Wavelet Domain Multi-fractal Analysis for Static and Dynamic Texture Classification

Hui Ji, Xiong Yang, Haibin Ling and Yong Xu

Abstract—In this paper, we propose a new texture descriptor for both static and dynamic textures. The new descriptor is built on the wavelet-based spatial-frequency analysis on two complementary wavelet pyramids: the standard multi-scale one and the so-called *wavelet leader* one. The introduced wavelet pyramids essentially capture the local texture responses in multiple high-pass channels in a multi-scale and multi-orientation fashion, in which there exists a strong power-law relationship for natural images. Such a power-law relationship is characterized by the so-called *multi-fractal analysis*. In addition, two more techniques, scale normalization and multi-orientation image averaging, are introduced to further improve the robustness of the proposed descriptor. Combining these techniques, the proposed descriptor enjoys both high discriminative power and robustness against many environmental changes. We apply the descriptor for classifying both static and dynamic textures. Our method has demonstrated excellent performance in comparison with the state-of-the-art approaches in several public benchmark datasets.

Index Terms—Texture, dynamic texture, wavelet, wavelet leader, multi-fractal analysis.

I. Introduction

Understanding visual textures, either static or dynamic, plays an important role in many computer vision and image processing tasks such as image and scene classification, video understanding, visual retrieval and image-guided diagnosis. Despite decades of research efforts on texture modeling, it remains a challenging problem, partly owing to the geometrical and/or photometric variations in texture patterns caused by environmental changes (e.g. [14], [22], [58]) including non-rigid surface deformation, viewpoint changes, illumination variation, rotation, scaling, occlusion and etc. Thus, a desired texture descriptor should not only capture highly discriminative information but also be robust to environmental changes.

The existing texture analysis methods are performed in either spatial domain, or frequency domain, or both of them. In the spatial domain, texture descriptors are often built on the top of some statistical measures of local texture patterns in terms of textons ([29]). These methods have achieved certain degree of insensitivity to occlusions and cluttering by

sacrificing some holistic texture information. An alternative solution is to seek such holistic information in the frequency domain, which has been shown to be effective on capturing essential characteristics of scene textures in [36]. To use information in both frequency and spatial domains, wavelet-based representations have been proposed by many researchers for texture analysis, retrieval and classification (Sec. II).

In this paper, we propose a novel texture descriptor that combines wavelet-based representation and multi-fractal analysis to gain both strong descriptive power and robustness against environmental changes. In the proposed approach, we first represent texture patterns by using traditional low-pass and high-pass wavelet coefficients, as well as the recently proposed wavelet leaders [48]. Then, instead of directly using these measurements, we apply multi-fractal spectrum (MFS) analysis on these wavelet coefficients to extract robust texture descriptors. Two additional processes: scale normalization and orientation averaging are introduced to further improve the robustness of our approach to scale and orientation changes. Integrating all these ingredients, the proposed texture descriptor, named *wavelet-based MFS* (WMFS), encodes rich descriptive information while enjoying strong robustness against environmental changes. The WMFS is applied to the classifications of both static and dynamic textures. Experimental evaluations on four public static texture datasets and one public dynamic texture dataset (with five different breakdowns) show clearly the effectiveness of the proposed method.

In summary, we make several contributions in this work. First, the novel combination of wavelet, wavelet leaders and MFS makes the proposed WMFS texture descriptor very powerful for texture classification tasks. Second, the introduced scale normalization and rotation averaging further improves the robustness of our approach. Third, the WMFS is applicable to both static and dynamic textures. In the experiments, WMFS performs better than or at least as good as other previously reported texture descriptors. In the rest of the paper, Sec. II reviews related works and overviews the proposed texture descriptor. Sec. III introduces background knowledges of the wavelet, wavelet leader and multi-fractal spectrum. Then, the proposed texture representation, for both static and dynamic, is detailed in Sec. IV. At last, the experimental evaluations are reported in Sec. V and the conclusion is drawn in Sec. VI.

II. Related works and our approach

There has been an abundant literature on texture classification. In the following, we only summarize recent studies that are most relevant to our work.

A. Static Texture Classification

Many modern texture classification systems (e.g. [9], [31], [50], [57], [27], [55]) model texture patterns using the statistics of spatial texton distribution. The basic idea is to extract local patches through robust feature detectors or random sampling. Then these patches are quantized into a texton dictionary and from which some statistical measurements (mostly histogram based) are derived. The main advantage of such local feature based descriptors lies in the robustness of local features to geometric and illumination changes, as well as partial occlusions. One representative work is the texture description proposed by Lazebnik et al. [28] which is based on the histogram of affine-invariant regions. Such a texture descriptor shows strong robustness to in-plane image rotations, scale changes and affine transformations. The promising performance has been shown on the applications of texture classification and retrieval. In Zhang et al. [55], a similar bag-of-words texture representation has been used for texture classification on several benchmark datasets.

Besides the histogram, an alternative statistical tool is the fractal/multi-fractal analysis (e.g. [23], [43], [44], [51], [52], [53]). One main attractive property of fractal/multi-fractal analysis is its ability of capturing the self-similarities [33] of spatial distribution of textons, which is another essential characteristic of texture patterns. Varma and Garg [43] proposed to use dense local fractal features to represent textons. In [51], [53], Xu et al. proposed to use multi-fractal analysis to characterize the spatial distribution of pixels of different types. In their approaches, pixels are first partitioned into different sets using either local density measurements or the SIFT-based orientation templates. Then the fractal dimension for each pixel set is estimated and combined together to form a multi-fractal spectrum that encodes statistical characterizations on how different types of pixels are distributed.

Spectral information of texture images has also been studied in the past for texture analysis, especially after the invention of wavelet transform (e.g. [2], [11], [24], [40], [45], [46], [47], [48]). Do et al. [11] used the marginal distribution of wavelet coefficients for texture retrieval. Arivazhagan et al. [1] used some advanced statistical features extracted from both low- and high-frequency components of discrete wavelet transform (DWT) for texture classification. Arneodo et al. [2] proposed the modulus maxima of a continuous wavelet transform (MMWT) for image analysis. For multi-component texture images including color images and multispectral images, various similarity measures are proposed for wavelet-based statistics of images, e.g., the Kullback-Leibler divergence measure [45] and Rao geodesic distance [46]. Coupled with a Bayesian framework, a joint statistical model is proposed in [24] to utilize the magnitudes of the dual complex wavelet coefficients for texture retrieval. In these approaches, the wavelet coefficients are used as the local features and then the texture retrieval and analysis is done via different statistical measurements or via different distance functions.

One closely related work that combines the wavelet transform and the fractal analysis is the so-called *wavelet leader* method by Wendt et al. [47], [48]. Instead of using the standard wavelet coefficients, they used wavelet leaders defined as the maximum of the neighboring high-pass wavelet coefficients in spatial-scale space. Three measurements are derived from the wavelet leaders: scaling exponents, multi-fractal spectra and Hölder exponents. However, some important texture primitives in these approaches are missing, e.g. low-frequency information. Also, the sensitivity to environmental changes is not well addressed. As a result, its performance is inferior to many state-of-the-art texture descriptors. Therefore, we propose a new wavelet-fractal approach that combines the information of textures in both spatial and spectral domains with the strong robustness against environmental changes. In our approach, a multi-orientation wavelet pyramid is used as the representation of texture images to combine both wavelet coefficients and wavelet leaders, which provides a solid foundation for multi-fractal analysis with rich information. With the additional help from the scale normalization process, the proposed wavelet-based multi-fractal analysis generates texture descriptors that are insensitive to geometrical and photometric changes.

A preliminary conference version of this work appeared in [54]. The main extensions in this journal version include the extension of WMFS to dynamic texture along with related evaluations, the evaluation of WMFS on two more datasets, and some other experiments to validate the use of wavelet leaders and the use of 'averaging' for rotation invariance.

B. Dynamic Texture Classification

Dynamic texture (DT) is defined as video sequences of moving scenes that exhibit certain stochastic stationary properties. Such video sequences are pervasive in real world, e.g., sequences of rivers, waterfall, foliage, smoke, clouds and fire. Compared with static texture, DT is more challenging to analyze, owing to the additional difficulties of characterizing the stochastic dynamic behavior of DT. Traditional DT classification systems (e.g. [6], [7], [12], [41], [42]) often explicitly model the underlying physical process and then distinguish different DTs by the values of the associated model parameters. For example, the *linear dynamical system* (LDS) is used in [42] to characterize DT processes via the analysis on the resulting Stiefel manifold. Then the LDSs of textures are compared using the Martin distance. In Chan and Vasconcelos [7], probabilistic kernels are combined with both the appearance and motion components for classifying DT sequences. Dynamic characteristics of DT are usually measured using normal flow (e.g. [8], [37], [38]). In [37] and [38], normal flow based statistical measurements have been used for DT analysis. Ghanem and Ahuja [18] introduced a Fourier phase based model to capture both motion and appearance information for DT recognition.

The development of a universal physical process for all DTs certainly is a very difficult task. Another type of methods bypass such difficulties by using appearance and structure based discriminative methods for DT classification ([10], [17], [41], [49], [56]). Wildes and Bergen [49] constructed spatiotemporal filters specifically tuned up for local DT structures with a small number of image patterns and motion patterns. Zhao and Pietikäinen [56] proposed a DT descriptor that extends local

binary pattern (LBP) from the 3D volume by computing LBP of a DT sequence in three orthogonal planes. Ravichandran et al. [41] combined both local DT structure analysis and generative models of DT for DT classification. Derpanis and Wildes [10] proposed an approach for DT recognition by matching distributions of spatial-temporal orientation structure. Chan and Vasconcelos [6] used kernel PCA to learn a non-linear kernel dynamic texture and applied it to video classification. In [5], an approach based on the hierarchical EM algorithm is used for clustering DTs. Ghanem and Ahuja [17] combined elementary distances using the maximum margin distance learning (MMDL) for DT classification.

