Correspondence

Texture Classification Using Refined Histogram

L. Li, C. S. Tong, *Member, IEEE*, and S. K. Choy

***Abstract*—In this correspondence, we propose a novel, efficient, and effective Refined Histogram (RH) for modeling the wavelet subband detail coefficients and present a new image signature based on the RH model for supervised texture classification. Our RH makes use of a step function with exponentially increasing intervals to model the histogram of detail coefficients, and the concatenation of the RH model parameters for all wavelet subbands forms the so-called RH signature. To justify the usefulness of the RH signature, we discuss and investigate some of its statistical properties. These properties would clarify the sufficiency of the signature to characterize the wavelet subband information. In addition, we shall also present an efficient RH signature extraction algorithm based on the coefficient-counting technique, which helps to speed up the overall classification system performance. We apply the RH signature to texture classification using the well-known databases. Experimental results show that our proposed RH signature in conjunction with the use of symmetrized Kullback–Leibler divergence gives a satisfactory classification performance compared with the current state-of-the-art methods.**

***Index Terms*—Histogram, statistical modeling, texture classification.**

I. Introduction

Texture classification is one of the fundamental problems in computer vision and has a wide variety of potential applications (see, e.g., [33]). Examples include classification of regions in satellite images [8] and detection of defects in industrial surface inspection [30]. In medical image analysis, texture signature is also applied to classification of pulmonary disease [31] and diagnosis of leukemic cells [32]. As a result, it has received a lot of research interest and a number of approaches have been proposed in the past decades. Among these approaches, texture characterization based on wavelet analysis, which provides a multiresolution and orientation representation that is consistent with the human visual system [1], may be the most popular. Essentially, the widely used wavelet signatures include the energy signature [3], [9], [35], [36], histogram-based signature [3], [4], [7], [13], [15], [34], co-occurrence signature [3], [8], hidden Markov model [10], [11], level sets [24], and so on. In addition to wavelet transforms, texture features extracted from the Gabor filters [25]–[28] are the popular texture representations, which are often adopted in the literature. Other widely used signal processing or filtering features such as Bessel K forms [23], spectral histograms [20]–[22], and independent component analysis features [29] were applied successfully to texture classification, synthesis, and segmentation.

Among the wavelet histogram-based signatures, the Generalized Gaussian Density (GGD) signature [3]–[6], [12]–[14], which is induced from a well-established GGD model for the wavelet subband detail coefficients, is widely used in various areas. Many studies showed that the GGD signature together with the use of Kullback–Leibler Divergence (KLD) gives a high classification/retrieval rate in texture classification/retrieval. Moreover, the GGD model uses only two parameters (scale and shape parameters) to characterize the wavelet subband leading to small storage requirement. However, the GGD signature has several disadvantages. First, Meignen and Meignen [14] showed that the GGD signature extracted in the Maximum Likelihood (ML) framework does not exist for some images (usually when the number of detail coefficients is sufficiently small). Even if it exists, a transcendental equation [4], which involves the gamma and digamma functions, needs to be solved numerically to obtain the estimator for each subband. Thus, a considerable amount of time is necessary and numerical problems may occur when the shape estimator is close to zero. Second, as shown in [7], the performance of GGD signature with the use of L_1-metric (or in general L_p-metric) as a similarity measure is poor compared to that of GGD with KLD. But when the KLD is employed, the computational cost to compute the KLD between two GGDs is high that will slow down the overall classification process. Third, it has been shown [15] that there are no sufficient statistics for the GGD parameters since the Fisher–Neyman factorization [19] does not exist. This means the usual GGD estimators (e.g., ML or moment estimators) cannot capture all the information about the model parameters that are contained in the data, and, therefore, the direct use of GGD estimators as the signature is not well justified.

The modeling of wavelet detail subband histograms via the Product Bernoulli Distributions (PBD) [7], [15] has also received a lot of interest. The PBD model makes use of a binary bit representation for wavelet subband histograms and the so-called Bit-plane Probability (BP) signature is constructed based on the model parameters. The PBD has been shown to perform similarly to the GGD in terms of modeling histogram and classification performance. Essentially, the main merits of BP approach are its efficiency for signature extraction and similarity measurement based on Euclidean metric, and the statistical justification of the model parameters for use in image processing applications [15]. However, it has two main problems. First, the interpretation of the BP signature is not clear since each element in the BP signature is the probability of 1-bit occurrence for each bit-plane and so there is no intuitive explanation (or physical meaning) for these "probabilities". Second, the Euclidean metric used for measuring distance between BP signatures and image models is not well justified.

Motivated by the advantages and disadvantages of the GGD and PBD, we aim at combining the key advantages of the GGD and PBD to construct our model. In this correspondence, we present a Refined Histogram (RH) to model the wavelet subband detail coefficients and propose a new image signature, namely, the RH signature, based on the RH model parameters for supervised texture classification. Our proposed RH model is established by using a step function with exponentially increasing intervals to approximate the wavelet detail histogram and the RH signature is constructed by the concatenation of the RH model parameters for all wavelet subbands. Our RH signature can be extracted efficiently by making use of the coefficient-counting technique that involves only multiplication and counting. In particular, we shall investigate some statistical properties of the RH signature and these properties can justify its sufficiency to characterize wavelet subbands. Finally, we study and compare the classification performance of the RH method with some existing approaches on the well-known texture databases.

Manuscript received April 03, 2009; revised December 21, 2009. First published January 26, 2010; current version published April 16, 2010. This work was supported in part by the HKBU's Centre for Mathematical Imaging and Vision and in part by the RGC GRF under Grant HKBU 202108. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Maya R. Gupta.

The authors are with the Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong (e-mail: creatorlarryli@gmail.com; cstong@hkbu.edu.hk; skchoy@math.hkbu.edu.hk).

Digital Object Identifier 10.1109/TIP.2010.2041414

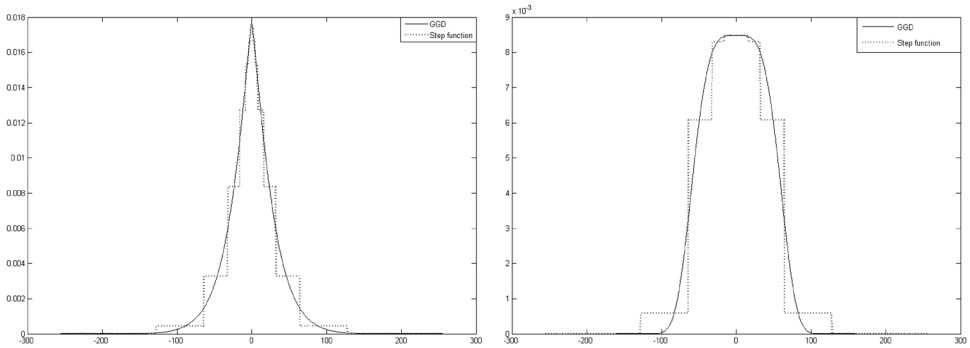


Fig. 1. Approximating the GGD by the step function.

