbol{\theta}^{(m)}\right)| < \epsilon$ with $\epsilon = 1^{-6}$ for example. In case the termination criteria is not met, set $m \leftarrow m+1$ and goto Step 2.

ACKNOWLEDGMENT

The authors like to thank Dr. Nick Kingsbury for providing the MATLAB code to the Dual-Tree Complex Wavelet Transform (v4.3). The full MATLAB code for our work is available under http://www.wavelab.at/sources.

REFERENCES

[1] N. Vasconcelos and A. Lippman, "Library-based coding: A representation for efficient video compression and retrieval," in *Proceedings of the Data Compression Conference (DCC'97)*, Snowbird, Utah, USA, 1997, pp. 121–130.
[2] ——, "A Unifying View of Image Similarity," in *Proceedings of the 15th International Conference on Pattern Recognition (ICPR'00)*, vol. 1, Barcelona, Spain, 2000, pp. 38–41.
[3] ——, "A probabilistic architecture for content-based image retrieval," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'00)*, Hilton Head, South Carolina, USA, 2000, pp. 216–221.
[4] M. Do and M. Vetterli, "Wavelet-based texture retrieval using Generalized Gaussian density and Kullback-Leibler distance," *IEEE Transactions on Image Processing*, vol. 2, no. 2, pp. 146–158, Feb. 2002.
[5] N. Vasconcelos and G. Carneiro, "What is the role of independence for visual recognition," in *Proceedings of the European Conference in Computer Vision (ECCV'02)*, Copenhagen, Denmark, 2002, pp. 297–311.
[6] M. Do and M. Vetterli, "Rotation invariant texture characterization and retrieval using steerable Wavelet-Domain Hidden Markov models," *IEEE Transactions on Multimedia*, vol. 4, no. 4, pp. 517–527, Dec. 2002.
[7] E. Simoncelli and W. Freeman, "The Steerable Pyramid: A flexible architecture for multi-scale derivative computation," in *Proceedings of the IEEE International Conference on Image Processing (ICIP'95)*, vol. 3, Washington, DC, USA, 1995, pp. 444–447.
[8] M. Do, "Fast approximation of Kullback-Leibler distance for dependence trees and Hidden Markov models," *IEEE Signal Processing Letters*, vol. 10, no. 4, pp. 115–118, Apr. 2003.
[9] G. Tzagkarakis and P. Tsakalides, "A statistical approach to texture image retrieval via alpha-stable modeling of wavelet decompositions," in *Proceedings of the 5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '04)*, Lisbon, Portugal, 2004.
[10] G. Tzagkarakis, B. Beferull-Lozano, and P. Tsakalides, "Rotation-invariant texture retrieval with Gaussianized Steerable Pyramids," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2702–2718, Sep. 2006.
[11] E. de Ves, A. Ruedin, D. Acevedo, C. Benavent, and L. Seijas, "A new wavelet-based texture descriptor for image retrieval," *Lecture Notes in Computer Science, Computer Analysis of Images and Patterns*, vol. 4673, pp. 895–902, Aug. 2007.
[12] N. A. Heckert and J. J. Filliben, *NIST Handbook 148: DATAPLOT Reference Manual*. National Institute of Standards and Technology Handbook Series, 2003, vol. 1.
[13] T. Lee, "Image representation using 2D Gabor wavelets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959–971, Oct. 1996.
[14] T. Randen and J. Husoy, "Filtering for texture classification: A comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 291–310, Apr. 1999.
[15] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America*, vol. 2, no. 7, pp. 1160–1169, Jul. 1985.
[16] N. Kingsbury, "The Dual-Tree Complex Wavelet Transform: a new technique for shift-invariance and directional filters," in *Proceedings of the 8th IEEE DSP Workshop*, Bryce Canyon, Utah, USA, Aug. 1998, pp. 9–12.
[17] I. Selesnick, R. Baraniuk, and N. Kingsbury, "The Dual-Tree Complex Wavelet Transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, Nov. 2005.
[18] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. Wiley-Interscience, 2001.
[19] L. Devroye, L. Gyorfi, and L. Lugosi, *Probabilistic Theory of Pattern Recognition*. Springer, 1996.
[20] N. Vasconcelos, "On the efficient evaluation of probabilistic similarity functions for image retrieval," *IEEE Transactions on Information Theory*, vol. 50, no. 7, pp. 1482–1496, Jul. 2004.