Our solution can be viewed as an appearance-based discriminative approach. The main difference from previous studies lies in that the discriminative information used in our approach is closely associated with stochastic self-similarities existing in a wide range of dynamic processes that generate those dynamic textures. It is noted that the effectiveness of multi-fractal analysis on capturing such self-similarities in dynamic processes has been demonstrated in various dynamic processes in nature (see e.g. [4], [19]).

C. Overview of the Proposed Approach

In this section, we give an overview of the structure of the proposed approach. See Fig. 1 for an illustration of the basic procedure of computing the proposed texture descriptor.

Scale normalization. The first component aims at improving the robustness of the wavelet coefficients to scale changes. Motivated by recent progresses in invariant feature detection (e.g. [16]), we propose to use scale-normalized texture images as the input for wavelet transform. The estimate of the texture scales is derived from the statistics of scale-invariant patches. In our implementation, the Laplacian blob detector [13] is used to collect scales of local patches, followed by a global scale estimation of the whole textured image.

Multi-orientation wavelet and wavelet leaders. Wavelet coefficients are known to encode both low-frequency and high-frequency information of textures in a multi-scale manner. However, some statistical measurements on wavelet coefficients, such as negative moments, can be unstable since a significant percentage of wavelet coefficients of natural images tend to be small. Thus, in addition to traditional wavelet coefficients, a modified version of the so-called *wavelet leaders* [26] technique is included as one additional measurement of texture images. The purpose is for facilitating the robust computation of multi-fractal spectrum of textures that relies on both positive and negative moments. Moreover, to suppress the orientation sensitivity of wavelet transform, we propose to average the wavelet coefficients over multiple oriented instances of texture images. In summary, the above multi-orientation wavelet and wavelet leader pyramids provide two complementary measurement sources that are stable for statistical computation and encode rich information regarding texture patterns.

Multi-fractal analysis. Based on the above representation, a *multi-fractal spectrum* (MFS) is estimated for each individual wavelet domain, including the low-frequency domain, the high-frequency domain and the wavelet leader domain. The

texture descriptor is then defined as the combination of the multi-fractal spectra estimated in all three domains.

Temporal analysis for dynamic texture. Dynamic texture analysis differs from static texture analysis on its additional temporal dimension. Most of the previous systems (see Sec. II-B) study DTs using methods radically different from those for static textures. In our approach, we show that the proposed framework for static texture can be easily extended to analyze dynamic textures with small modifications. We treat a DT sequence as three different sequences, and each of them composes 2D slides along different spatio-temporal axis. The wavelet-based MFS descriptors of these three orthogonal sequences are combined together to form a texture descriptor for the given DT sequence.

III. WAVELETS AND MULTIFRACTAL ANALYSIS

Before presenting the detailed description of our approach, we first give an introduction of two mathematical tools upon which our approach is built.

A. Wavelets and Wavelet Leaders

There has been extensive literature on wavelet and its applications. Interested readers are referred to [32] for more details. Given an image I, the *discrete wavelet transform* (DWT) decomposes I into one low-frequency channel $D_S(I)$ at the coarsest scale and multiple high-frequency channels at multiple scales $W_{o,s}(I)$, $o = 1, 2, 3$, $s = 1, 2, \ldots, S$, where S is the number of scales ($S = 3$ is used in our work) and o indicates the index of three filter orientations, i.e., horizontal, vertical and diagonal. Thus, we have three high-frequency channels ($o = 1, 2, 3$) at each scale s, which encode the discontinuities of the image along horizontal, vertical and diagonal directions in a multi-scale fashion. In our implementation, the 2D tensor product of Daubechies' DB2 wavelet ([32]) is chosen for its nearly anti-symmetry.

To improve the robustness of certain statistical measurements of regular wavelet coefficients, we use the additional wavelet-based measurements, the so-called *wavelet leaders* ([48]) which is first proposed for multi-fractal analysis of images. Following [48], the wavelet leaders are defined as the maximum of all wavelet coefficients in terms of magnitude in the local spatial neighborhood and scale neighborhood at coarser scales. More specifically, for a pixel $\mathbf{x} \in \mathbb{R}^2$ at scale s_0, its corresponding wavelet leader coefficient is defined as

$$L_{s_0}(I; \mathbf{x}) = \max_{1 \le s \le s_0} \max_{1 \le o \le 3} \max_{\mathbf{y} \in \Omega(\mathbf{x})} |W_{o,s}(I; \mathbf{y})|, \qquad (1)$$

where $\Omega(\mathbf{x})$ is the square neighborhood of $\mathbf{x}$, and $W_{o,s}(I; \mathbf{y})$ is the wavelet coefficient at $\mathbf{y}$. Thus, for an input image I, the wavelet and wavelet leader transformations produce a set of two-dimensional outputs denoted as

$$\mathcal{WL}(I) = \{D_S(I), W_{o,s}(I), L_s(I) : 1 \le o \le 3, 1 \le s \le S\}.$$

Fig. 2 illustrates the high-frequency wavelet coefficients and wavelet leaders of a sample texture image. It is seen that a large amount of small wavelet coefficients are removed when converting wavelet coefficients to wavelet leaders, which
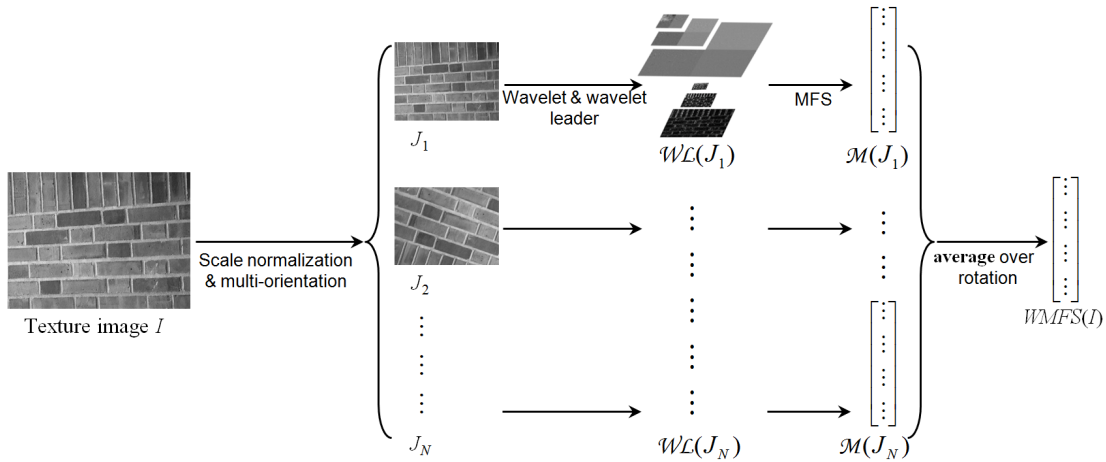
Fig. 1. Flow chart of the proposed wavelet-based multi-fractal spectrum texture representation.
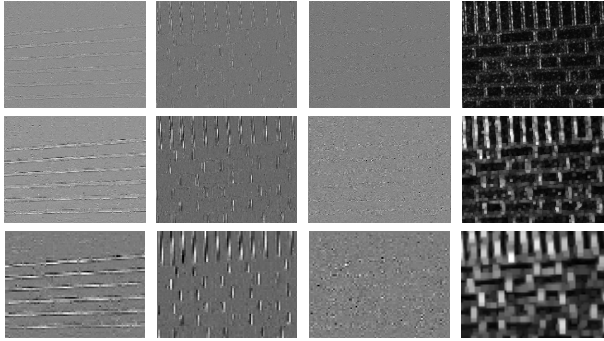

Fig. 2. The high-pass wavelet coefficients and wavelet leaders of a sample textured image Each row represents wavelet coefficients of horizontal, vertical, diagonal directions and wavelet leader respectively.

makes some statistical measurements applicable to the data, e.g. the negative-order moment. It is mathematically justified that the wavelet leaders enable an accurate measurement of the multi-fractal properties of 2D measuring fields [21]. Meanwhile, such a conversion from wavelet coefficients to wavelet leaders does not remove too much information of texture images thanks to the multi-scale maximum nature of the edgels in the wavelet domain (See [48] for more details).

B. Multi-fractal Analysis

Fractal dimensions ([33]) have been widely used in physics and geophysics to capture self similarities in many real world phenomena. Based on the concept of "measurement at scale δ", fractal dimensionality studies the irregularity of a given point set E in the space by measuring its power-law behavior with respect to the scale δ: $m(\delta) \propto \delta^p$, where $m(\delta)$ is some measurement of the given point set E at scale δ. The exponential quantity p is the so-called *fractal dimension* and denoted by $\dim(E)$. One popular definition of the fractal dimension is the *box-counting* fractal dimension ([15]) defined as follows. Given a ℓ-dimensional point set $P = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\} \subset \mathbb{R}^\ell$, let the space $\mathbb{R}^\ell$ be covered by a mesh of ℓ-dimensional

hypercubes with side length r (i.e., r-mesh ℓ-dimensional hypercubes) and a counting function $c(P, r)$ is defined as the number of r-mesh ℓ-dimensional hypercubes that intersect P. Then the *box-counting* fractal dimension $\dim(P)$ is defined as

$$\dim(P) = \lim_{r \to 0} \frac{\log c(P, r)}{-\log r}. \tag{2}$$

In practice, for a point set coming from an image of size $m \times m^1$, to approximate the process of $r \to 0$, we estimate the slope of $\log c(P, r)$ for a side-length sequence, $r_n > r_{n+1} > \cdots > r_m > 0$, $(n < m)$, using the least squares method. In our texture classification context, we define $r_i = \frac{m-i+1}{m}$, $i = n, n+1, \ldots, m$.

