Texture Retrieval via the Scattering Transform

Alexander Sagel, Dominik Meyer, Student Member, IEEE and Hao Shen, Member, IEEE

Abstract—This work studies the problem of content-based image retrieval, specifically, texture retrieval. Our approach employs a recently developed method, the so-called Scattering transform, for the process of feature extraction in texture retrieval. It shares a distinctive property of providing a robust representation, which is stable with respect to spatial deformations. Recent work has demonstrated its capability for texture classification, and hence as a promising candidate for the problem of texture retrieval. Moreover, we adopt a common approach of measuring the similarity of textures by comparing the subband histograms of a filterbank transform via the Kullback-Leibler divergence. Despite the popularity of describing histograms using parametrized probability density functions, such as the Generalized Gaussian Distribution (GGD), it is unfortunately not applicable for describing most of the Scattering transform subbands, due to the complex modulus performed on each one of them. In this work, we propose to use the Weibull distribution to model the Scattering subbands of descendant layers. Our numerical experiments demonstrated the effectiveness of the proposes approach, in comparison with several state of the arts.

Index Terms—Content-based image retrieval, feature extraction, Kullback-Leibler divergence, Scattering transform, similarity measure, Weibull distribution.

I. INTRODUCTION

A. Overview

CONTENT-BASED image retrieval (CBIR) is a special case of image classification. It can be viewed as the process of assigning a query image to a set of image classes, where each class represents a database image. Content-based hereby refers to the mode of formulating the search query. As opposed to metadata-based image search, which relies on pre-labeling the images beforehand, a CBIR system retrieves the $n$ best matches within the database with respect to the visual similarity to the query image.

In CBIR, the concepts of feature extraction (FE) and similarity measure (SM) play crucial roles. FE converts an image into a feature vector of numerical values with the aim to produce a low-dimensional, yet sensible representation of the input image in the context of some particular application. An SM assigns a numerical value to a pair of two feature vectors. In this work, we assume that the SM is nonnegative and that a smaller SM value indicates a higher similarity and vice versa. A typical CBIR system is depicted in Fig. 1. The images to be extracted are processed via an FE algorithm and the extracted feature vectors (signatures) are stored in a database.

The query image is fed to the same FE algorithm and the output is compared with all the feature vectors in the database by means of the defined SM. As a result, the images with the lowest SM values are returned. Consequently, designing a CBIR system boils down to the construction of an FE/SM framework.

Nowadays, CBIR is employed to deal with big databases and vast amounts of data. Thus, in additional to high retrieval efficiency, execution speed is of great concern. This requires sufficient dimensionality reduction within the FE and a computationally efficient SM.

B. Related Work and Our Contribution

Extracting sensible information from images has been a field of research in digital signal processing for roughly half a century now. Naturally, the choice of the FE method depends on the specific application. For instance, when comparing images of landscapes, pictures taken in a forest will have different color distributions from those taken in a desert, so an FE based on the color properties of the image is a reasonable approach. On the other hand, images of objects are strongly characterized by shapes, so in this case the FE could rely on extracted contours. This work focuses on CBIR for textures. The earliest approaches for texture FE, such as the gray-level co-occurrence matrices (GLCM) and the biologically motivated proposed Tamura features have been presented in the 1970s and are often based on numerical measures for how nearby pixels relate to each other.

More recent approaches include incorporating spatial-frequency representations of the images, such as the Fast Wavelet Transform (FWT), Gabor frames or Curvelet frames. A conventional approach is designing the features based on the energy of the respective transform coefficients, but it can be beneficial to look at their statistical properties...
The likelihood of a set of random realizations \( \{x_1, \ldots, x_N\} \) for a PDF \( p \) is defined as
\[
L(x_1, \ldots, x_N | p) = \prod_{i=1}^{N} p(x_i),
\]
and is a measure for the probability that \( \{x_1, \ldots, x_N\} \) is subjected to \( p \). That is to say, \( L(x_1, \ldots, x_N | p_d) \) can be viewed as a measure of similarity between the query image and the database images. Since the natural logarithm \( \ln \) is a monotone function, this is also true for the log-likelihood \( \ln L \).

Let \( A \) denote the set of the PDFs of a number of database images. The best match for a query \( \{x_1, \ldots, x_N\} \) can thus be determined via
\[
p_d^* = \arg \max_{p_d \in A} \ln L(x_1, \ldots, x_N | p_d)
= \arg \max_{p_d \in A} \sum_{i=1}^{N} \ln p_d(x_i).
\]

Expression (2) is called a Maximum Likelihood (ML) solution. The Kullback-Leibler-Divergence (KLD) between two PDFs \( p \) and \( q \) with the support \( S \) is defined as
\[
D(p||q) = \int_{S} p(x) \ln \frac{p(x)}{q(x)} \, dx.
\]

Suppose that \( \{x_1, \ldots, x_N\} \) are subjected to the PDF \( p_q \). Then it can be derived from the law of large numbers, that minimizing for the KLD between \( p_q \) and \( p_d \) yields the same results as maximizing for the (log-)likelihood for \( \{x_1, \ldots, x_N\} \) for \( p_d \), provided that \( N \) is sufficiently large. Thus, (2) can be reformulated as
\[
p_d^* = \arg \min_{p_d \in A} D(p_q||p_d).
\]

The authors of [9] propose features based on the statistical moments of the coefficients generated by an FWT of a texture image. To this end, each FWT subband is modeled by a two-parameter GGD. The FE consists of obtaining the two GGD parameters for each subband. Since the KLD of two GGDs can be directly computed from the two GGD parameters, for every pair of images, the KLD for each subband can be estimated directly from the feature vectors. Summing up all the subband KLDs for a pair of images yields an SM. The summation is justified by the fact that the KLD of two joint distributions of statistically independent variables is the sum of the KLDs belonging to the respective marginal distributions. The described SM is hence the estimated KLD between joint distributions for all the subbands produced by an image.