More specifically, this correspondence makes the following contributions.

- We propose a novel RH model that based on the step function for wavelet detail histograms, and present a fast and effective RH signature extraction algorithm based on coefficient-counting technique, which can speed up the classification process.
- We show that the RH signature is the unique minimum variance unbiased estimators of the RH model parameters. This statistical property provides a sound basis for its use in computer vision applications.
- Instead of using the Euclidean metric for measuring the distance between RH signatures, which is not well justified, we use the Symmetrized Kullback–Leibler Divergence (SKLD) as the similarity measure. We reveal that the SKLD between two RH models is exactly the same as the SKLD between two RH features. This means that measuring the distance between images is equivalent to measuring the SKLD between RH model estimators, which justifies the use of SKLD over Euclidean metric.
- We apply the RH signature to supervised texture classification. Comparative experimental results indicate that the proposed RH signature with the use of SKLD gives a satisfactory performance compared with the current state-of-the-art approaches.

The outline of this correspondence is as follows. In the next Section, we introduce our RH model, the signature extraction algorithm, and discuss the statistical properties of the associated RH signature. Experimental results will be presented in Section III, which is followed by our conclusions.

II. Refined Histogram Model

A. Definition

Mallat [5] first noticed an unproven observation that the distributions of the wavelet detail coefficients are peaked at zero and decrease exponentially. Thus, modeling such kind of distribution is of special interest in digital image processing. Based on this observation, the distribution of the wavelet subband detail coefficients can be fitted by the Generalized Gaussian Density (GGD) which is of the form

$$f^{\mathrm{GGD}}(x;\alpha,\beta)=\frac{\beta}{2\alpha\Gamma(1/\beta)}e^{-(|x|/\alpha)^{\beta}},\quad \alpha,\beta>0 \tag{1}$$

where α is the standard deviation of the probability density function and β represents the shape of the GGD model. Although β provides more flexibility to control the model shape, the introduction of this parameter increases the model complexity which makes the parameter estimation (such as Maximum Likelihood [4]) more difficult. This motivates the use of a simple function to replace the GGD model (while preserving the shape of the GGD and its satisfactory classification performance) to approximate the distribution of wavelet detail coefficients.

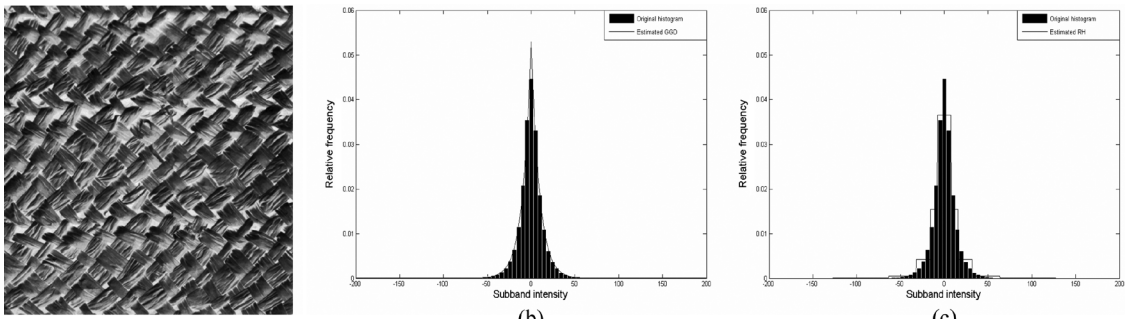
Since step function is one of the simplest functions, it naturally becomes our first choice for modeling the wavelet detail histograms. Intuitively, one good way to achieve this is to utilize the constant functions with adaptive intervals to approximate the GGD functions (see examples in Fig. 1). Because the GGD is an exponential function, intervals with exponentially increasing lengths (i.e.a^N) would be a suitable choice. In order to speed up the feature extraction (see Section II-B) which is one of the most crucial steps in practical classification problems, we choose $a=2$ as the base of the exponent, and thus, the length of the interval increases exponentially with the power of 2. Then the proposed Refined Histogram (RH) for the wavelet detail coefficients, x, with model parameters p_i's is defined as follows:

$$RH(x;p_1,p_2,\ldots,p_n)=\begin{cases}p_i/2^{N-i+1}, & 2^{N-i}\le|x|\le 2^{N-i+1}-1\\ p_n/2^{N-n+2}, & 0\le|x|\le 2^{N-n+1}-1\\ 0, & 2^N\le|x|\end{cases} \tag{2}$$

where N is an arbitrary positive integer, and n (with $0<n\le N$) is the number of quantization levels (or the number of steps for the step function) used to approximate the subband histogram. We remark that the larger the value of n, the better the approximation to the subband histogram. From a geometrical point of view, each p_i is the area of the rectangle with the width $2^{N-i+1}, i=1,2,\ldots,n-1$. Thus, $p_i/2^{N-i+1}$ is the height of the rectangle which is the function value of RH model for the interval $2^{N-i}\le|x|\le 2^{N-i+1}-1$. In addition, from a statistical viewpoint, it is easy to see that p_i is the probability of the coefficients in the corresponding interval, and thus, we have $\sum_{i=1}^{n}p_i=1$.

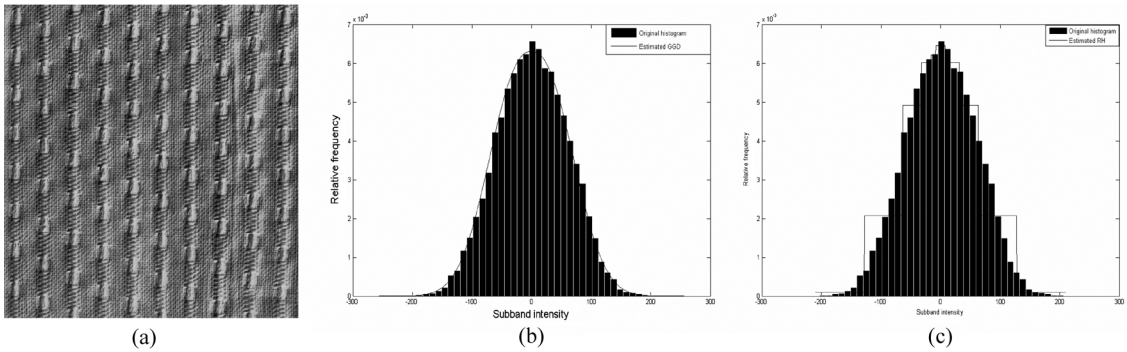
Note that the wavelet detail histogram is peaked at zero and decreases exponentially [5] when move away from the origin (see Figs. 2–3). This implies the variation of the histogram is smaller at its tails (or in some cases at its peak). In such a cases, one may combine some of the consecutive bins of the RH model with similar values into one bin in order to further reduce the model complexity.