As a generalization of fractal dimension, *multi-fractal spectrum* (MFS) analysis is a powerful tool to describe more complex patterns mixed by multiple objects with different fractal dimensions. In the multi-fractal analysis, an image domain Λ (e.g., the $m \times m$ grid points in the above example) is first divided into multiple point sets P_αs according to some categorization term α, i.e., $\Lambda = \bigcup_\alpha P_\alpha$ and $P_{\alpha_1} \bigcap P_{\alpha_2} = \emptyset$, for $\alpha_1 \neq \alpha_2$. The MFS is then given by the multi-fractal function $\dim(P_\alpha)$ vs. α. In the classical definition of the MFS, the categorization term α is defined according to the "density" function, such as the image intensity, to guide the partition of Λ (See [53] for more details).

For a texture image I defined on Λ, given a 2D coefficient matrix $X \in \mathcal{WL}(I)$, d disjoint point sets $P_1, P_2, \ldots, P_d$ are first generated from X using different thresholds $\alpha_0 < \alpha_1 < \ldots < \alpha_d$, i.e., $P_i = \{\mathbf{p} \in \Lambda : \alpha_{i-1} \leq X(\mathbf{p}) < \alpha_i\}$. Then the MFS of X is calculated as

$$\mathrm{MFS}(X) = (\dim(P_1), \dim(P_2), \ldots, \dim(P_d))^\top. \tag{3}$$

For implementation, we use the algorithm introduced in [53] that generates a d-dimensional MFS feature, where $d = 26$, $[\alpha_0, \alpha_d] = [1, 4]$, and the interval $[1, 4]$ is equally divided into 26 subintervals with breakpoints $\{\alpha_i\}, i = 0, 1, 2, \ldots, d$.

It is noted that there are many other types of fractal dimensions defined for the point sets. Our approach is not

---

[1] For simplicity, we assume that the image has equal length in each dimension, otherwise we can normalize it to make so.

Algorithm 1 Multi-orientation wavelet and wavelet leader-based MFS

Input: texture image I

Output: descriptor WMFS(I)

1) Estimating scale of I using statistics from local scale invariant features (Sec. IV-A)
2) $J \leftarrow$ scale normalization of I
3) $\{J_1, J_2, \ldots, J_N\} \leftarrow$ multiple orientations of J
4) Compute the multi-orientation wavelet and wavelet leader pyramids $\mathcal{WL}(J_k)$, $k = 1, \ldots, N$ (Sec. IV-B)
5) Compute the MFSs $\{\mathcal{M}(J_k) : 1 \leq k \leq N\}$ using the box-counting method (Sec. IV-C) and formula (6)
6) WMFS$(I) \leftarrow$ average over $\{\mathcal{M}(J_k)\}$ (Sec. IV-C)
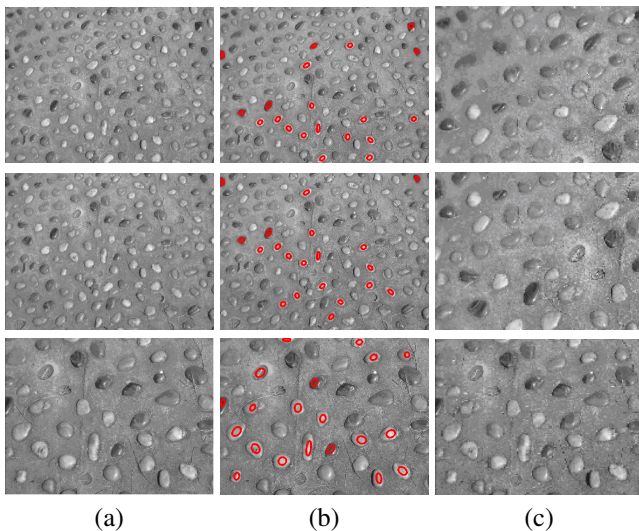

(a)   (b)   (c)

Fig. 3. (a) Three textures in different scales, the top and the middle texture images are nearly in the same scale; (b) images of twenty detected local scale invariant patches; (c) scale-normalized images.

limited to the box-counting fractal dimension. The main reason we choose the box-counting fractal dimension is for its implementation simplicity and computational efficiency.

IV. OUR PROPOSED TEXTURE DESCRIPTOR

Our proposed texture descriptor is built upon the multi-fractal analysis in the wavelet pyramid domain. It can be viewed as the collection of multiple MFS vectors extracted from components of the multi-orientation wavelet pyramids, which include both wavelet coefficients and wavelet leaders of multiple oriented instances of a given texture image. The proposed approach is illustrated in Fig. 1 and outlined in Algorithm 1. In the rest of this section, we will detail each step in the algorithm.

A. Scale Estimation and Normalization

The global scale variation caused by view-point changes is very common in practice. In order to gain robustness to such global scale changes, the feature space upon which the texture descriptor is built should also be robust to such variation. Unfortunately, the plain wavelet pyramid is sensitive

to scale changes. Thus, we propose to first scale-normalize the original texture image before applying the wavelet transform. The basic idea is to estimate the scale of the image using the statistics of local scale-invariant texture patches. Thanks to recent progresses in robust local patch extraction, there are many good local patch detectors with strong scale-invariance [16], [34], [35]. The scale-invariant blob detector used in this paper is the affine-adapted version [35] of the scale-invariant Laplacian-of-Gaussian (LOG) region detector [30]. We use the implementation from Dorko[2] with default setting. See more implementation details in [13].

Inspired by these works, we estimate the global scale of given texture image using the average area of detected invariant elliptic patches. Specifically, for an input texture image I, we first use the local patch detector to extract local elliptic patches and keep the T largest patches, denoted as $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_T\}$. Then scale t of the image is estimated as

$$t = \sqrt{\frac{1}{T} \sum_{i=1..T} \text{area}(\mathbf{p_i})}. \qquad (4)$$

In our implementation, the 20 largest patches are used for each image to estimate its associated scale t, i.e., T is fixed at 20 for all images. We then compare t with a predefined reference scale t_0. Finally, a scale normalized image J is generated by scaling I by the factor of t_0/t. See Fig. 3 for the illustration of the scale normalization. The estimated scales of three given images in Fig. 3 (a) are 1:1:2 respectively. The images are then rescaled accordingly in Fig. 3 (c) and they appear to be on the same scale, which illustrates that the scale-normalization process can effectively remove the scale variations.

B. Multi-orientation Wavelet Pyramid

Another weakness for wavelet transform when used for texture description is its sensitivity to large image rotations. We propose a simple approach to overcome the sensitivity. The basic idea is to use multiple wavelet transforms with different orientation selectivities to encode images under different rotation views, which is equivalent to apply standard 2D tensor wavelet transform on multiple instances of the same image rotated by different angles.

More specifically, given a scale normalized image J, we generate a sequence of images $\{J_1, J_2, \ldots, J_N\}$ for N uniformly sampled orientations. In particular, J_k is generated by rotating J with angle $k\theta$, where θ is the angle unit chosen in the implementation. We then calculate the wavelet and wavelet leader pyramids for each rotated instance:

$$\{\mathcal{WL}(J_k) : 1 \leq k \leq N\}, \qquad (5)$$

where $\mathcal{WL}(J_k)$ is the set of wavelet and wavelet leader coefficient matrices for J_k as defined in Sec. III.

Clearly, the wavelet set $\{\mathcal{WL}(J_k)\}$ defined above is invariant to the rotation up to the tolerance $\theta/2$ when discarding the element order in the set. For the angle interval θ, it can be seen that a too large θ decreases the robustness of the resulting feature while a too small θ decreases the discrimination power and computation efficiency. In our experiments, we use $\theta = \frac{\pi}{8}$ as it achieves a good balance.
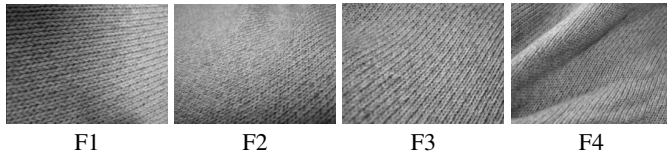
[2]http://lear.inrialpes.fr/people/dorko/downloads.html

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

6

| F1 | F2 | F3 | F4 |

Fig. 4. Four fabric textures with rotation, viewpoint, scale, illumination and non-rigid surface deformation.