In ML based CBIR, the KLD was employed as a measure of probability that a match is correct. However, modeling the subband histograms with a parametrized PDF can be also motivated bilogically. In fact, it has been observed [19] that the gray-value distributions of the filterbank outputs of a texture correspond to perceivable texture features. This gives rise to a variety of similar FE/SM approaches. In particular, replacing the FWT by any other filterbank transform might yield more descriptive features. Furthermore, the KLD could be replaced by any of the vast number [20] of SMs for PDFs. For instance,
the symmetrized KLD between two distributions \( p \) and \( q \) is defined as
\[
D_{\text{sym}}(p||q) = \frac{D(q||p) + D(p||q)}{2},
\]
and has the advantage that the symmetry property
\[
D_{\text{sym}}(p||q) = D_{\text{sym}}(q||p)
\]
always holds.

C. The Scattering transform

1) Definition: Let \( \theta \in L^2(\mathbb{R}^2) \) be a rotationally symmetric window function with low-pass characteristics. Let \( \eta \in \mathbb{R}^2 \setminus \{0\} \) and \( J \in \mathbb{Z} \) be fixed and \( \mathcal{R} \subset SO(2) \) a finite group of rotation matrices. With \( \psi(x) = \theta(x)e^{i\eta^T x} \), we define the wavelet \( \psi_{j,R}(x) \) as
\[
\psi_{j,R}(x) = 4^{-j}\psi(2^{-j}Rx), \quad j \in \{J, J-1, \ldots\}, R \in \mathcal{R}.
\]
Further, let us assume a low-pass and rotationally symmetric scaling function \( \phi \in L^2(\mathbb{R}^2) \) and define
\[
\phi_J(x) = 4^{-J}\phi(2^{-J}x),
\]
such that for the respective Fourier transforms \( \hat{\psi}, \hat{\phi} \),
\[
|\hat{\phi}(2^j\omega)|^2 + J \sum_{j=-\infty}^{\infty} |\hat{\psi}(2^jR\omega)|^2 = 1
\]
holds, for almost all \( \omega \in \mathbb{R}^2 \). Then the set
\[
\mathcal{D} = \{\phi_J(\mathbf{u}-\mathbf{x})\}_{\mathbf{u}\in\mathbb{R}^2} \cup \{\psi_{j,R}(\mathbf{u}-\mathbf{x})\}_{\mathbf{u}\in\mathbb{R}^2, j\in\mathbb{Z} \land j \leq J, R \in \mathcal{R}}
\]
constitutes a Parseval-tight frame \( \bar{\mathcal{P}} \). It will be assumed to span \( L^2(\mathbb{R}^2) \) in the following. Note that for two signals \( f, g \), the equation
\[
(f * g)(\mathbf{u}) = \langle f(\mathbf{x}), \bar{g}(\mathbf{u}-\mathbf{x}) \rangle
\]
holds.

The core of the Windowed Scattering transform (WST) \[15\] is a dyadic wavelet decomposition \( U_{\phi,\psi,J,R}[f; j, R] \) of the input signal \( f \in L^2(\mathbb{R}^2) \) with the complex modulus performed on the band-pass components, defined as
\[
U_{\phi,\psi,J,R}[f; j, R] = \{ |\psi_{j,R} * f|, \quad j \leq J, \\
\phi_{j,R} * f, \quad j > J.
\]
The modulus operation \( \cdot \) traverses some of the energy of the band-pass signals towards lower frequencies. Therefore, \( U_{\phi,\psi,J,R} \) can be applied to the output signals \( |\psi_{j,R} * f| \) again. This yields a tree structure like the one depicted in Fig. 2. Note that Eq. (12) yields an infinite (but countable) number of output signals. However, in practice we deal with bandlimited input signals \( f \). Hence, there exists an integer \( J_l \) with \( J_l < J \), such that Eq. (12) needs to be evaluated for \( j \geq J_l \) only, which corresponds to a tree with finite breadth.

Basically, the idea of the WST is to apply \( U_{\phi,\psi,J,R} \) successively to the input signal and only keep the low-pass signals, i.e. to neglect signals represented by the black nodes in Fig. 2. The WST along the path \( \mathbf{p} = (j_1, R_1), \ldots, (j_m, R_m) \) is a representation consisting solely of low-pass signals, avoiding the error \( \|S_{\phi,\psi,J,R}[f; \mathcal{P}_m] - S_{\phi,\psi,J,R}[f; \mathcal{P}]\|_2 \). Actually reconstructing \( f \) from its WST involves a phase recovery problem. However, it is known [24] that digital realizations of wavelet representations similar to those employed in the WST are uniquely determined by their complex modulus. Hence, there exists an integer \( J_l \) with \( J_l < J \), such that Eq. (12) needs to be evaluated for \( j \geq J_l \) only, which corresponds to a tree with finite breadth.

Since WST is a representation consisting solely of low-pass signals, it is stable with respect to small spatial translations and deformations. Specifically, let \( \tilde{f} \) be a deformed or translated version of \( f \). Under certain mild assumptions about the underlying wavelet frame, it has been proven that the error \( \|S_{\phi,\psi,J,R}[f; \mathcal{P}_m] - S_{\phi,\psi,J,R}[f; \mathcal{P}]\|_2 \) is bounded asymptotically, cf. [15]. Note, that we are given some freedom of choice in the wavelet frame which allows for considerable flexibility in terms of parameters such as frequency selectivity or directionality.