Figs. 2 and 3 compare the RH (with $N=n=8$) with GGD (1) for modeling the wavelet detail histograms using two texture images. In Fig. 2, the original histogram is the horizontal detail subband histogram at level 1, whereas in Fig. 3, the original histogram is the vertical detail subband histogram at level 2. The original histograms for these two examples are similar to a Laplacian distribution and a Gaussian distribution. Thus, the GGD model fits well both the observed histograms [with $\beta\approx 1$ in Fig. 2(b) and $\beta\approx 2$ in Fig. 3(b)]. The RH model [Figs. 2(c) and 3(c)] in these cases also gives a satisfactory approximation in the sense that the range and shape information of the original histogram are captured. For histogram-based image classification problems, it is expected that both RH and GGD should perform similarly in terms of classification performance (see Section III-B).



(a) (b) (c)

Fig. 2. (a) Texture. Approximating the original histogram (Laplacian-like distribution) using (b) GGD and (c) RH.



(a) (b) (c)

Fig. 3. (a) Texture. Approximating the original histogram (Gaussian-like distribution) using (b) GGD and (c) RH.

B. Signature Extraction

In this subsection, we present the RH signature extraction algorithm. Here, the RH signature is constructed by concatenating the estimated RH model parameters for all wavelet detail subbands.

After J-level orthogonal wavelet transforms, the original image is decomposed into $3J$ high-pass subbands and one low-pass subband. Because we are only interested in high-pass subbands, we simply refer the high-pass subband to as subband hereafter. Note that the subband histogram is typically symmetrical about zero and decreases exponentially from zero; thus, most wavelet coefficients would probably lie in the range $[-2^N+1, 2^N-1]$ for some N. Hence, given a sufficiently large N, the set of coefficients $x \in [-2^N+1, 2^N-1]$ is good enough to characterize the subband information, and neglecting the coefficients $x \notin [-2^N+1, 2^N-1]$ presumably does not have much influence on subband histogram modeling. In the rest of this correspondence, we shall refer the term "L wavelet coefficients after dropping those coefficients that do not lie on that range" to as "L wavelet coefficients of a subband".

Given a particular subband with L wavelet coefficients $(x_1, x_2, \ldots, x_L)$, there are three approaches to estimate RH parameters. In the first approach, we use the definition of the RH model directly to estimate the parameters $(p_i)_{i=1,2,\ldots,n}$. Let m_i be the number of the absolute value of wavelet coefficients $|x_j|$, $j = 1, 2, \ldots, L$, in the i^{th} interval, then

$$m_i = \begin{cases} \#\left\{|x_j| : 2^{N-i} \le |x_j| \le 2^{N-i+1} - 1 \right. \\ \qquad \left. i = 1, 2, \ldots, n-1\right\} \\ \#\left\{|x_j| : 0 \le |x_j| \le 2^{N-n+1} - 1\right\}. \end{cases} \tag{3}$$

Since p_i is the probability of wavelet coefficients in the i^{th} interval, it follows from the definition of RH model that $\hat{p}_i = m_i/L$.

In the second approach, we estimate the RH parameters via a more efficient binary bit-counting technique, rather than using the definition of RH model directly. Due to the symmetrical property of the subband histograms, estimating the RH parameters from a set of wavelet coefficients $x \in [-2^N+1, 2^N-1]$ is in fact equivalent to estimating the RH parameters from the absolute coefficients $|x| \in [0, 2^N-1]$ (i.e., positive half-plane). The set of absolute coefficients $|x| \in [0, 2^N-1]$ is then quantized into a list of L nonnegative integers that lie in the same range followed by expanding each integer into N binary bit-planes. Given a predefined quantization level, $n (0 < n \le N)$, we start counting the integers from the most significant bit-plane to the least significant bit-plane. We shall count the number of 1's in the first bit of all L integers (call it $\overline{m}_1$). Then those integers whose first bit equals 1 will be "removed" from the list, and thus, only $L - \overline{m}_1$ integers are remained in the list. In general, we count the number of 1's in the i^{th} bit, $\overline{m}_i$, from the remaining $L - \sum_{k=1}^{i-1} \overline{m}_k$ integers, and all integers whose i^{th} bit equals 1 will be removed accordingly. This process continues until $i = n-1$ and finally only $\overline{m}_n$ integers are remained in the list and we have $\overline{m}_n = L - \sum_{k=1}^{n-1} \overline{m}_k$. It is easy to see that $\overline{m}_i$ is exactly the same as the m_i defined in (3). Therefore, we have $\hat{p}_i = m_i/L = \overline{m}_i/L$.

Note that the second approach relies on the bit-counting to estimate the RH model parameters. In many high-level programming languages, it is not always possible to (directly) manipulate the bits in the data structure. Therefore, it would be an important and welcome advantage if we can convert the bit-counting technique to coefficient- or integer-counting technique, and in the meantime, the estimator extracted from both techniques is exactly the same. Our third approach is based on this conversion by slightly modifying the second approach. In each iteration, instead of counting the 1-bit occurrence on each bit-plane, we multiply all the integers by 2 and count the number of the integers which are larger than $2^N - 1$ followed by removing these integers from the list. It is easy to see that such integer-counting is in fact equivalent to bit-counting. As a result, we obtain $\hat{p}_i = \overline{m}_i/L$. It is important to note that the third approach can be applied to quantized or unquantized wavelet coefficients, which provides more flexibility to extract the RH estimators compared with the first and second approaches. In this correspondence, we shall use the third approach for RH signature extraction.

To extract the RH estimators based on the absolute wavelet coefficients $|x| \in [0, 2^N - 1]$, it is necessary to determine N. For typical 8-bit texture images, we find experimentally that, on average, more

than 97.5% of absolute coefficients for any particular subband lie in $[0, 2^8 - 1]$ (see also 8-bit textures in Figs. 2 and 3). This suggests setting $N = 8$ is sufficient for characterizing most subband information, and thus, we shall use $N = 8$ in all the following experiments.

Having introduced the RH model and the model parameter estimation algorithm, we characterize a particular detail subband S_j by the n-D RH feature as

$$RH(S_j) = \left(\hat{p}_1^j, \hat{p}_2^j, \ldots, \hat{p}_n^j\right). \tag{4}$$

We denote the concatenation of two RH features for S_j and S_k using the direct sum operator $\oplus$:

$$RH(S_j) \oplus RH(S_k) = \left(\hat{p}_1^j, \hat{p}_2^j, \ldots, \hat{p}_n^j, \hat{p}_1^k, \hat{p}_2^k, \ldots, \hat{p}_n^k\right).$$

Given an image I with $3J$ high-pass subbands, the image is then characterized by the RH signature defined as follows:

$$RH(I) = RH(S_1) \oplus RH(S_2) \oplus \ldots \oplus RH(S_{3J}). \tag{5}$$

C. Statistical Properties

The main theme of this subsection is to study the statistical properties of the RH model parameters. We shall show that the RH signature always exists and is unique. Moreover it is the minimum variance unbiased estimators of the RH parameters. These properties justify the use of the RH signature in computer vision applications.