C. Wavelet-based MFS

As described in previous sections, given an image I, we compute its multi-orientation wavelet and wavelet leader pyramids $\{\mathcal{WL}(J_k)\}$ from its scale-normalized version for all sampled orientations. The MFS is calculated for each component in $\mathcal{WL}(J_k)$ using the box-counting method, and we define:

$$\mathcal{M}(J_k) = \biguplus_{X \in \mathcal{WL}(J_k)} \text{MFS}(X), \qquad (6)$$

where $\biguplus$ denotes the concatenation of MFS vectors of all components of $\mathcal{WL}(J_k)$. Thus, we get a bag of MFS for the given texture image $I : \{\mathcal{M}(J_1), \mathcal{M}(J_2), \ldots, \mathcal{M}(J_N)\}$. Instead of using the bag of MFSs as the texture descriptor, we construct our wavelet-based MFS (WMFS) by averaging the MFS over its multiple-orientation instances, i.e.

$$\text{WMFS}(I) = \frac{1}{N} \sum_{k=1}^{N} \mathcal{M}(J_k). \qquad (7)$$

Averaging over orientations aims at the robustness of the descriptor to global rotation. Note that other than 'average', there are other statistics that are rotation invariant such as 'maximum'. In our empirical study we found that 'average' outperforms 'maximum' for texture classification, and we therefore choose it for achieving rotation invariance. In summary, for each coefficient channel, we get a 26-dimensional MFS feature vector. Since three scales are used (i.e., $S = 3$) in wavelet decomposition, there are total 13 wavelet coefficient channels, including one low-pass channel and nine high-pass channels and three wavelet leader channels. Thus, the dimension of the texture feature WMFS(I) is 26×13=338.

Four texture images are shown in Fig. 4 from the same class with different viewpoints, different scales and different surface distortions The associated WMFS with/without scale normalization and rotation averaging are shown in Fig. 5. The standard deviation of each feature for each of the 25 randomly selected dimensions is plotted in Fig 6 (a). The variation of features across four texture images of the same class is less when using the scale normalization and orientation averaging, which demonstrates the invariance of the proposed WMFS to scale and orientation. This is further demonstrated in the distance matrices among four features presented in Fig. 6 (b) and (c), where the intra-class distances are reduced when using the scale normalization and rotation averaging.

D. Dynamic Texture Classification

In this section, we extend the proposed WMFS descriptor from 2D static texture to 3D dynamic texture, owing to the existence of the multi-scale self-similarities, fractal structure observed in a large spectrum of dynamic nature images. For example, it is shown in [4], [19] that the amplitudes of temporal frequency spectra of many video sequences, including camera movements, weather and biological evolution indeed obey the power-law in terms of the frequency (the scale of sinusoids).

There are two natural ways to extend the proposed WMFS for DT. One is to treat a DT sequence as a 3D volume and extend the WMFS for 2D image data to handle 3D volume data. This approach, while theoretically sound, faces challenges such as dealing with varying frame rates or motion speed. The second method, which we propose here, is to capture the self-similarity behavior of 2D slices along three orthogonal x, y and t axes in 3D DT volume. Then we calculate WMFS for each 2D slice along three axes. For each axis, the mean of the WMFS of 2D slices is obtained. At last, the WMFS of DT sequence is defined by concatenating the three mean fractal dimension vectors with respect to three axes. Here we only show the calculation of WMFS for DT sequence along t axis, which is shown in Fig. 7. Suppose there are m 2D slices along t axis, denoted as $V_t = \{I_1, I_2, \ldots, I_m\}$, where each I_i is a 2D slice, we take the mean of WMFS of all slices as the descriptor for V_t. Specifically, we have

$$\text{WMFS}(V_t) = \frac{1}{m} \sum_{i=1}^{m} \text{WMFS}(I_i) . \qquad (8)$$

WMFS(V_x) and WMFS(V_y) can be calculated in the same manner, that is, WMFS(V_x) and WMFS(V_y) are the means of WMFS of all slices along x and y axes in DT volume respectively. Define $\text{WMFS}_{DT}(V) = [\text{WMFS}(V_t), \text{WMFS}(V_x), \text{WMFS}(V_y)]$. Then the vector WMFS_{DT} is our proposed descriptor for the DT sequence V. It is noticed that WMFS(V_t) is about the average of the DT structure in spatial domain over time and the temporal information of DT is not used. Since the wavelet coefficients in the high-pass channels of a 2D slice along x- or y-axis are obtained by applying high-pass filters on the 2D slice, these wavelet coefficients can be viewed as the normal flow along y- and x-axis using multi-scale high-pass filters. Thus, The vector WMFS(V_x) and WMFS(V_y) are capturing the variations of the normal flow along y-axis and along x-axis. The performances of WMFS(V_t) and WMFS_{DT} are compared in experimental section; see Table V.

To further improve the stability of the resulting descriptor, the mean of WMFS over all slices is used. The resolution of the frames of the video sequence is usually much lower than that of static image. Thus the slices may not contain enough pixels to yield a stable estimation the MFS. By taking the mean, it effectively suppresses the variations of the computed MFS vectors of all slices. Also, the temporal behavior of DT is characterized by the WMFS of the 2D slices along x and y axes, but without explicitly using the optical flow information. Such an approach has its advantages over those optical flow based approaches since the optical flow field of the sequence of low resolution is hard to be estimated reliably.

Only two scales are used in wavelet decomposition when calculating WMFS for DT (i.e., $S = 2$) due to the limitation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
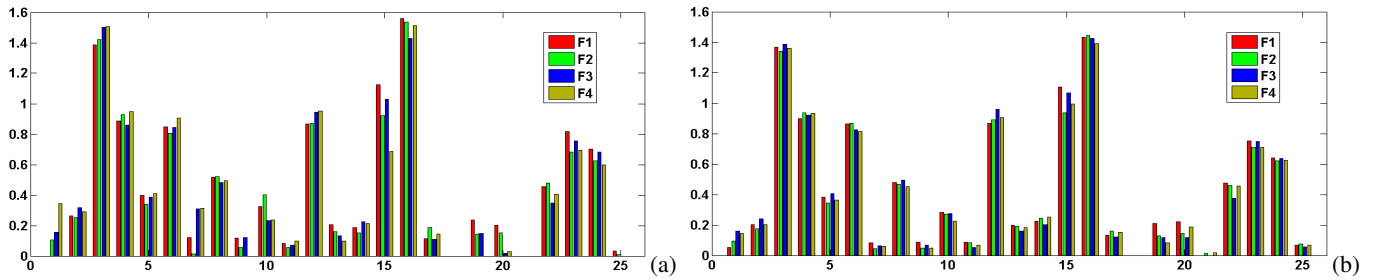
7

Fig. 5. WMFS feature vectors of four texture images shown in Fig. 4, where the x-axis denotes the dimension indices (25 out of 338 are randomly chosen for better illustration) and the y-axis denotes the values of corresponding fractal dimensions. (a) WMFS without scale normalization and rotation averaging; (b) WMFS with scale normalization and rotation averaging.
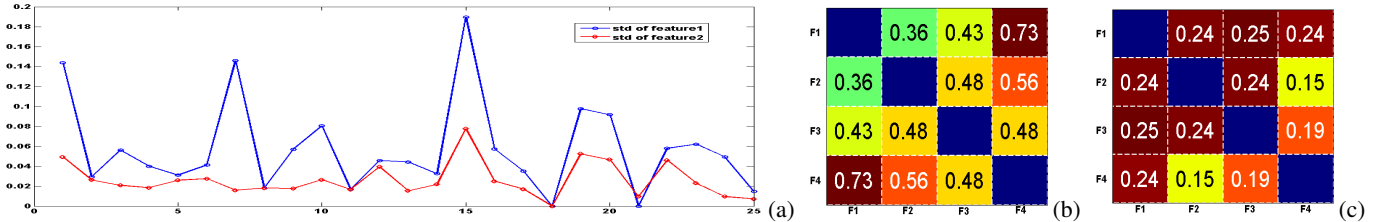

Fig. 6. Comparison of WMFS with and without scale normalization and orientation averaging. (a) The standard deviations of the estimates shown in Fig. 5, where the blue curve stands for the WMFS without scale normalization and orientation averaging, and the red curve stands for the WMFS with them; (b-c) the distance matrices of the four texture images in Fig. 4 using the WMFS *without*(b) and *with*(c) scale normalization and orientation averaging.
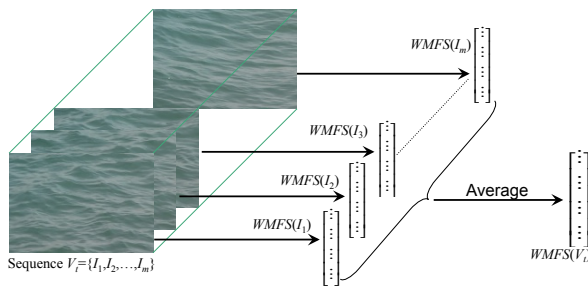

Fig. 7. WMFS for dynamic textures along t-axis.

Fig. 10. Sample images from the ALOT dataset.

of the dimensions of the slices. In summary, there are nine wavelet coefficient channels including one low-pass channel, six high-pass channels and two wavelet leader channels. Thus the dimension of $WMFS_{DT}$ texture feature is $26 \times 9 \times 3 = 702$.

V. EXPERIMENTAL EVALUATION

A. Static Texture Classification

1) Texture Datasets: We evaluate the proposed texture descriptor on three texture datasets of modest size: the UMD texture dataset[3]; the UIUC texture dataset[4]; the KTH-TIPS dataset [20], and one large-size dataset the ALOT dataset[5] Both the UMD dataset and the UIUC dataset consist of 1000 uncalibrated, unregistered images: 40 samples for each of 25 different textures. The image resolution of the UMD dataset is 1280×900 and the image resolution of the UIUC dataset is

---

640×480. Significant viewpoint changes and scale differences present in both datasets. The sample images from these two datasets can be found in Fig. 8. The KTH-TIPS texture dataset [20] consists of 810 images: 81 samples for each of 10 different classes with a low resolution around 200×200. The dataset has significant viewpoint, illumination and small scale changes. The sample images of the KTH-TIPS are shown in Fig. 9. The ALOT dataset is a large-scale dataset, which consists of 25000 color images: 100 samples for each of 250 different textures with a resolution of 1536×1024 pixels. To speed up the computation, we only use the grayscale version of the images downsampled by half to compute the WMFS descriptor during the classification. Some sample images from the dataset are shown in Fig. 10.