Due to its stability and flexibility, the WST is a convenient signal representation and can be used as the foundation for the FE from images.

3) Normalized WST: In order to reduce redundancy and to increase invariance to distortions, the Normalized WST (NWST) [23] was introduced. Let us define \( \mathcal{P}_0 \) to be the “empty” path corresponding to the layer \( m = 0 \) and \( \tilde{p} \in \mathcal{P}_M \) to be the predecessor of
p ∈ ℙₘ, i.e. p = ((j₁, R₁),..., (jₘ, Rₘ)) implies ū = ((j₁, R₁),..., (jₘ₋₁, Rₘ₋₁)).

The NWST is defined as
\[ S_{φ,ψ,J,R}[f;p] = \begin{cases} φ · f & \text{if } p = p₀, \\ \frac{S_{φ,ψ,J,R}[f;p]}{|S_{φ,ψ,J,R}|} & \text{otherwise.} \end{cases} \quad (18) \]

In words, each subband of the WST is normalized by the respective parent subband. In practice, a small constant is added to the denominator in order to avoid division by zero.

III. STATISTICAL SCATTERING CBIR

A. Subband Modeling

We propose to model the gray-value distributions of the different WST subbands with parametrized PDFs and describe the images in terms of their respective parameters to obtain a complete FE mechanism on top of the WST.

1) Root layer: We refer to the sets of WST subbands belonging to the same path length as layers. The root layer, i.e. the layer with \( m = 0 \), consists solely of the signal
\[ S_{φ,ψ,J,R}[f;p₀] = φ · f. \quad (19) \]

The low-pass filter \( φ · f \) serves as a blurring filter, implying that \( φ · f \) is a low-resolution version of \( f \). Hence, the gray-value distribution of \( φ · f \) is expected to have a shape similar to that of \( f \). Typically, histograms of natural images can be modeled by multimodal PDFs. As a less elaborate alternative which is widely used as a model for bandpass components, cf. \([25]\), we propose to utilize the GGD, i.e.
\[ p_{GGD}(x|α, β) = \frac{β}{2αΓ(1/β)} e^{-(|x|/α)^β}, \quad (20) \]

where \( Γ(t) = \int₀^{∞} x^{t-1} e^{-x} \, dx \), \( t > 0 \), and the parameter \( α \in ℜ^+ \) determines the scale of the distribution while \( β \in ℜ^+ \) describes the shape. Despite the simplicity of the GGD, its two parameters provide effective measures of perceivable visual features such as contrast.

The GGD parameters \( α \) and \( β \) can be estimated as a solution of an ML problem. Detailed derivation of a numerical algorithm for estimating the parameters \((α, β)\) is given in Appendix A.

2) Descendant layer: The layers \( m \geq 1 \) contain signals of the form
\[ S_{φ,ψ,J,R}[f;p] = φ_j · |ψ_{j,R}_m · | · · · |ψ_{j,R}_1 · f | · · · |, \quad p ≠ p₀. \quad (22) \]

Clearly, the modulus eliminates all negative values which is why a symmetric model such as the GGD is not appropriate anymore. The histograms of signals from descendant layers suggest a PDF that occupies only positive values and can describe a skew to the left, which makes the Weibull Distribution (WD) a nearby choice. Fig. 5 exemplarily shows histograms of WST subbands of different textures with a fitted WD curve. Despite variations in skewness and spread, the WD fittings describe the actual histograms fairly well. The PDF of the WD is defined for \( x ≥ 0 \) as
\[ p_{WD}(x|λ, k) = λk · (λx)^{k-1} e^{-(λx)^k}. \quad (23) \]

Analogous to the GGD, \( λ \in ℜ^+ \) is the scale parameter, whereas \( k \in ℜ^+ \) determines the shape.

The above argument is purely empirical. The following discussion aims to justify further the choice of the WD as a model for WST subbands in the layers \( m = 1 \): Like for the layer \( m = 0 \), we neglect the impact of the low-pass filter \( φ · f \). The approximately analytical band-pass filters \( ψ_{j,R} \) produce signals with similar zero-mean distributions in their real and imaginary parts. Moreover, it is known that bandpass components of a natural image can be well modeled by a GGD. It is therefore natural to assume the GGD with the same parameters for the distribution of the real and imaginary parts of the band-pass signals \( ψ_{j,R}_m · | · · · |ψ_{j,R}_1 · f | · · · | \) involved in the WST. By neglecting statistical dependence between real and imaginary components, let us define a complex
In order to determine the PDF for the modulus $x = |z| \geq 0$, we need to fix $|z|$ and integrate $p_{CGGD}$ over the circles centered at 0 in the complex plane:

\[ p_{CGGD, \text{abs}}(x|\alpha, \beta) = \int_{-x}^{x} p_{CGGD}(z|\alpha, \beta)\, dz \]

\[ = \left( \frac{\beta}{\alpha \Gamma(1/\beta)} \right)^2 2 x e^{-\alpha x^2/2} \int_0^{\pi/2} e^{-\beta \sqrt{\alpha^2 - x^2}} \, d\varphi \]  

(25)

Evaluating the integral analytically is demanding, if not impossible. However, it is known that the modulus of a complex variable with statistically independent and Gaussian distributed real and imaginary parts abides the Rayleigh distribution, which can be easily verified by substituting $\alpha = \sqrt{2}\sigma$ and $\beta = 2$ into Eq. (25), i.e.