Property 1: The RH signature always exists and is unique.

This is obvious from the definition of RH model (2) and from Section II-B that $\hat{p}_i = m_i/L$. Unlike the GGD estimators which may not exist in some cases, the RH signature always exists and is unique, and the calculation does not have any numerical problems.

Property 2: The Maximum Likelihood (ML) estimator of parameter p_i is $\hat{p}_i = m_i/L$.

Given a particular wavelet subband with L independent coefficients $\widetilde{x} = (x_1, x_2, \ldots, x_L)$ and let $\theta = (p_1, p_2, \ldots, p_n)$, we define the log-likelihood function as

$$\begin{aligned} l(\widetilde{x}; \theta) &= \log \prod_{j=1}^{L} RH(x_j; \theta) = \log\left(p_1^{m_1} p_2^{m_2} \cdots p_n^{m_n}\right) \\ &= \sum_{i=1}^{n} m_i \log p_i \end{aligned}$$

with $L = \sum_{i=1}^{n} m_i$. The ML estimator of p_i can be obtained by maximizing the following unconstrained Lagrange function

$$f(\theta, \lambda) = \sum_{i=1}^{n} m_i \log p_i + \lambda\left(\sum_{i=1}^{n} p_i - 1\right)$$

where λ is the Lagrange multiplier. If $(\hat{\theta}, \hat{\lambda})$ is the maximizer of $f(\theta, \lambda)$, the gradients in both sets of variables must vanish. Thus

$$\begin{cases} \frac{\partial f(\theta,\lambda)}{\partial p_i} = 0 \\ \frac{\partial f(\theta,\lambda)}{\partial \lambda} = 0 \end{cases} \Rightarrow \begin{cases} \frac{m_i}{\hat{p}_i} + \hat{\lambda} = 0 \\ \sum_{i=1}^{n} \hat{p}_i = 1. \end{cases}$$

Solving the above equations gives $\hat{p}_i = m_i/L$.

Property 3: The ML estimator $(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n)$ is an unbiased estimator of $(p_1, p_2, \ldots, p_n)$.

Let M_i be a random variable to denote the number of coefficients in the i^{th} interval. Each coefficient has a probability of p_i in the i^{th} interval and we have totally L such independent coefficients. Thus, M_i follows a binomial distribution, $M_i \sim B(L, p_i)$, and we have $E(M_i) = p_i L$, for $i = 1, 2, \ldots, n$. Using *Property 2*, then $E(\hat{p}_i) = p_i$ which proves that the ML estimator $\hat{p}_i$ is an unbiased estimator of p_i.

Property 4: The ML estimator $(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n)$ is a sufficient statistic for $(p_1, p_2, \ldots, p_n)$.

To show this, the likelihood function can be written as

$$l'(\widetilde{x}; \theta) = \prod_{j=1}^{L} RH(x_j; \theta) = \prod_{i=1}^{n} p_i^{m_i} = \prod_{i=1}^{n} p_i^{\hat{p}_i L} = \left(\prod_{i=1}^{n} p_i^{\hat{p}_i}\right)^L.$$

By the Fisher–Neyman factorization theorem [19], the result follows.

Property 5: The ML estimator $(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n)$ is a complete statistic for $(p_1, p_2, \ldots, p_n)$.

Since $M_i \sim B(L, p_i)$ as shown in *Property 3*, the completeness of $(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n)$ directly follows from the completeness of the binomial distribution.

Because of the *Properties 1–5*, by the Lehmann-Scheffé theorem [19], $(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n)$ is the unique minimum variance unbiased estimator of $(p_1, p_2, \ldots, p_n)$. Therefore, the RH signature, which is the concatenation of the RH features for all subbands, should well characterize the detail subband information compared with the GGD signature, which is not a sufficient statistic as shown in [15].

D. Similarity Measurement

After signature extraction, we need to define some distances to compare the discrepancy between two images. Here, we shall use two distances, namely, the L_1-metric and the Symmetrized Kullback–Leibler Divergence (SKLD) [16], for the RH signature in the following experiments.

Given two images I_1 and I_2 with $3J$ detail subbands $\Omega_1 = (S_1^{I_1}, S_2^{I_1}, \ldots, S_{3J}^{I_1})$ and $\Omega_2 = (S_1^{I_2}, S_2^{I_2}, \ldots, S_{3J}^{I_2})$, respectively. The L_1-metric between two images is given by

$$\mathrm{L}_1(I_1, I_2) = \sum_{\substack{p_{I_1} \in RH(\Omega_1) \\ p_{I_2} \in RH(\Omega_2)}} |RH(I_1) - RH(I_2)|$$

where $RH(I)$ is defined by (5), and p_{I_k} is all the RH model estimators for the image I_k, $k = 1, 2$.

For the SKLD, we first define the SKLD between two n-D RH features for subbands S_1 and S_2 as

$$\begin{aligned} SKLD\left(RH(S_1), RH(S_2)\right) &= \sum_{i=1}^{n} p_i^1 \log \frac{p_i^1}{p_i^2} + \sum_{i=1}^{n} p_i^2 \log \frac{p_i^2}{p_i^1} \\ &= \sum_{i=1}^{n} \left(p_i^1 - p_i^2\right) \log \frac{p_i^1}{p_i^2} \end{aligned} \tag{6}$$

where $RH(S)$ is defined by (4). Based on (6), the SKLD between the two images is given by

$$SKLD(I_1, I_2) = \sum_{i=1}^{3J} SKLD\left(RH\left(S_i^{I_1}\right), RH\left(S_i^{I_2}\right)\right) \tag{7}$$

assuming that all detail subbands are statistically independent. It is important to note that the SKLD between two RH models is exactly the same as the SKLD between two RH features [see (6)], i.e., $SKLD(RH_1, RH_2) = SKLD(RH(S_1), RH(S_2))$. Since we have chosen to represent images via the RH models, this provides a nice relationship that measuring the distance between images is equal to measuring the SKLD between RH model parameters, which, in turn, justifies the use of SKLD for RH signature in classification or retrieval applications. Unlike the Product Bernoulli Distributions (PBD model) and the associated Bit-plane Probability (BP) signature as shown in [7] and [15], the distance between two images is computed by

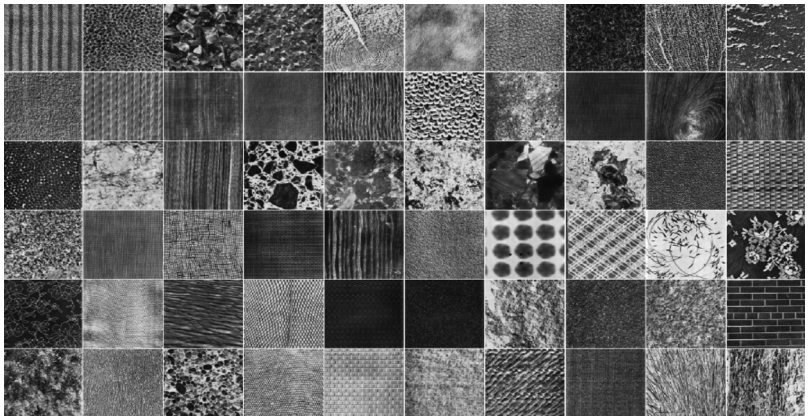


Fig. 4. Sixty Brodatz texture images [17].

comparing directly the distance between two BP signatures based on the L_1-metric (or weighted L_1-metric). However, measuring the L_1-metric between two BP signatures is not equivalent to measuring the distance between the PBD models, and hence, the use of L_1-metric for BP signatures to compute the discrepancy between images is not well justified.