2) Experimental Setup: In our classification experiments, the training set is selected as a fixed size random subset of the class, and all remaining images are used as the test set. The reported classification rates are the average over 200 random subsets. We use the support vector machine (SVM) implemented by Pontil et al. [39] as the classifier with RBF kernels. The 1-vs-all classification strategy is adopted in our implementation. The cost factor of the SVM is the

---

[3]http://www.cfar.umd.edu/~fer/High-resolution-data-base/hr_database.htm
[4]http://www-cvr.ai.uiuc.edu/ponce_grp/data/index.html
[5]http://www.science.uva.nl/~aloi/public_alot

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

8

C1: 96.5	C2: 95.8	C3: 99.9	C4: 97.7	C5: 99.9
C6: 99.8	C7: 100	C8: 94.7	C9: 99.8	C10: 99.9
C11: 99.7	C12: 100	C13: 100	C14: 100	C15: 98.4
C16: 100	C17: 98.5	C18: 98.7	C19: 100	C20: 96.4
C21: 97.5	C22: 97.4	C23: 97.5	C24: 99.6	C25: 99.5

C1: 97.0	C2: 95.5	C3: 91.0	C4: 97.8	C5: 96.8
C6: 95.3	C7: 100	C8: 94.0	C9: 97.8	C10: 99.3
C11: 99.5	C12: 98.5	C13: 98.3	C14: 95.0	C15: 100
C16: 100	C17: 100	C18: 100	C19: 88.5	C20: 99.5
C21: 97.8	C22: 99.8	C23: 99.8	C24: 99.8	C25: 100

Fig. 8. Classification rates (%) by the WMFS descriptor on the UMD dataset (left) and the UIUC dataset (right).

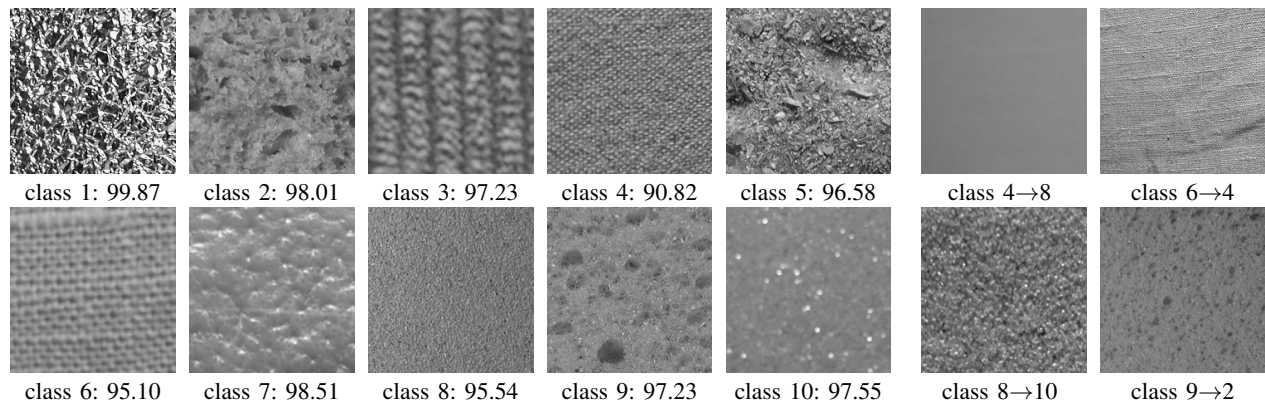| class 1: 99.87	class 2: 98.01	class 3: 97.23	class 4: 90.82	class 5: 96.58	class 4→8	class 6→4
class 6: 95.10	class 7: 98.51	class 8: 95.54	class 9: 97.23	class 10: 97.55	class 8→10	class 9→2

Fig. 9. Sample images in the KTH-TIPS texture dataset. The first five columns show samples and the corresponding classification rates (%) from different classes by our WMFS method; while the remained columns (images from different classes) show failed examples. Class labels before and after "→" indicate true and classified class labels respectively. For example, "4→8" indicates an image from class 4 is misclassified as in class 8. The same notation is used in other figures as well.

number of images in the dataset, and the the range of the RBF kernel parameter is $[0.001, 0.05]$. The parameters used in the SVM are determined by the standard cross-validation. The parameters of the proposed texture descriptor are set as follows. The number of orientations is 16; the wavelet is 'DB2'; the MFS parameters are the same as in [53]. Three scales ($S = 3$) are used for the UMD dataset, the UIUC dataset and the ALOT dataset; and only two scales ($S = 2$) for the KTH-TIPS dataset due to its low resolution.

Due to the implementation complexity of many existing methods and required parameter tune-up for optimal performance, we compare our methods against those methods which either have the code available online, or have classification results reported on the tested datasets. For the UMD and UIUC datasets, 20 samples for each class are used for training and the rest are used for testing. We compared our method (WMFS) against four existing methods:

- The first one, *(H+L)(S+R)*, proposed in [28], uses a patch-based approach by clustering elliptic regions followed by a normalization to circles. Two types of descriptors, spin image and RIFT (Rotation Invariant Feature Transform), are then defined on each region. The resulting descriptor is the histogram of clusters of these local descriptors, which is compared by earth mover's distance.
- The second method, *VG-fractal*, introduced in [43], uses the local density function of various image measurements to produce a 13-dimensional local descriptor. The descriptor is the histogram of clusters of these local descriptors.
- The third method, *MFS* is based on the multi-fractal analysis framework by [53]. The pixel classification is
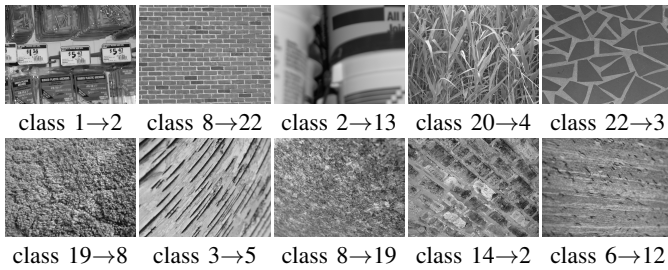
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

9

class 1→2 class 8→22 class 2→13 class 20→4 class 22→3

class 19→8 class 3→5 class 8→19 class 14→2 class 6→12

Fig. 11. Examples of the UMD dataset (first row) and the UIUC dataset (second row) on which the WMFS fails.

TABLE I
CLASSIFICATION RATES (%) ON THE UMD AND THE UIUC DATASETS.

Dataset	VG-fractal [43]	MFS [53]	(H+L)(S+R) [28]	OTF [51]	WMFS w/o leader	WMFS
UMD	96.36	93.93	96.95	98.49	97.60	98.68
UIUC	92.31	92.74	97.02	97.40	95.45	97.62

TABLE II
CLASSIFICATION RATES (%) ON THE KTH-TIPS DATASET

Method	Hayman et al.[20]	(H+L)(S+R) [28]	Zhang et al.[55]	WMFS
Rate (%)	94.8 ± 1.2	91.3 ± 1.4	96.1 ± 1.2	96.54 ± 1.07

using the local density function at each point. Several measurements are used to define multiple local density functions. The descriptor is obtained by combining MFSs defined on these local density functions.

- The fourth method, *OTF*, is an extension of the above MFS method [51] by using multi-scale local orientation histogram for pixel classification.

On the KTH-TIPS dataset, we compare our method with three texture classification methods from [20], [28], [55] respectively. The method in [20] combines filter bank responses of images and a SVM for classification, and the method in [55] uses a bag-of-visual-words framework with SVM. We follow the experiment setup in [55] and randomly divide the dataset into a training set and a testing set. Then we report the average result over 100 runs. On the ALOT dataset, we use the same parameters as those for the UMD and UIUC datasets, except that the step of the scale normalization is omitted as there is little change in the scales of the images in the ALOT dataset.

Our proposed approach is implemented in MATLAB 2007B and the experiments are done on a computer with a 2 GHz Intel Pentium Dual CPU and 2G memory. The running time on the UIUC dataset is as follows. The running time of the proposed feature extraction is about 37 seconds per image. After feature extraction, the features are input to a SVM classifier. Our proposed approach does not require expensive clustering. Thus, the running time for each trial is about 0.2 seconds with 25 classes. During each trial, we used 20 images per class for training, and other 20 images per class for testing.