\[ p_{CGGD, \text{abs}}(x|\sqrt{2}\sigma, 2) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)} \]  

(26)

The usage of the GGD for modeling FWT subbands was motivated by the need to model histograms which are more or less peaked in shape than it is the case for the Gaussian distribution. For this purpose, the additional shape parameter $\beta$ was introduced. This corresponds to introducing an additional shape parameter to the Rayleigh distribution. Altering the peakedness of the real and imaginary part corresponds roughly to altering the skewness of the respective modulus distribution. From Eq. (25), we can observe that $p_{CGGD, \text{abs}}(0|\alpha, \beta) = 0$, so a good alternative to $p_{CGGD, \text{abs}}$ for modeling the WST subbands would be a two-parameter generalization of the Rayleigh distribution which is controllable in its skewness without changing its value at 0. For $k > 1$, these requirements are fulfilled by the WD. Similarly, we employ a numerical approach to estimate the Weibull parameters via minimizing the corresponding ML function. The derivation of the algorithm can be found in Appendix B.

\begin{itemize}
\item \textbf{Scattering Similarity}
\end{itemize}

1) PDF Similarity with parametrized KLD: The extracted GGD and WD parameters constitute a feature vector for each image. Following the approach from [9], we utilize the KLD as defined in [3] as a measure of similarity between two PDFs. The equation (3) can be further simplified when evaluating the KLD for two GGDs, $p(x) = p_{CGGD}(x|\alpha_1, \beta_1)$, $q(x) = p_{CGGD}(x|\alpha_2, \beta_2)$, or two WDs, $p(x) = p_{WD}(x|\alpha_1, \beta_1)$, $q(x) = p_{WD}(x|\alpha_2, \beta_2)$). With the application of some standard integration rules, the KLD for two GGDs amounts to

\[ D_{CGGD}(\alpha_1, \beta_1||\alpha_2, \beta_2) = \ln \frac{\beta_1 \alpha_2 \Gamma(1/\beta_2)}{\beta_2 \alpha_1 \Gamma(1/\beta_1)} - \frac{1}{\beta_1} + \frac{\alpha_1}{\alpha_2} \frac{\Gamma((\beta_2 + 1)/\beta_1)}{\Gamma(1/\beta_1)} \]  

(27)

as indicated in [9]. With $\gamma \approx 0.57722$ being the Euler-Mascheroni constant, it can be derived that the KLD between two WDs is given by

\[ D_{WD}(\lambda_1, k_1||\lambda_2, k_2) = \ln \frac{k_1 \lambda_2^k}{k_2 \lambda_1^k} - 1 + \left( \frac{\lambda_1}{\lambda_2} \right)^k \frac{k_1}{k_1 + 1} + (k_1 - k_2) \left[ \ln \lambda_1 - \gamma \right] \frac{1}{k_1} \]  

(28)

see [26] for details.

2) KLD Sum: An SM for two images $I_1, I_2$ can be defined by summing up all the KLDs for the layer $m = 0$ (GGD) and the remaining layers $1 \leq m \leq M$ (WD). To this end let us assume an indexing $i \in \{0, \ldots, N\}$ of all the subbands of a WST with $i = 0$ denoting the subband at the layer $m = 0$. Let us further denote the probability that the gray value $x_i$ appears at a pixel in the subband $i$ as $p_{1,i}(x_i)$ for image $I_1$ and as $p_{2,i}(x_i)$ for image $I_2$. The joint probabilities that a combination $(x_1, \ldots, x_N)^T$ of gray values appears for the whole WST representation shall be given by $p_1(x_0, \ldots, x_N)$ and $p_2(x_0, \ldots, x_N)$, respectively. Let us assume statistical independence of the WST subbands, i.e.

\[ p_n(x_0, \ldots, x_N) = \prod_{i=0}^{N} p_{n,i}(x_i), \quad n \in \{1, 2\}. \]  

(29)

Then, the expression

\[ D_{WST}(I_1||I_2) = D(p_1||p_2) = D(p_{1,0}||p_{2,0}) + \sum_{i=1}^{N} D(p_{1,i}||p_{2,i}) \]

(30)

\[ = D_{CGGD}(\alpha_1, \beta_1||\alpha_2, \beta_2) + \sum_{i=1}^{N} D_{WD}(\lambda_1, k_1||\lambda_2, k_2) \]

describes the KLD between $p_1$ and $p_2$, where the last equation of (30) expresses the KLD in terms of the GGD and WD parameters of the different subbands, extracted from the images $I_1, I_2$. 

\[ \]
C. Implementing the Framework

The two previous sections provide the theoretical basis for building a complete CBIR based on WST Subband histograms. In order to extract the features, the WST (Section II-C) is performed on each image. Each subband is then subjected to ML to extract the GGD parameters for layer \( m = 0 \) and the WD parameters for all the other layers. If the number of subbands is \( N \), this amounts to feature vectors of size \( 2N \). In order to compare the query feature vector to the feature vectors in the database, the parametrized KLDs are evaluated for each subband via \( (27) \) and \( (28) \). Their sum \( (30) \) yields the SM. As most of these steps can be performed independently of each other, the whole system is well suited to be implemented in a parallel manner, for example in applications for Big Data.