III. Texture Classification

A. Experimental Setting

We apply our proposed RH signature (5) with the use of SKLD (7) for supervised texture classification and refer our method to as RH+SKLD for simplicity. We have used 60 512×512 gray-scale Brodatz [17] texture images (see Fig. 4) for our experiments. As can be seen, there are significant variations for some textures and some of them are very similar to each other. Thus, classification of this dataset is challenging. Each of the 60 texture images is then divided into 16 128×128 nonoverlapping samples creating a database of 960 texture samples. For each sample, 3-level orthogonal wavelet transforms with Daubechies 10 filter bank is used. In the training phase, T (we let T vary from 2 to 8) samples are randomly selected from each of 60 classes for training and the remaining samples are used for testing. We utilize the K-Nearest Neighbor ($K\mathrm{NN}$) with $K = 1$[1] as our default classifier due to its simplicity, efficiency, and effectiveness. Therefore, the input test sample (with its signature) is compared with the training dataset and it will be assigned to the class corresponding to the closest training sample. In particular, we randomly select training samples from the dataset, and repeat the experiment 10 times to obtain the average classification performance.

B. Classification Results

Fig. 5(a) plots the average classification rate of the RH signature (with error bars) with respect to the quantization level ($n = 3, 4, \ldots, 8$) using the L_1-metric and SKLD at 8-training samples case, and Fig. 5(b) plots the corresponding time used for classification. As can be seen, the classification rate of $\mathrm{RH} + \mathrm{L}_1$ and RH+SKLD increases with the quantization level, n, and the RH+SKLD outperforms the $\mathrm{RH} + \mathrm{L}_1$ by 1.0%–5.0%, which suggests SKLD performs better than L_1-metric [but the time used for $\mathrm{RH} + \mathrm{L}_1$ is shorter as L_1-metric is not expensive computationally compared with SKLD, see Fig. 5(b)]. In these two approaches, we see that the difference for the classification rate between $n = 3$ and $n = 8$ cases is 10.0%–15.0%, which implies the quantization level plays an essential role to classification performance. Note that when $n \geq 5$, the improvement for classification rate is insignificant. This is mostly because when n increases, the RH signature can capture more histogram details, and thus, the classification rate increases. However, when it is too large (e.g., $n \geq 6$), we only have more refined details for the peak of the histogram, which may be unnecessary. In such a case, the classification performance will not be further improved significantly. In addition, we remark that the larger the number of quantization levels, the higher the dimensionality of the RH signature. Therefore, the computation time used for classification would be longer [see Fig. 5(b)]. The error bars are also shown in Fig. 5(a) where each error bar is a distance of one standard deviation above and below the average classification rate. The value of standard deviation is about 1.5 for both methods, which means the variation of the classification rates is small and, hence, affirms the robustness of the RH signature.

Next, we compare our proposed $\mathrm{RH}(n = 8) + \mathrm{SKLD}$ with two existing model-based texture classification methods, namely, the Generalized Gaussian Density (GGD) signature with the use of the Kullback–Leibler Divergence (KLD) [4], and the Bit-plane Probability (BP) signature with the use of the L_1-metric [7]. We refer these two methods to as GGD+KLD and $\mathrm{BP} + \mathrm{L}_1$, respectively. To extract the BP signature, each wavelet coefficient is quantized into an integer followed by expanding it into M binary bit-planes. Following [7] and [15], we choose $M = 8$. The BP signature is then obtained by the concatenation of the probabilities of 1-bit occurrence across all bit-planes and subbands. In case of GGD signature, we use the Brent's method [2] to compute the ML estimators [4] of GGD parameters. As in the RH approach, 3-level orthogonal wavelet transforms with Daubechies 10 filter bank is used for BP and GGD signatures extraction, and $K\mathrm{NN}$ classifier with $K = 1$ is applied. For J-level wavelet decomposition with $3J$ detail subbands, the number of parameters for RH, BP, and GGD signatures are $3nJ$, $3MJ$, and $6J$, respectively. Fig. 5(c) shows the average classification rate with respect to the number of samples used in the training phase. One first remarks that our proposed method and GGD+KLD perform similarly (in fact, $\mathrm{RH}(n = 8) + \mathrm{SKLD}$ performs slightly better than GGD+KLD by 1.0%–2.0%), while these two methods outperform $\mathrm{BP} + \mathrm{L}_1$ by 4.0%–6.0%. These results imply that the use of step function for the RH model to approximate subband histogram (and GGD) is successful. The poor performance for the $\mathrm{BP} + \mathrm{L}_1$ is presumably because the PBD model assumes all bit-planes are statistically independent, and the BP signature does not take into account the bit-plane correlations [7], [15]. Also note that in this comparison, we choose $n = M = 8$; thus, the dimensions of RH and BP signatures are exactly the same. However, the RH approach outperforms BP approach by at least 4.0%, which implies RH signature is more expressive than BP signature in terms of characterizing subband information. For the GGD signature, the size of GGD signature is 4 times smaller than that of RH and BP. Although GGD

[1]We also considered the range of K from 2 to 8. The results showed that $K = 1$ gives the best results, and thus, we shall simply report the $K = 1$ case in our experiments.