3) Experimental Results: Tables I summarizes the classification rates of tested methods on the UMD dataset and the UIUC dataset. Fig. 12 (a) and (b) show the classification rates vs. the class indices on both datasets respectively. The classification rates for individual class are shown in Fig. 8. Overall, the proposed WMFS performs slightly better than current state-of-the-art methods on these two datasets. Some texture images on which our method failed are shown in Fig. 11. A more close inspection shows that the WMFS handles the scale and orientation changes better than the others such as the OTF method. But it performs slightly worse on those textures with significant underlying surfaces with severe distorted geometric features than the OTF method, for example the straight edges are distorted to curvy edges. Such sensitivity comes from the fact that the local filter based wavelet measurements are less robust to such geometric changes than the the texton pattern based OTF method. The performances of the

WMFS descriptor computed without wavelet leaders shown in Fig. 12 for the UMD and UIUC datasets illustrate that the inclusion of wavelet leaders indeed improves the performance of the WMFS descriptor on the texture classification. Table II summarizes the classification rates of different methods on the KTH-TIPS dataset, where our method again achieves the best classification performance. Examples of correctly and incorrectly classified images are shown in Fig. 9. Table III shows the classification rates of our method and two other methods on the ALOT dataset. It shows that our method still performs reasonably well on the large size dataset. We also conducted the t-test analysis on the recognition results on the UMD and UIUC datasets to evaluate the statistical significance of the performance differences between the proposed one and other methods. The t-statistic values of the results by the methods for comparison are summarized in Table IV. It is seen that, on both datasets, the WMFS noticeably outperforms the VG-Fractal method and the MFS method, but its improvements over the (H+L)(S+R) method and the OTF method are marginal.

B. Dynamic Texture Classification

We test the proposed texture descriptor on the UCLA dynamic texture dataset [42], which has been widely used for evaluating dynamic texture analysis systems. The UCLA DT dataset contains DT sequences from 50 classes in grayscale. Each class has four grayscale DT sequences and each sequence includes 75 frames with 160×110 pixels. To add more challenges and reduce some ambiguity, the dataset is reorganized

TABLE III
CLASSIFICATION RATE (%) VS TRAINING SIZE ON THE ALOT DATASET

Training size	10	20	30	40	50
WMFS	82.95	89.33	93.57	95.98	96.94
OTF	81.04	89.71	93.45	94.89	95.60
MFS	71.35	78.89	82.57	84.46	85.64

TABLE IV
T-STATISTIC VALUES OF THE RESULTS ON THE UMD AND THE UIUC DATASETS

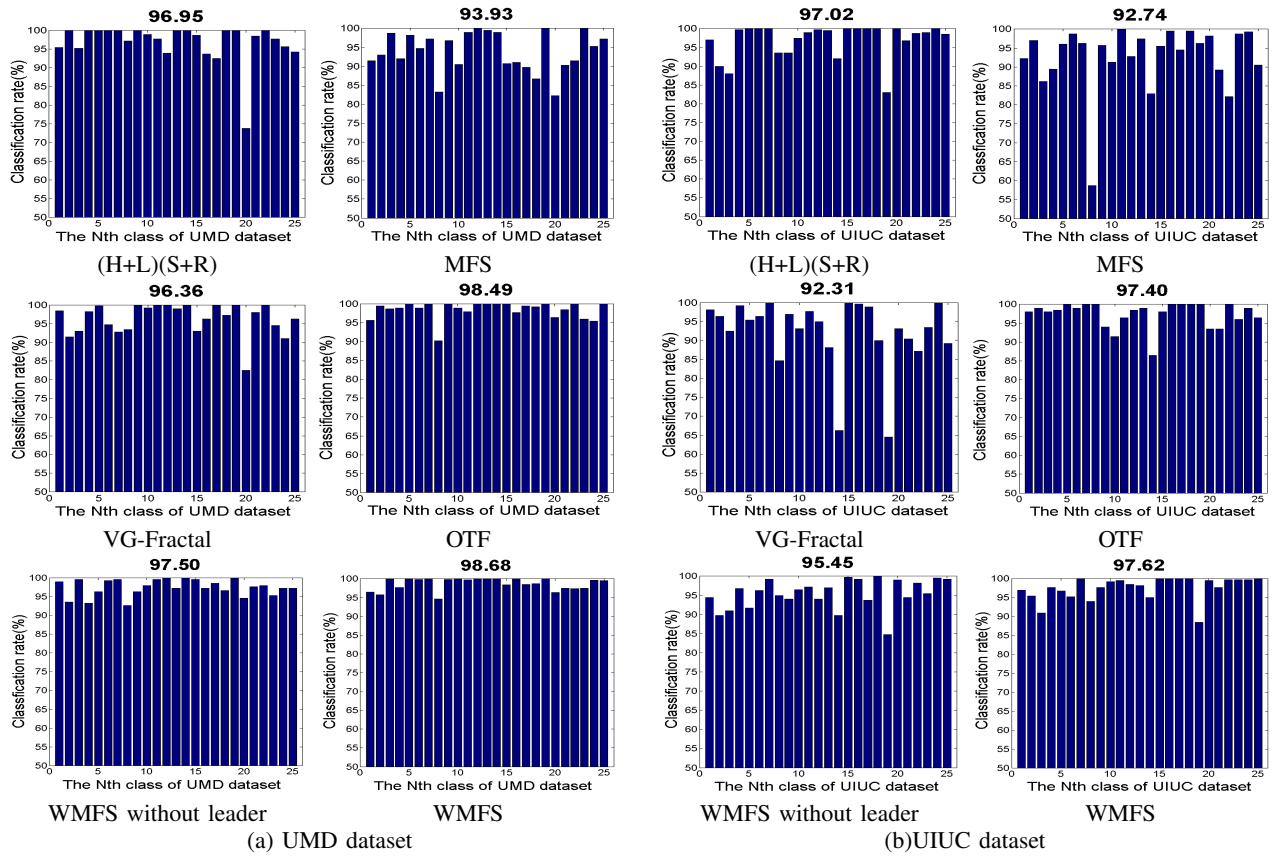| Dataset	WMFS vs. VG-fractal	WMFS vs. MFS	WMFS vs. (H+L)(S+R)	WMFS vs. OTF
UMD	3.00	5.51	1.65	0.53
UIUC	3.42	2.97	1.28	0.26

Fig. 12. Classification rate vs. class index on the UMD and the UIUC datasets. The number on the top of each sub-figure is the average rate.

Fig. 13. Sample key frames in the UCLA DT-8 dataset.

into five different breakdowns in [5], [6], [10], [17], [41], termed DT-50, DT-SIR, DT-9, DT-8 and DT-7 respectively. Some sample frames from all five cases are shown in Fig. 13, Fig. 15 and Fig. 16. Using both NN (nearest-neighbor) and SVM classifiers, we conduct experiments on five different breakdowns for evaluation, in comparison with state-of-the-art methods [5], [6], [10], [17], [41]. For the implementation of the $WMFS_{DT}$, the step of scale estimation and normalization is omitted for computational efficiency as the scale changes are small in most DTs. The other parameters are set as the same as those for the static textures, except that two scales are used (i.e. S=2) since image resolution in DT is much lower than that of static texture image. Table V summarizes the classification rates of the proposed method, together with results reported in the previous works.

1) DT-50 and the Results: The original 50 classes of DT sequences in UCLA DT dataset are used for the experiments of DT classification. The state-of-the-art recognition rates are, 99.00% using NN [17], and 97.50% using SVM [6]. We follow the same setup as [17] and [6]. Specifically, the training set

TABLE V
CLASSIFICATION RATES (%) ON THE UCLA DATASET (SUPERSCRIPT "M" IS FOR RESULTS USING MAXIMUM MARGIN LEARNING AND 1NN [17])

Method	DT-7		DT-8		DT-9		DT-50		DT-SIR
Classifier	1NN	SVM	1NN	SVM	1NN	SVM	1NN	SVM	1NN
[41]	–	–	70.00	80.00	–	–	–	–	–
[10]	92.30	–	–	–	–	–	81.00	–	42.30
[17]	–	–	–	–	95.60^m	–	99.00^m	–	–
[6]	–	–	–	–	–	–	89.50	97.50	–
[5]	–	–	88.00	–	–	–	–	–	–
$WMFS_{V_t}$	89.30	96.95	91.89	95.53	91.85	95.15	93.00	99.40	45.50
$WMFS_{DT}$	96.87	98.45	97.18	96.96	96.95	97.11	99.12	99.75	61.25

is randomly selected using 75% (3 sequences for each class) of the whole dataset and the rest is used as the testing set. The experiment contains 200 random trials, and we achieve the average recognition rates of 99.12% using NN and 99.75% using SVM, which outperform previously tested methods. See Table V for the comparison of the results.

2) DT-SIR and the Results: To eliminate the effects owing to biases in identical viewpoint selection, the "shift-invariant recognition (SIR)" is set in [10] as follows. Each of the original 200 DT sequences is spatially divided into left and right, non-overlapping halves, resulting in totally 400 new sequences. The experimental setup follows that in [10]. Shown in Table V, we achieve the average recognition rate of 61.25% using NN, which is significantly higher than 42.30% in [10].

3) DT-9 and the Results: In the DT-50 breakdown, all four sequences in each class are captured with the same viewing parameters (e.g., identical viewpoint). To evaluate the performance across viewpoint change, a nine-class breakdown [17], named DT-9, is created by merging same classes in DT-50 from different viewpoints. The resultant nine classes are: boiling water (8), fire (8), flowers (12), fountains (20), plant (108), sea (12), smoke (4), water (12) and waterfall (16), where the number denotes the number of DT sequences in each class.

Our experiment is conducted using the same configuration in [17]. A training set is randomly selected 50% from the whole dataset (each class has two training sequences) and the rest is used as the testing set for each class. We run the experiment on 200 randomly generated training/testing sets. The average recognition rates are 96.95% by NN classifier and 97.11% by SVM classifier. See Table V for the classification rates of different methods in the DT-9 breakdown. The confusion matrices by NN and SVM are shown in Fig. 14.