IV. Numerical Experiments

A. Experimental Settings

1) Implementation: All experiments were performed in a Matlab 2014a environment. The code reproducing the key results is available online. In addition to our approach, the FWT + GGD method was implemented according to \( [9] \) with two minor changes in order to further increase performance. First, the low-pass components were included in the FE. Second, the symmetrized version \( (5) \) of the KLD was employed. As a further development of the FWT + GGD method, we implemented the Dual-Tree Complex Wavelet Transform (DT-CWT) + GGD approach which is almost identical to the former, with the exception that instead of FWT subbands, the real parts of the DT-CWT \( (27) \) subbands, obtained via the DT-CWT Transform Pac\( \text{o} \) version 4.3, were fed to the ML estimator. For our approach, the WST was performed by means of the ScatNe\( \text{t} \) toolbox, version 0.2. The Weibull parameters were extracted via \( \text{wblfit} \) from the Statistical Toolbox. Likewise, the similarity was measured via the symmetrized KLD.

2) Choice of parameters: Along with the regular WST, all experiments were also conducted with its normalized version as defined in \( [18] \). The directionality parameter, corresponding to the cardinality of the rotation group \( R \), is fixed to be \( L = 4 \). All computations were performed with double precision and the subbands of the WST are critically sampled according to the Shannon-Nyquist sampling theorem. The bandwidth of the low-pass filter \( \psi_j \) is chosen in a way such that it produces signals of size \( 16 \times 16 = 256 \), in accordance with \( [9] \). The WST is evaluated for the maximum path lengths \( M = 2 \) and \( M = 3 \). In theory, a linear increase in \( M \) leads to an exponential increase in the number of subbands. However, since the modulus operation hardly traverses any energy toward higher frequencies, they become negligible with increasing path length. Apart from the mentioned, default settings of ScatNet were used. In particular, the Morlet Wavelet was used as the band-pass filter \( \psi \) on each layer of the WST. For the FWT + GGD experiments, the 4-tap orthogonal Daubechies filter, cf. \( (27) \), was employed. The DT-CWT transform was run with the options \texttt{antonini} and \texttt{qshift_a}. In both cases, the size of the smallest subband was \( 16 \times 16 \), which corresponds to three levels of decomposition.

3) Data: The database D1 is the same as used in \( [9] \) and is based on the VisTex database. It consists of 40 texture images of the size \( 512 \times 512 \), each divided into 16 non-overlapping \( 128 \times 128 \) patches. This amounts to 640 image samples of 40 different textures depicted in Fig. 4. The database D2 was generated from images of the following subset of the UIUC texture database\( ^3 \) Bark1, Bark2, Wood2, Wood3, Water, Marble, Floor1, Floor2, Pebbles, Wall, Brick1, Glass1, Glass2, Carpet1, Carpet2, Wallpaper, Fur, Knit, Cudoroy, Plaid. Two \( 640 \times 480 \) images from each class are used to create five overlapping \( 256 \times 256 \) patches which are then scaled down to half the edge size. Hence, we get a database of 20 different texture classes each containing ten \( 128 \times 128 \) patches. Fig. 5 depicts one patch from each utilized image. All image patches are normalized to zero mean and unit energy, in order to avoid any bias caused by the overall lighting condition of each original texture image. The set of all patches generated from the same texture is considered a class. Its cardinality will be denoted by \( c \) in the following. Consequently, \( c = 16 \) for D1 and \( c = 10 \) for D2. For each image patch, the \( c−1 \) most similar patches were retrieved. The retrieval rate for each patch is defined as the ratio of the number of retrieved patches from the same class to \( c−1 \). The overall retrieval rate is the average of the retrieval rates for all the images in the database.

B. Results

1) Overview: The retrieval rates are summarized in Table I. As a benchmark, the results are compared to those produced by two other subband histogram methods, FWT + GGD and DT-CWT + GGD. All methods have in common that they first obtain a statistical model for each texture image and then measure the similarity by means of the KLD.

While WST + WD fails in comparison to DT-CWT + GGD on Database D1, NWST + WD is able to outperform all of the competing frameworks by 1.84% for \( M = 2 \) and 2.96% for \( M = 3 \). Database D1 is widely used as a benchmark for CBIR retrieval. In order to provide a sense for the state of the art, Table II summarizes the results from recent publications on comparable approaches: Discrete Wavelet Frames (DWF) + GGD\( ^9 \), DWT + Generalized Gamma Distribution (GTD)\( ^[17] \), Wavelet Domain Hidden Markov Models (WD-HMM)\( ^[16] \) and Rotated Complex Wavelets\( ^[26] \). To the authors’ best knowledge, the best published result for this experiment so far was produced by Rotated Complex Wavelets with a retrieval rate of 82.81% which is still outperformed by 0.53% by NWST + WD with \( M = 3 \).

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Method & DWF+GGD & FWT+GGD & WD-HMM & Rotated Wavelets \\
\hline
Retrieval Rate & 78.12% & 78.40% & 80.05% & 82.81% \\
\hline
\end{tabular}
\caption{Performance of State of The Art Methods on Database D1}
\end{table}