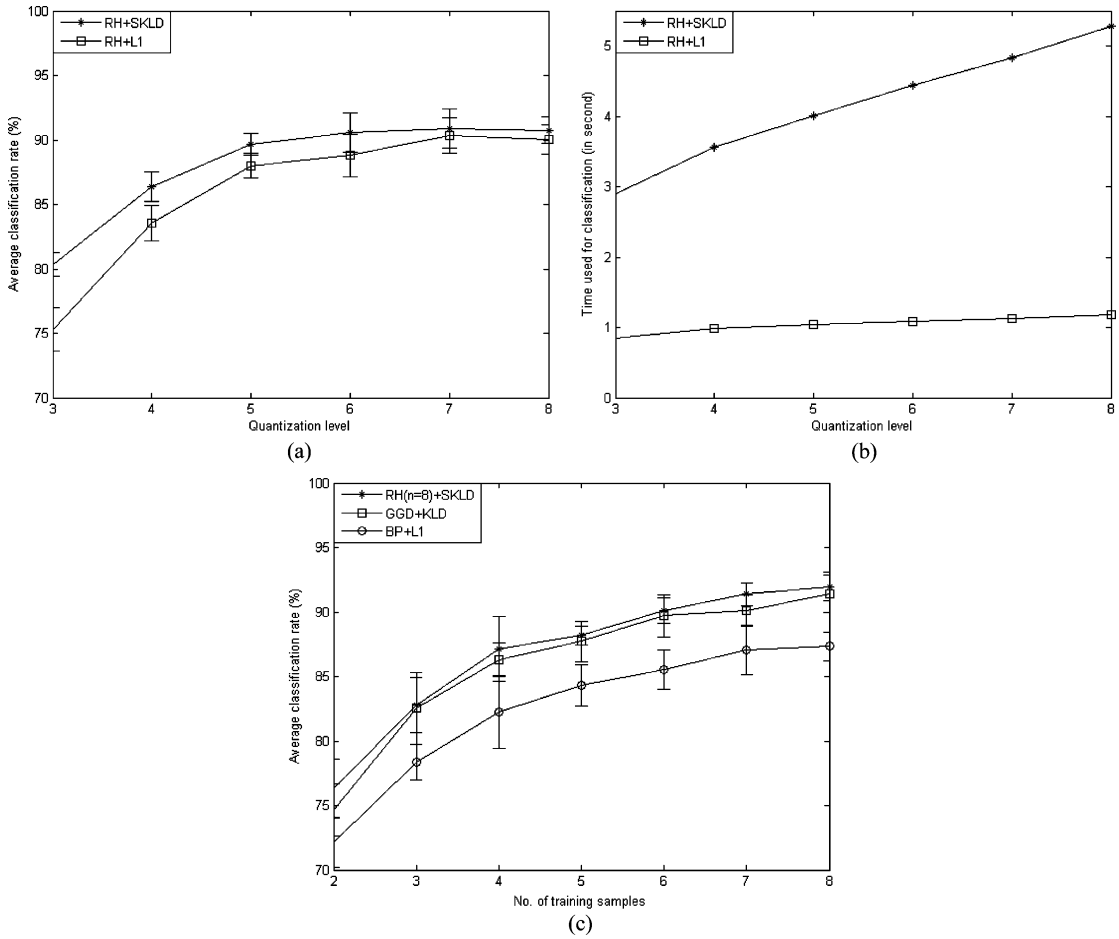


Fig. 5. (a) Classification performance of RH signature using different similarity measures. (b) Time used for classification. (c) Comparative classification performance versus the number of training samples.

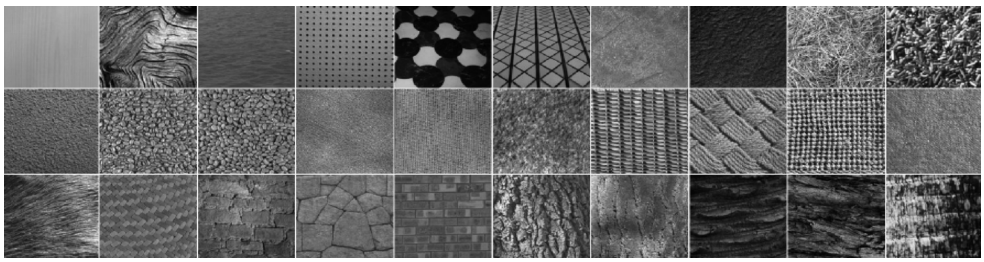


Fig. 6. Vistex texture dataset [18].

and RH approaches perform similarly, the computational cost for feature extraction and similarity measurement for GGD is far higher than that for RH and BP (see Section III-C). If we take into account the computational cost and classification performance, our proposed method is perhaps the best method compared with GGD+KLD and $\mathrm{BP} + \mathrm{L}_1$.

In addition to the 60-texture dataset used in the previous experiment, we shall apply the aforementioned methods to the well-known Vistex dataset [18] of 30 512 × 512 texture images as shown in Fig. 6, which has also been used in [13] and [3]. Each image is divided into 16 128 × 128 nonoverlapping samples, and we have in total 480 samples. The experimental setting used in this experiment is exactly the same as the previous experiment. Fig. 7 compares the performance of all the classification methods using different number of training samples with error bars. We see that the general trend is similar to the previous experiment [see Fig. 5(c)]: our proposed $\mathrm{RH}(n = 8) + \mathrm{SKLD}$ performs slightly better than GGD+KLD by 1.0%–2.0%, and outperforms $\mathrm{BP} + \mathrm{L}_1$ by about 5.0%–8.0%, for any number of training samples. The standard deviation of the error bar is about 1.5–2.0 for $\mathrm{RH}(n = 8) + \mathrm{SKLD}$ and GGD+KLD, while the $\mathrm{BP} + \mathrm{L}_1$ approach has standard deviation of about 3.0.

To provide additional justification of the proposed method, we compare our approach with the spectral histogram method proposed in [21]. The spectral histogram consists of the marginal distributions of filter responses that have been shown to be an effective texture characterization. We have applied our method to the same 40 textures with the same experimental setting as in [21]. Given an integration scale (the size of the input image window), each of the 40 texture images is partitioned into a set of nonoverlapping samples and the dataset is then randomly divided into a separate set of samples for testing and training (half of the samples are used for training and the remaining samples are used for testing). Seven filters including the intensity and Gabor filters (only cosine components are used) with different scales and orientations are automatically selected using a filter selection algorithm and are used to compute the spectral histogram. To measure the similarity between spectral histograms, the authors used a Chi-square statistic.

TABLE I
TIME FOR FEATURE EXTRACTION (TFE FOR WHOLE DATASET) AND SIMILARITY MEASUREMENT (TSM AT 8-TRAINING SAMPLES CASE) FOR 60-TEXTURE DATASET

Method	BP+L_1	GGD+KLD	RH(n=3)+SKLD	RH(n=4)+SKLD
TFE	6.0	105.6	1.5	2.0
TSM	1.2	152.6	2.8	3.5
Method	RH(n=5)+SKLD	RH(n=6)+SKLD	RH(n=7)+SKLD	RH(n=8)+SKLD
TFE	2.4	2.8	3.1	3.3
TSM	4.0	4.4	4.9	5.3

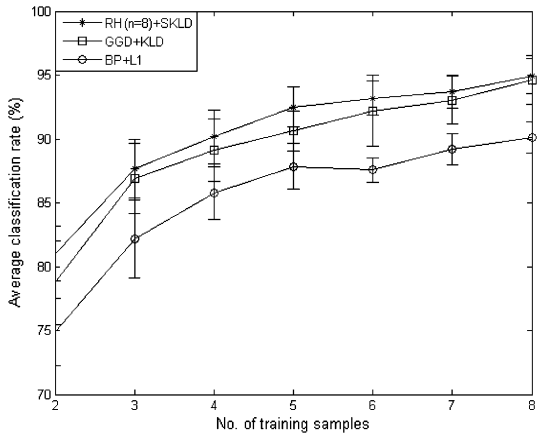


Fig. 7. Comparative classification performance.