4) DT-8 and the Results: The DT-8 breakdown [41] contains all classes of DT-9 except "plants", which has much more sequences than other classes. Our experiment on DT-8 is configured similarly as that on DT-9: we trained on 50% of DT-8 dataset and tested on the rest. See Table V for the classification rates of different methods on UCLA DT-8 dataset and the confusion matrices of the proposed methods are given in Fig. 14. The figure clearly shows the superior performance of the proposed approach.

5) DT-7 and the Results: The last breakdown on which we evaluate the proposed approach is DT-7 containing seven classes. It is proposed in [10] by first cutting spatially each sequence in the UCLA dataset into left and right halves and then dividing all 400 resulting sequences into seven semantic categories: flames (16), fountain (8), smoke (8), turbulence (40), waves (24), waterfall (64), vegetation (240). See Table V for the comparison of the classification rates of different methods on UCLA DT-7 dataset. The confusion matrices are in Fig. 14. Again, our approach shows superior performance.

6) More Discussions on the Proposed DT Descriptor: Fig. 15 shows the sequences from the two classes that are misclassified by our method in DT-50 breakdown: the sequences in class *fire* and class *smoke* are misclassified as the class *water*. The main reason is that there exist some flames in certain *fire* sequences which are similar to the light reflections in some *water* sequences; and the background with smokes in some *fire* sequences is nearly the same as the wave patterns in many *water* sequences. See Fig. 15 (a) for an illustration. Also, it is seen from Fig. 15 (b) that the similarity between the sequences from *smoke* and *water* makes it a challenging task to distinguish sequences from these two classes.

It is seen that the classification rates on DT-SIR are overall

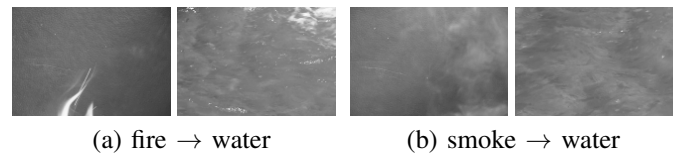(a) fire → water   (b) smoke → water

Fig. 15. (a) Sample sequences of the DT-50 dataset on which our proposed method fails; (b) sample sequences of DT-50 & DT-9 & DT-8 datasets on which our proposed method fails.

much lower than those on the other four breakdowns. It is because the DT-SIR eliminates the repetitive instances of DT sequences under the same viewpoint and significantly increased the variations of the DT sequences under different viewpoints. Considering the good robustness of multi-fractal analysis to view changes, it is not surprising to see that our method outperforms the other method. However, there are still a lot of room for future improvement on the robustness to view changes. Some failed sample sequences from DT-SIR by our approach are shown in Fig. 16.

For the other cases, Fig. 14 shows the misclassification rates of our method between any two classes using the confusion matrices. For the DT-9 breakdown, all the misclassification rates are below 5% except for the class *smoke* whose misclassification rate is 7.86%. The reason is the same as that in the DT-50 breakdown: the sequences in the class *smoke* are hard to distinguish from the sequences in the class *water*. The similar phenomenon is also observed in the DT-8 breakdown with misclassification rate 8.43% for the class *smoke*. The other seven misclassification rates of the proposed method are less than 4% in the DT-8 breakdown. For the DT-7 breakdown, the misclassification rates are below 2% for all classes except for the class *smoke*, which is misclassified as the class *turbulence* with an error rate 20.17%. The error is mainly due to the high inter-class similarity between these two.

To demonstrate how the temporal information of DT will bring additional discriminative information for DT, we compare the classification performances of $\text{WMFS}(V_t)$ and WMFS_{DT} in Table V. Recall that $\text{WMFS}(V_t)$ only captures the multi-fractal structure of DT sequences in spatial domain while WMFS_{DT} captures the multi-fractal structure of DT in spatial-temporal domain. The additional temporal information captured by WMFS_{DT} significantly improves the classification performance, especially on the DT-SIR breakdown which emphasizes the significant viewpoint variations. Such a phenomenon is not surprising as the spatial appearances of the inner-class DT sequences may vary significantly and only temporal information can help detecting the similarities of the inner-class DT sequences on their temporal dynamics.

VI. SUMMARY AND CONCLUSIONS

Motivated by the strong multi-scale power-law relationship in natural texture patterns, in this paper, we propose a new texture descriptor, wavelet-based multi-fractal spectrum (WMFS), for both static and dynamic textures. The power-law relationship is characterized by extracting the multi-fractal structure from the wavelet and wavelet leader pyramids of
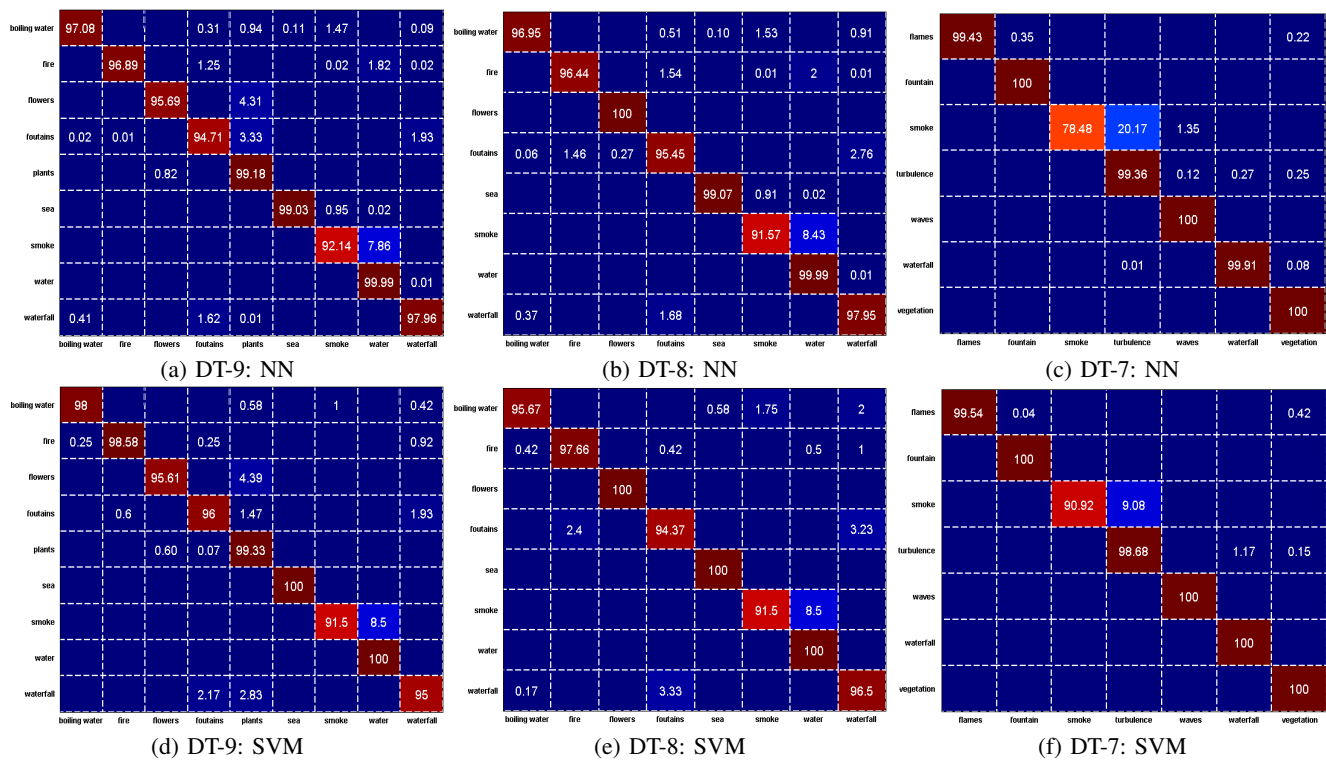
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

12

Fig. 14.   Confusion matrices of our method on the UCLA DT-9, DT-8 and DT-7 breakdowns.

SIR:class 15→ SIR:class 33   SIR:class 24→ SIR:class 22   SIR:class 44→ SIR:class 46   DT7: smoke→ DT7: turb.

Fig. 16.   Sample sequences from the datasets DT-SIR & DT-7 on which our proposed method fails.

images. Additionally, two more techniques, scale normalization and multi-orientation averaging are introduced to further improve the robustness of the wavelet-based multi-fractal analysis to scale and rotation variations, which are not well addressed by existing fractal analysis based approaches. By combining these techniques, we developed a texture descriptor with both high discriminative power and robustness against many environmental changes. There are several advantages of the proposed texture descriptor, including (a) high inter-class discriminability, (b) strong robustness to inner-class variations and many environmental changes, and (c) easy implementation as it avoids many complicated processing steps often used by other modern methods, e.g. clustering, texton generation and cross-bin comparisons.

Our proposed WMFS texture descriptor is based on the statistical measure on the distribution of different types of pixels. Thus, it also suffers from the often seen weakness as many other statistical methods, that is, it requires sufficient pixels to have an accurate and stable estimation. As a result, the WMFS does not work very well on the static texture images of very low resolution. It is empirically observed that when the image resolution is less than 64×64, it does not perform as well as some state-of-the-art methods such as (H+L)(S+R) method. However, such a weakness is not severe for DT recognition. The reason is that we are taking the average of the WMFSs over many 2D slices as the resulting descriptor which effectively suppresses the estimation variations. In the future, we would like to investigate more effective multi-fractal analysis tools for static texture of low resolution. Also, we are interested in studying the application of the proposed WMFS based DT descriptor in object recognition and action recognition for surveillance videos.