1https://www.ldv.ei.tum.de/uploads/media/texture_retrieval_scattering_14.zip
2http://www-sigproc.eng.cam.ac.uk/Main/NGK
3http://www.di.ens.fr/data/software/scatnet/
4http://vismod.media.mit.edu/vismod/imagery/VisionTexture/distribution.html
5http://www-cvr.ai.uiuc.edu/ponce_grp/data/
Fig. 4. Database D1 from left to right and top to bottom, with row and column in brackets: Bark0 (1,1), Bark6 (1,2), Bark8 (1,3), Bark9 (1,4), Brick1 (1,5), Brick4 (1,6), Brick5 (1,7), Buildings9 (1,8), Fabric0 (2,1), Fabric4 (2,2), Fabric7 (2,3), Fabric9 (2,4), Fabric11 (2,5), Fabric14 (2,6), Fabric15 (2,7), Fabric17 (2,8), Fabric18 (3,1), Flowers5 (3,2), Food0 (3,3), Food5 (3,4), Food 8 (3,5), Grass1 (3,6), Leaves8 (3,7), Leaves10 (3,8), Leaves11 (4,1), Leaves12 (4,2), Leaves16 (4,3), Metal0 (4,4), Metal2 (4,5), Misc2 (4,6), Sand0 (4,7), Stone1 (4,8), Stone4 (5,1), Terrain10 (5,2), Tile1 (5,3), Tile4 (5,4), Tile7 (5,5), Water5 (5,6), Wood1 (5,7), and Wood2 (5,8).

Fig. 5. Database D2 (patches) from left to right and top to bottom: Bark1, Bark2, Wood2, Wood3, Water, Marble, Floor1, Floor2, Pebbles, Wall, Brick1, Glass1, Glass2, Carpet1, Carpet2, Wallpaper, Fur, Knit, Curdoroy, Plaid.
|                | FWT + GGD | DT-CWT + GGD | WST + WD M=2 | WST + WD M=3 | NWST + WD M=2 | NWST + WD M=3 |
|---------------|-----------|--------------|--------------|--------------|--------------|--------------|
| Database D1   | 77.40%    | 80.38%       | 78.93%       | 77.92%       | 82.22%       | 83.34%       |
| Database D2   | 53.28%    | 59.89%       | 67.00%       | 65.72%       | 64.56%       | 63.67%       |

**TABLE I**

**PERFORMANCE OF WST + WD AND NWST + WD IN COMPARISON WITH FWT + GGD AND DT-CWT + GGD ON DATABASES 1 AND 2**

Database D2 is considerably smaller than D1, but involves more variation within the classes, for instance, in terms of camera angle and deformation. Again, the WST greatly improves the retrieval performance. However, this time the regular WST along with shorter path lengths produces the best results with a significant margin of 7.11% in comparison with DT-CWT + GGD. Nevertheless, in both experiments, NWST + WD consistently outperforms comparable approaches not involving Scattering.

2) Impact of Directionality: In general, natural textures contain strongly directional features. Hence, it is often beneficial to incorporate a frame of higher directionality as it is done for the Rotated Wavelets approach, for example. The DT-CWT + GGD method which mostly draws its benefits from increased directional selectivity is consistently and significantly outperformed by NWST + WD by at least 1.84% for D1 and 3.78% for D2, despite the same directionality properties. That is to say, there are reasons to believe that the success of the WST based methods is not exclusively due to increased directionality.

3) Impact of Textural Structure: To further analyze the capabilities of the involved methods, consider Fig. 6. It shows the retrieval rates for each class in Database D1 produced by DT-CWT + GGD as well as by WST + WD and NWST + WD with \(M = 2\). The results produced by FWT + GGD were omitted in order to preserve comparability in terms of directionality characteristics. DT-CWT + GGD performed well in comparison to the WST based techniques for the classes Brick5, Buildings9, Metal0 and Wood2 and badly for Bark9, Fabric4, Fod0 and Leaves10. In general, it appears that DT-CWT + GGD is most suitable to flat and (piecewise) smooth surfaces, while the two methods involving Scattering seem to be preferable for surfaces that are rather uneven, be it
because they are rough or because they are bent or dented. This correlates with the premise that Scattering coefficients are less sensitive to spatial deformations, since on the one hand, images of rough surfaces carry significant fractions of high-frequency components and on the other hand spatial deformation in the image is often caused by locally bending or denting the depicted material. More specifically, NWST + WD performed comparably well for the classes Bark0, Bark6, Flowers5, Terrain10 and regular WST for the classes Bark8, Fabric11, Grass1, Stone4. This suggests that WST + WD works better on fine structures while NWST + WD yields better results for coarse structures.

This assumption is corroborated by the following experiment. Fig. 7 depicts the retrieval rates for the Database D2 after blurring it with a Gaussian filter of different bandwidths. Unsurprisingly, the retrieval performance declines with stronger blurring for all methods. However, it is evident that the retrieval rate for WST + WD declines significantly faster than the other two. This can be interpreted as a consequence of the distortion invariance of normalized Scattering coefficients. More importantly, this confirms that the regular WST method relies on fine features since they are most heavily affected by blurring. On the other hand, images of coarse patterns and flat and piecewise smooth surfaces are less affected by blurring such that the other two methods appear to be relatively robust.

V. CONCLUSION AND OUTLOOK

In this work, we propose a subband histogram FE method based on the statistics of the WST of a texture image. Our derivation and analysis demonstrate how to extract statistical features from a WST representation of a texture image by means of matching a Weibull model via ML and how to measure the similarity of the respective feature vectors via the KLD. The proposed techniques come in handy when it comes to reducing the enormous degree of redundancy introduced by the WST since they represent each subband by as much as two real numbered parameters. The experiments show that the proposed method can outperform comparable algorithms based on filter bank transforms in terms of retrieval accuracy when designed as a CBIR system. Regular WST coefficients seem to work better on fine structures and Normalized WST coefficients on coarse structures.

Since approaches for rotation invariant Scattering representations are already available [16, 17], it appears feasible to extend the presented ideas towards a rotation invariant CBIR framework which would allow for more general problem settings.