The effectiveness of the method is measured by the classification gain, G, which is defined as $G = V(1 - CR)$, where V is the total number of texture classes in the dataset, and CR is the classification error rate. The experiments are repeated 100 times to obtain the average, best, and worst performances. In our experiment, we shall only compare with the average classification performance and the comparison is done on two integration scales: 32×32 and 64×64. The classification gains of our method are 36.8 (corresponding to a classification error of 8.0%) and 37.5 (corresponding to an error of 6.3%) for integration scales 32×32 and 64×64, respectively, while the classification gain for the spectral histogram approach is at least 37.0 (corresponding to an error less than 7.5%) [see Fig. 5(a) of [21]]. This comparison shows that our method is very competitive to the state-of-the-art spectral histogram approach in terms of classification performance.

C. Computational Cost

All the experiments in this correspondence have been implemented on a workstation with double Xeon Quadcore E5320 CPUs (a 2 G RAM and 64-bit version of Fedora 7) in Matlab environment.

Table I reports the Time used for Feature Extraction (TFE for 960 samples) and Similarity Measurement (TSM at 8-training samples case) for GGD, BP, and RH (with quantization level $n = 3, 4, \ldots, 8$) approaches for the 60-texture dataset. For TFE, 3-level wavelet decomposition (as suggested in [14] to increase the likelihood for the existence of GGD signature) with Daubechies 10 filter bank is applied to each 128×128 sample (takes about 23.0 seconds for discrete wavelet transforms for the whole dataset). As can be seen, the TFE for the RH is at least 30 times faster than that for the GGD. This is because estimating the GGD shape parameter in the ML framework needs to solve a transcendental equation [4] (consisting of gamma and digamma functions) using the Brent's method which is computationally expensive, while the extraction of the RH only requires multiplication and counting. Also note that the TFE for the RH is at least 2 times faster than that for the BP. This is mainly depends on the number of quantization levels, n, since the smaller the n, the lower the dimension of the RH, leading to shorter TFE. When the number of quantization levels equals the number of bit-planes used to extract the BP (i.e., $n = 8$ in our case), the dimensions of the BP and RH are exactly the same. In such a case, the TFE for RH is about 2 times faster than that for BP. There are two main reasons: First, for n-D RH feature [see (4)], only $n - 1$ estimators ($\hat{p}_i$'s) are needed to compute as the last estimator can be obtained by $\hat{p}_n = 1 - \sum_{i=1}^{n-1} \hat{p}_i$ (see Sections II-A and II-B), whereas all n estimators must be counted for the BP. In general, for J-level wavelet decomposition with $3J$ high-pass subbands, only $3(n-1)J$ estimators need to be computed for RH whereas $3nJ$ estimators need to be computed for BP. Thus, the BP needs to compute extra $3J$ estimators compared with the RH, leading to longer TFE. Second, in the RH signature extraction algorithm, some coefficients will be removed from the list of detail coefficients after obtaining an estimator in each iteration. Thus, the size of list becomes smaller and counting coefficients becomes faster. However, the BP signature in [7] and [15] is obtained by counting the 1-bit occurrence for each bit-plane across all subbands. This means extracting each element of BP must count all the quantized coefficients in the list, and thus, TFE must be longer. For the TSM, the $\mathrm{BP} + \mathrm{L}_1$ is more efficient than GGD+KLD and RH+SKLD, while RH+SKLD is about 30 times faster than the GGD+KLD. Although the TSM for $\mathrm{BP} + \mathrm{L}_1$ is shorter, the total time (TFE+TSM) used for $\mathrm{BP} + \mathrm{L}_1$ and RH+SKLD are in fact comparable. As far as the TFE, TSM, and classification performance [see Fig. 5(c) and Fig. 7] are concerned, our proposed RH+SKLD outperforms $\mathrm{BP} + \mathrm{L}_1$ and GGD+KLD.

TABLE II
TIME FOR FEATURE EXTRACTION (TFE FOR WHOLE DATASET) AND SIMILARITY MEASUREMENT (TSM AT 8-TRAINING SAMPLES CASE) FOR 30-TEXTURE VISTEX DATASET

Method	BP+L_1	GGD+KLD	RH(n=8)+SKLD
TFE	2.6	35.5	1.5
TSM	0.3	38.0	1.3

The TFE and TSM for the 30-texture Vistex dataset are summarized in Table II. The results are similar to Table I: $\mathrm{BP} + \mathrm{L}_1$ and $\mathrm{RH}(n = 8) + \mathrm{SKLD}$ are comparable in terms of the total time (TFE+TSM), and these two methods are more efficient than GGD+KLD.

Lastly, we point out that the classification performance of the $\mathrm{RH}(n = 5) + \mathrm{SKLD}$ is similar to the GGD+KLD but it is about 38 times much faster. Of course, in general, one can increase the dimension of RH signature to obtain better classification rate, and in fact, the user can choose the RH signature with arbitrary dimension (note that $0 < n \leq N$). This nice property provides a favorable way to implement the whole classification process on the fly. For example, for the on-line feedback classification/retrieval system with dynamic database, the system can first compute the RH signature with quantization level equals 2 and then outputs the result. If better classification/retrieval is necessary, the user can adjust the dimension from two to higher dimension until the better result is achieved. Given a predefined N, extracting high dimension RH signature from low dimension can be done by simply multiplication and counting (see Section II-B). Since the adjustment for extracting the RH signature from low dimension to high dimension can be realized very efficiently, it is expected that the use of RH signature is more flexible than the GGD and BP in many other applications.

IV. Conclusion

This correspondence presents the Refined Histogram (RH) and the associated RH signature, which combines the key advantages of GGD and BP signatures, based on the modeling of wavelet detail histograms. The RH model makes use of the step function with exponentially increasing intervals to approximate the detail histogram and to capture the histogram characteristics, whereas the RH signature can be extracted efficiently using the coefficient-counting technique that involves only multiplication and counting. We study the statistical properties of RH signature and show that it is the unique minimum variance unbiased estimators of the RH model parameters, which provides a sound basis for its use in computer vision applications. In addition, we adopt the Symmetrized Kullback–Leibler Divergence (SKLD) as the similarity measure and reveal that the SKLD between two RH models is exactly the same as the SKLD between two RH features. This is important and useful in the sense that we do not need to estimate the probability mass functions (of RH models) when comparing distances. Finally, we employ the RH signature to supervised texture classification. Experimental results indicate that the RH signature with the use of SKLD has a satisfactory performance, compared with the current state-of-the-art generalized Gaussian density, bit-plane probabilities, and spectral histogram approaches.