[21] R. Clarke, *Transform Coding of Images*. Academic Press, 1985.
[22] S. Nadarajah, "A generalized Normal distribution," *Journal of Applied Statistics*, vol. 32, pp. 685–694, Sep. 2005.
[23] K. Krishnamoorthy, *Handbook of Statistical Distributions with Applications*. Chapman & Hall, 2006.
[24] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1964.
[25] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using Hidden Markov models," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 886–902, Apr. 1998.
[26] A. Cohen and B.-J. Whitten, *Parameter estimation in reliability and life space models*. Marcel-Dekker, 1988.
[27] G. Moser, J. Zerubia, and S. Seprico, "SAR amplitude probability density function estimation based on Generalized Gaussian model," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1429–1442, Jun. 2006.
[28] E. Kuruoglu and J. Zerubia, "Modeling SAR images with a generalization of the Rayleigh distribution," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 527–533, Apr. 2004.
[29] R. Kwitt and A. Uhl, "Modeling the marginal distributions of complex wavelet coefficient magnitudes for the classification of Zoom-Endoscopy images," in *Proceedings of the IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'07)*, Rio de Janeiro, Brazil, 2007, pp. 1–8.
[30] ——, "Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval," in *Proceedings of the 15th IEEE International Conference on Image Processing (ICIP'08)*, San Diego, California, USA, 2008, pp. 933–936.
[31] J. Mathiassen, A. Skavhaug, and K. Bo, "Texture similarity measure using Kullback-Leibler divergence between Gamma distributions," in *Proceedings of the 7th European Conference on Computer Vision-Part III (ECCV'02)*, Copenhagen, Denmark, 2002, pp. 133–147.
[32] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.
[33] P. Moreno, P. Ho, and N. Vasconcelos, "A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications," in *Proceedings of the Neural Information Processing Conference (NIPS'03)*, Vancouver, Canada, 2003.
[34] M. Varanasi and B. Aazhang, "Parameteric Generalized Gaussian density estimation," *Journal of the Acoustical Society of America*, vol. 86, no. 4, pp. 1404–1415, Oct. 1989.
[35] K. Birney and T. Fischer, "On the modeling of DCT and subband image data for compression," *IEEE Transactions on Image Processing*, vol. 4, no. 2, pp. 186–193, Feb. 1999.
[36] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
[37] R. Krupinski and J. Purczynski, "Approximated fast estimator for the shape parameter of Generalized Gaussian distribution," *Signal Processing*, vol. 86, pp. 205–211, Feb. 2006.
[38] K.-S. Song, "A globally convergent and consistent method for estimating the shape parameter of a Generalized Gaussian distribution," *IEEE Transaction on Information Theory*, vol. 52, no. 2, pp. 510–527, Feb. 2006.
[39] S. Choi and R. Wette, "Maximum likelihood estimation of the parameters of the Gamma distributon and their bias," *Technometrics*, vol. 11, no. 4, pp. 683–690, Nov. 1968.
[40] M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*, 3rd ed., ser. Wiley Series in Probability and Statistics. Wiley, 2000.
[41] G. Nason and B. Silverman, "The Stationary Wavelet Transform and some statistical applications," *Lecture Notes in Statistics*, vol. 103, pp. 281–300, 1995.
[42] B. Manjunath and W. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, Aug. 1996.
[43] "MIT vision and modeling group," [Online], available: http://vismod.media.mit.edu.
[44] I. Daubechies, *Ten Lectures on Wavelets*, ser. Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1982, no. 61.
[45] N. Kingsbury, "Complex wavelets for shift-invariant analysis and filtering of signals," *Journal of Applied Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, May 2001.
[46] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.
[47] R. Gray, "Vector quantization," *IEEE Accoustics, Speech, and Signal Processing (ASSP) Magazine*, vol. 1, no. 2, Apr. 1984.
[48] J. Goldberger, S. Gordon, and H. Greenspan, "An efficient image similarity measure based on approximations of the KL-divergence between two Gaussian mixtures," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV'03)*, Nice, France, 2003, pp. 487–493.

Roland Kwitt received the M.S. degree in telecommunication systems engineering from the University of Applied Sciences Salzburg in 2005 and the M.S. degree in computer sciences from the University of Salzburg in 2007. He is currently concluding his PhD thesis at the Department of Computer Sciences (University of Salzburg) where he has been a research assistant since 2007. His research interests include statistical image analysis, pattern recognition, image retrieval and multimedia security.

Andreas Uhl received the M.S. degree in mathematics and the PhD degree in computer science from the University of Salzburg in 1992 and 1996, respectively. He is currently an associate professor at the Department of Computer Sciences at the University of Salzburg where he heads the Multimedia Processing and Security Lab. His research interests include image and video processing and compression, wavelets, media security, medical imaging, biometrics, and number-theoretical numerics.