The summation (30) of the KLDs was partly motivated by assuming statistical independence of the subbands. However, neither does statistical independence hold for the WST subbands, nor for the subbands of other filterbank transforms, in practice. Accounting for statistical dependence in the model can lead to significant improvements in efficiency [10, 12]. Another approach could be based on recalling that analyzing the subband histograms is also motivated biologically. Thus, weight factors could be introduced to the summands of (30) in order to account for the different significance of each subband.

APPENDIX

A. Estimation of GGD parameters

In this section, we describe an optimization algorithm for estimating the GGD parameters. Recall the log-likelihood of a GGD \( p(x|\alpha, \beta) \) for a set of realizations \( \{x_1, \ldots, x_N\} \) in the following equation

\[
f_{\text{GGD}}(\alpha, \beta) = \ln L_{\text{GGD}}(x_1, \ldots, x_N|\alpha, \beta) = N \left[ \ln \frac{\beta}{2\alpha} - \ln \Gamma \left( \frac{1}{\beta} \right) \right] - \sum_{i=1}^{N} \left( \frac{|x_i|}{\alpha} \right)^{\beta}. \tag{31}\]

Then the parameters \((\alpha^*, \beta^*)\) for approximating the distribution of the realizations can be identified by solving the following maximization problem

\[
(\alpha^*, \beta^*) = \arg\max_{(\alpha, \beta)} f_{\text{GGD}}(\alpha, \beta). \tag{32}\]

In this work, we employ a numerical optimization approach to solve the maximization problem given in Eq. (32). We firstly take the first derivative of \( f_{\text{GGD}} \) as

\[
\frac{\partial f_{\text{GGD}}}{\partial \alpha} = \alpha^{\beta-1} \beta \sum_{i=1}^{N} |x_i|^{\beta} - \frac{N}{\alpha}, \tag{33}\]

and

\[
\frac{\partial f_{\text{GGD}}}{\partial \beta} = \frac{N}{\beta} + N \Gamma'(1/\beta) - \sum_{i=1}^{N} \left( \ln \frac{|x_i|}{\alpha} \right) \left( \frac{|x_i|}{\alpha} \right)^\beta. \tag{34}\]

Setting the right hand side of Eq. (33) equal to 0 uniquely determines

\[
\alpha^* = \sqrt{\frac{1}{N} \sum_{i=1}^{N} |x_i|^\beta}. \tag{35}\]

for \( \beta > 0 \). Let us denote the Digamma function by \( \Psi(z) = \Gamma'(z)/\Gamma(z) \), similarly, setting the right hand side of Eq. (34) to be equal to 0 leads to

\[
1 + \frac{\Psi(1/\beta)}{\beta} - \frac{\beta}{N} \sum_{i=1}^{N} \left( \frac{|x_i|}{\alpha} \right)^\beta \ln \frac{|x_i|}{\alpha} = 0. \tag{36}\]

By substituting Eq. (35) into Eq. (36), we have

\[
1 + \frac{\Psi(1/\beta)}{\beta} - \frac{\beta}{N} \sum_{i=1}^{N} |x_i|^\beta \ln |x_i| + \frac{\beta}{N} \sum_{i=1}^{N} |x_i|^\beta = 0, \tag{37}\]

which characterizes the critical points of the function \( f_{\text{GGD}} \) in terms of \( \beta \) for \( \alpha = \alpha^* \). Iterative numerical algorithms, such as the Newton-Raphson method, can be used to solve the above equation for \( \beta \), cf. [9]. Due to the summand \(-N \ln \Gamma(1/\beta)\), however, we cannot guarantee concavity of \( f_{\text{GGD}} \). Nevertheless, a choice of good initial point \( \beta_0 \), as proposed in [9], [25], can increase the probability of finding the global maximum of the function \( f_{\text{GGD}} \), i.e. via solving the following equation for \( \beta_0 \) via interpolation,

\[
\frac{N \left( \sum_{i=1}^{N} |x_i|^2 \right)^2}{\sum_{i=1}^{N} |x_i|^2} = \frac{(\Gamma(2/\beta_0))^2}{(1/\beta_0) \Gamma(3/\beta_0)}. \tag{38}\]
Once the solution $\beta^*$ is found, it can be substituted into Eq. (35) for $\alpha^*$.

### B. Estimation of Weibull parameters

In this section, we employ a numerical optimization approach to estimate the Weibull parameter. The log-likelihood amounts to

$$f_{WD}(k, \lambda) = \ln L_{WD}(x_1, \ldots, x_N|k, \lambda) = Nk \ln \lambda + N \ln k + (k - 1) \sum_{i=1}^{N} \ln x_i - \sum_{i=1}^{N} (\lambda x_i^k).$$  

(39)

The optimal parameters $(k^*, \lambda^*)$ are identified by solving

$$(k^*, \lambda^*) = \operatorname{argmax}_{(k, \lambda)} f_{WD}(k, \lambda),$$  

(40)

by employing a similar technique as demonstrated in Appendix A. The first derivatives of $f_{WD}$ with respect to $k$ and $\lambda$ are computed as

$$\frac{\partial f_{WD}}{\partial k} = \frac{N}{k} + \sum_{i=1}^{N} \ln x_i + \left( N - \lambda \sum_{i=1}^{N} x_i^k \right) \ln \lambda - \lambda \sum_{i=1}^{N} (\ln x_i) x_i^k,$$

(41)

and

$$\frac{\partial f_{WD}}{\partial \lambda} = k \left( \frac{N}{\lambda} - \lambda^{k-1} \sum_{i=1}^{N} x_i^k \right).$$  