While the Refined Histogram achieves promising results in texture classification, the present work can be extended in several directions. For the RH model, the step function with exponentially increasing intervals is adopted, which implies that it performs well only in some special histograms (such as monotonically decreasing subband histograms). We plan to extend the RH to model any random histograms by adjusting the lengths of the intervals so that the improved model can adaptively provide optimal intervals to approximate histograms. Another direction is to apply the RH signature to region-based unsupervised texture/nontexture segmentation problem. This can be done by employing the K-means type algorithm to cluster image regions (which are characterized by the RH signature) with similar texture. To improve segmentation result, boundary localization based on pixel-wise classification technique proposed in [22] can be adopted. Those issues will be explored in our future work.

References

[1] J. Beck, A. Sutter, and R. Ivry, "Spatial frequency channels and perceptual grouping in texture segregation," *Comput. Vis. Graphics, Image Process.*, vol. 37, no. 2, pp. 299–325, Feb. 1987.

[2] R. P. Brent, *Algorithms for Minimization Without Derivatives*. Englewood Cliffs, NJ: Prentice-Hall, 1973.

[3] G. V. de Wouwer, P. Scheunders, and D. V. Dyck, "Statistical texture characterization from discrete wavelet representation," *IEEE Trans. Image Process.*, vol. 8, no. 4, pp. 592–589, Apr. 1999.

[4] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian density and Kullback–Leibler distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.

[5] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[6] B. Aiazzi, L. Alparone, and S. Baronti, "Estimation based on entropy matching for generalized gaussian PDF modeling," *IEEE Signal Process. Lett.*, vol. 6, no. 6, pp. 138–140, Jun. 1999.

[7] M. H. Pi, C. S. Tong, S. K. Choy, and H. Zhang, "A fast and effective model for wavelet subband histograms and its application in texture image retrieval," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3078–3088, Oct. 2006.

[8] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, no. 6, pp. 610–621, Nov. 1973.

[9] T. Chang and C. C. J. Kuo, "Texture analysis and classification using tree-structured wavelet transform," *IEEE Trans. Image Process.*, vol. 2, no. 4, pp. 429–441, Oct. 1993.

[10] M. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using Hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[11] G. Fan and X. G. Xia, "Improved Hidden Markov models in wavelet-domain," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 115–120, Jan. 2001.

[12] M. K. Varanasi and B. Aazhang, "Parametric generalized gaussian density estimation," *J. Acoust. Soc. Amer.*, vol. 86, no. 4, pp. 1404–1415, Oct. 1989.

[13] S. K. Choy and C. S. Tong, "Supervised texture classification using characteristic generalized gaussian density," *J. Math. Imag. Vis.*, vol. 29, no. 1, pp. 35–47, Sep. 2007.

[14] S. Meignen and H. Meignen, "On the modeling of small sample distributions with generalized gaussian density in a maximum likelihood framework," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1647–1652, Jun. 2006.

[15] S. K. Choy and C. S. Tong, "Statistical properties of bit-plane probability model and its application in supervised texture classification," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1399–1405, Aug. 2008.

[16] S. Kullback and R. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

[17] Brodatz Texture Images [Online]. Available: http://www.ux.uis.no/~tranden/brodatz.html

[18] Vistex Texture Images [Online]. Available: http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html

[19] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*, 5th ed. Upper Saddle River, NJ: Prentice-Hall, 1995.

[20] X. Liu and D. L. Wang, "A spectral histogram model for Texton modeling and texture discrimination," *Vis. Res.*, vol. 42, no. 23, pp. 2617–2634, Oct. 2002.

[21] X. Liu and D. L. Wang, "Texture classification using spectral histograms," *IEEE Trans. Image Process.*, vol. 12, no. 6, pp. 661–670, Jun. 2003.

[22] X. Liu and D. L. Wang, "Image and texture segmentation using local spectral histograms," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3066–3077, Oct. 2006.

[23] A. Srivastave, X. Liu, and U. Grenander, "Universal analytical forms for modeling image probabilities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1200–1214, Sep. 2002.

[24] J. F. Aujol, G. Aubert, and L. Blanc-Feraud, "Wavelet-based level set evolution for classification of textured images," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1634–1641, Dec. 2003.

[25] S. E. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on Gabor filters," *IEEE Trans. Image Process.*, vol. 11, no. 10, pp. 1160–1167, Oct. 2002.

[26] J. Han and K. K. Ma, "Rotation-invariant and scale-invariant Gabor features for texture image retrieval," *Image Vis. Comput.*, vol. 25, no. 9, pp. 1474–1481, Sep. 2007.

[27] J. Melendez, M. Garcia, and D. Puig, "Efficient distance-based per-pixel texture classification with Gabor wavelet filters," *Pattern Anal. Appl.*, vol. 11, no. 3–4, pp. 365–372, Sep. 2008.

[28] K. Muneeswaran, L. Ganesan, S. Arumugam, and K. R. Soundar, "Texture image segmentation using combined features from spatial and spectral distribution," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 755–764, May 2006.

[29] X. W. Chen, X. Zeng, and D. V. Alphen, "Multi-class feature selection for texture classification," *Pattern Recognit. Lett.*, vol. 27, no. 14, pp. 1685–1691, Oct. 2006.

[30] F. S. Cohen, Z. Fan, and S. Attali, "Automated inspection of textile fabrics using textural models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 803–808, Aug. 1991.

[31] R. N. Sutton and E. L. Hall, "Texture measures for automatics classification of pulmonary disease," *IEEE Trans. Comput.*, vol. C-21, no. 7, pp. 667–676, Jul. 1972.

[32] H. Harms, U. Gunzer, and H. M. Aus, "Combined local color and texture analysis of stained cells," *Comput. Vis. Graph., Image Process.*, vol. 33, no. 3, pp. 364–376, Mar. 1986.

[33] C. H. Chen, L. F. Pau, and P. S. P. Wang, *Handbook of Pattern Recognition and Computer Vision*, 2nd ed. Singapore: World Scientific, 2000.

[34] S. K. Choy and C. S. Tong, "Statistical wavelet subband characterization based on generalized gamma density and its application in texture retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 281–289, Feb. 2010.

[35] S. K. Choy, C. S. Tong, and Z. Z. Zhao, "A novel and effective multistage classification system for microscopic starch grain images," *Microsc. Res. Tech.*, vol. 73, no. 1, pp. 77–84, Jan. 2010.

[36] C. S. Tong, S. K. Choy, Z. Z. Zhao, Z. T. Liang, and H. Chen, "Identification of starch grains in microscopic images based on granulometric operations," *Microsc. Res. Tech.*, vol. 70, no. 8, pp. 724–732, Aug. 2007.