(42)

Setting Eq. (42) to be equal to 0 uniquely determines $\lambda^*$ for $k > 0$. Solving for $\lambda$ yields

$$\lambda^* = \left( \frac{1}{N} \sum_{i=1}^{N} x_i \right)^{-\frac{1}{k}}.$$  

(43)

Substituting 43 in Eq. (41) and equating it with 0 characterize the critical points of $f_{WD}$ in terms of $k$ for $\lambda = \lambda^*$, i.e.

$$\frac{N}{k} + \sum_{i=1}^{N} \ln x_i - \frac{N \sum_{i=1}^{N} (\ln x_i) x_i^k}{\sum_{i=1}^{N} x_i^k} = 0.$$  

(44)

Since the function $f_{WD}$ as defined in Eq. (39) is concave with respect to $k > 0$, a numerical algorithm such as the Newton-Raphson method can be used to obtain $k^*$. Substituting this solution into Eq. (43) yields $\lambda^*$.

### REFERENCES

[1] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” Systems, Man and Cybernetics, IEEE Transactions on, no. 6, pp. 610–621, 1973.

[2] H. Tamura, S. Mori, and T. Yamawaki, “Textural features corresponding to visual perception,” Systems, Man and Cybernetics, IEEE Transactions on, vol. 8, no. 6, pp. 460–473, 1978.

[3] T. Randen and J. Husoy, “Filtering for texture classification: a comparative study,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 21, no. 4, pp. 291–310, Apr 1999.

[4] J. Smith and S.-F. Chang, “Transform features for texture classification and discrimination in large image databases,” in Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference, vol. 3, Nov 1994, pp. 407–411 vol.3.

[5] B. S. Manjunath and W.-Y. Ma, “Texture features for browsing and retrieval of image data,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 18, no. 8, pp. 837–842, 1996.

[6] I. Sumana, M. Islâm, D. Zhang, and G. Lu, “Content based image retrieval using curvelet transform,” in Multimedia Signal Processing, 2008 IEEE 10th Workshop on, Oct 2008, pp. 11–16.

[7] G. Van de Wouwer, P. Scheunders, and D. Van Dyck, “Statistical texture characterization from discrete wavelet representations,” Image Processing, IEEE Transactions on, vol. 8, no. 4, pp. 592–598, 1999.

[8] N. Vasconcelos, “On the efficient evaluation of probabilistic similarity functions for image retrieval,” Information Theory, IEEE Transactions on, vol. 50, no. 7, pp. 1482–1496, July 2004.

[9] M. N. Do and M. Vetterli, “Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance,” IEEE Trans. Image Processing, vol. 11, no. 2, pp. 146–158, 2002.

[10] ——, “Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden markov models,” Multimedia, IEEE Transactions on, vol. 4, no. 4, pp. 517–527, 2002.

[11] D.-Y. Po and M. Do, “Directional multiscale modeling of images using the contourlet transform,” in Statistical Signal Processing, 2003 IEEE Workshop on, Sept 2003, pp. 262–265.

[12] G. Tzagkarakis, B. Beferull-Lozano, and P. Tsakalides, “Rotation-invariant texture retrieval with gaussianized steerable pyramids,” Image Processing, IEEE Transactions on, vol. 15, no. 9, pp. 2702–2718, Sept 2006.

[13] S. Choy and C. Tong, “Statistical wavelet subband characterization based on generalized gamma density and its application in texture retrieval,” Image Processing, IEEE Transactions on, vol. 19, no. 2, pp. 281–289, Feb 2010.

[14] M. Allili, “Wavelet modeling using finite mixtures of generalized gaussian distributions: Application to texture discrimination and retrieval,” Image Processing, IEEE Transactions on, vol. 21, no. 4, pp. 1452–1464, April 2012.

[15] S. Mallat, “Group invariant scattering,” Communications on Pure and Applied Mathematics, vol. 65, no. 10, pp. 1331–1398, 2012.

[16] L. Sifre and S. Mallat, “Combined scattering for rotation invariant texture analysis,” in European Symposium on Artificial Neural Networks, 2012.

[17] ——, “Rotation, scaling and deformation invariant scattering for texture discrimination,” in Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013, pp. 1233–1240.

[18] J. Bruna, “Scattering representations for recognition,” Ph.D. dissertation, CMAP, Ecole Polytechnique, 2012.

[19] J. R. Bergen and M. S. Landy., “Computational modeling of visual texture segregation,” in Computational Models of Visual Processing, ser. A Bradford book, M. Landy and J. Movshon, Eds. MIT Press, 1991.

[20] S.-H. Cha, “Comprehensive survey on distance/similarity measures between probability density functions,” International Journal of Mathematical Models and Methods in Applied Sciences, vol. 1, no. 4, pp. 300–307, 2007.

[21] S. Mallat, A Wavelet Tour of Signal Processing - The Sparse Way, Elsevier Science, 2008, pp. 89–376.

[22] J. Bruna and S. Mallat, “Invariant scattering convolution networks,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 35, no. 8, pp. 1872–1886, 2013.

[23] ——, “Deep scattering spectrum,” IEEE Transactions on Signal Processing, 2013.

[24] I. Wallspurger, A. d’Aspremont, and S. Mallat, “Phase recovery, maxcut and complex semidefinite programming,” Mathematical Programming, pp. 1–35, 2012.

[25] S. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 11, no. 7, pp. 674–693, Jul 1989.

[26] C. Bauckhage, “Computing the Kullback-Leibler Divergence between probability density functions,” in Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013, pp. 1233–1240.