REFERENCES

[1] S. Arivazhagan and L. Ganesan, "Texture Classification Using Wavelet Transform", *Pattern Recog. Lett.*, vol. 24, pp. 1513-1521, 2003.

[2] A. Arneodo, N. Decoster, P. Kestener and S. Roux, "A Wavelet-based Method for Multifractal Image Analysis: From Theoretical Concepts to Experimental Applications", *Advances in Imaging and Electron Physics,* vol. 126, pp. 1-98, 2003.

[3] R. Azencott, S. Ping and L. Younes, "Texture Classification Using Windowed Fourier Filters", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 2, pp. 148-153, 1997.

[4] V. A. Billock, G. C. Guzman and J. S. Kelso, "Fractal Time and 1/f Spectra in Dynamic Images and Human Vision", *Physics D*, vol. 148, pp. 136-146, 2001.

[5] A. Chan, E. Coviello and G. Lanckriet, "Clustering Dynamic Textures with the Hierarchical EM Algorithm", in *Proc. IEEE Conf. on CVPR*, San Francisco, USA, 2010, pp. 2022-2029.

[6] A. Chan and N. Vasconcelos, "Classifying Video with Kernel Dynamic Textures", in *Proc. IEEE Conf. on CVPR*, Minneapolis, USA, 2007.

[7] A. Chan and N. Vasconcelos, "Probabilistic Kernels for the Classification of Auto-Regressive Visual Processes", in *Proc. IEEE Conf. on CVPR*, San Diego, USA, 2005, pp. 846-851.

[8] D. Chetverikov and R. Pëteri, "A Brief Survey of Dynamic Texture Description and Recognition", in *Proc. Int'l Conf. on Comput. Recog. Sys.*, 2005, pp. 17-26.

[9] K. Dana and S. Nayar, "Histogram Model for 3D Textures", in *Proc. IEEE Conf. on CVPR*, Santa Barbara, USA, 1998, pp. 618-624.

[10] K. G. Derpanis and R. P. Wildes, "Dynamic Texture Recognition Based on Distributions of Spacetime Oriented Structure", in *Proc. IEEE Conf. on CVPR*, San Francisco, USA, 2010, pp. 191-198.

[11] 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, 2002.

[12] G. Doretto, A. Chiuso, Y. N. Wu and S. Soatto, "Dynamic Texture", *Int'l J. Comput. Vision*, vol. 51, no. 2, pp. 91-109, 2003.

[13] G. Dorkó and C. Schmid, "Selection of Scale Invariant Neighborhoods for Object Class Recognition", in *Proc. IEEE ICCV*, Nice, France, 2003, pp. 634-640.

[14] O. Drbohlav and A. Leonardis, "Towards Correct and Informative Evaluation Methodology for Texture Classification under Varying Viewpoint and Illumination", *Computer Vision and Image Understanding*, vol. 114, no. 4, pp. 439-449, 2010.

[15] K. S. Falconer, *Techniques in Fractal Geometry,* first edition, John Wiley, 1997.

[16] S. Garding and T. Lindeberg, "Direct Computation of Shape Cues Using Scale-adapted Spatial Derivative Operators", *Int'l J. Comput. Vision*, vol. 17, no. 2, pp. 163-191, 1996.

[17] B. Ghanem and N. Ahuja, "Maximum Margin Distance Learning for Dynamic Texture Recognition", in *Proc. ECCV*, Crete, Greece, 2010, pp. 223-236.

[18] B. Ghanem and N. Ahuja, "Phase Based Modelling of Dynamic Textures", in *Proc. IEEE ICCV*, Rio de Janeiro, Brazil, 2007, pp. 1-8.

[19] J. V. Hateren, "Processing of Natural Time Series of Intensity by the Blowfly Visual System", *Vision Research*, vol. 37, pp. 3407–3416, 1997.

[20] E. Hayman, B. Caputo, M. Fritz and S. O. Eklundh, "On the Significance of Real-world Conditions for Material Classification", in *Proc. ECCV*, Prague, Czech, 2004, pp. 253-266.

[21] S. Jaffard, "Wavelet Techniques in Multifractal Analysis", *Fractal Geometry and Applications: A Jubilee of Benoit Mandelbrot, M. Lapidus and M. van Frankenhuijsen Eds.*, Proceedings of Symposia in Pure Mathematics, vol. 72, no. 2, pp. 91-152, 2004.

[22] U. Kandaswamy, S. A. Schuckers and D. Adjeroh, "Comparison of Texture Analysis Schemes Under Nonideal Conditions", *IEEE Trans. Image Process. *, vol. 20, no. 8, pp. 2260-2275, 2011.

[23] L. M. Kaplan, "Extended Fractal Analysis for Texture Classification and Segmentation", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 11, pp. 1572-1585, 1999.

[24] R. Kwitt and A. Uhl, "Efficient Texture Image Retrieval Using Copulas in a Bayesian Framework", *IEEE Trans. Image Process.* vol. 20, no. 7, pp. 2063-2077, 2011.

[25] R. Kwitt and A. Uhl, "Image Similarity Measurement by Kullback-Leibler Divergences between Complex Wavelet Subband Statistics for Texture Retrieval", in *Proc. IEEE ICIP*, 2008, pp. 933-936.

[26] B. Lashermes, S. Jaffard and P. Abry, "Wavelet Leader Based Multifractal Analysis", in *Proc. ICASSP*, Philadelphia, USA, 2005, pp. 161-164.

[27] S. Lazebnik, C. Schmid and S. Ponce, "A Discriminative Framework for Texture and Object Recognition Using Local Image Features", in *Toward Category-Level Object Recognition, Springer-Verlag*, pp. 423 - 442, 2006.

[28] S. Lazebnik, C. Schmid and S. Ponce, "A Sparse Texture Representation Using Affine-invariant Regions", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 27, pp. 1265-1278, 2005.

[29] T. Leung and S. Malik, "Representing and Recognizing the Visual Appearance of Materials Using Three-dimensional Textons", *Int'l J. Comput. Vision*, vol. 43, no. 1, pp. 29-44, 2001.

[30] T. Lindeberg, "Feature Detection with Automatic Scale Selection", *Int'l J. Comput. Vision*, vol. 30, no. 2, pp. 77-116, 1998.

[31] S. Malik, S. Belongie, T. Leung and S. Shi, "Contour and Texture Analysis for Image Segmentation", *Int'l J. Comput. Vision*, vol. 43, no. 1, pp. 7-27, 2001.

[32] S. Mallat, *A Wavelet Tour of Signal Processing,* Academic Press, CA: San Diego, 1998.

[33] B. B. Mandelbrot, *The Fractal Geometry of Nature,* San Francisco, CA: Freeman, 1982.

[34] J. Matas, O. Chum, M. Urban and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions", in *Proc. BMVC*, Cardiff, UK, 2002, pp. 384-393.

[35] K. Mikolajczyk and C. Schmid, "Scale and Affine Invariant Interest Point Detectors", *Int'l J. Comput. Vision*, vol. 60, no. 1, pp. 63-86, 2004.

[36] A. Oliva and A. Torralba, "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope", *Int'l J. Comput. Vision*, vol. 42, no. 3, pp. 145-175, 2001.

[37] R. Pëteri and D. Chetverikov, "Dynamic Texture Recognition Using Normal Flow and Texture Regularity", *Patt. Recog. Image Ana.*, LNCS 3523/2005, pp. 223-230, 2005.

[38] R. Polana and R. Nelson, "Temporal Texture and Activity Recognition", *Computational Imaging and Vision*, vol. 9, pp. 87-124, 1997.

[39] M. Pontil and A. Verri, "Support Vector Machines for 3D Object Recognition", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 6, pp. 637-646, 1998.

[40] C. M. Pun and M. C. Lee, "Log-polar Wavelet Energy Signatures for Rotation and Scale Invariant Texture Classification", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 590-603, 2003.

[41] A. Ravichandran, R. Chaudhry and R. Vidal, "View-invariant Dynamic Texture Recognition Using a Bag of Dynamical Systems", in *Proc. IEEE Conf. on CVPR*, Miami, USA, 2009, pp. 1651-1657.

[42] P. Saisan, G. Doretto, Y. Wu and S. Soatto, "Dynamic Texture Recognition", in *Proc. IEEE on CVPR*, Kauai, USA, 2001, pp. 58-63.

[43] M. Varma and R. Garg, "Locally Invariant Fractal Features for Statistical Texture Classification", in *Proc. ICCV*, Rio de Janeiro, 2007, pp. 1-8.

[44] J. L. Vehel, P. Mignot and J. P. Berroir, "Multifractals, Texture, and Image Analysis", in *Proc. IEEE Conf. on CVPR*, Maui, USA, 1992, pp. 661-664.

[45] G. Verdoolaege, Y. Rosseel, M. Lambrechts and P. Scheunders, "Wavelet-based Colour Texture Retrieval Using the Kullback-Leibler Divergence between Bivariate Generalized Gaussian Models", in *Proc. IEEE ICIP*, Kairo, 2009, pp. 265-268.

[46] G. Verdoolaege and P. Scheunders, "Geodesics on the Manifold of Multivariate Generalized Gaussian Distributions with an Application to Multicomponent Texture Discrimination", *Int'l J. Comput. Vision*, vol. 95, no. 3, pp. 265-286, 2011.

[47] H. Wendt, P. Abry, S. Jaffard, H. Ji and Z. Shen, "Wavelet Leader Multifractal Analysis for Texture Classification",in *Proc. IEEE ICIP*, Kairo, 2009, pp. 3829-3832.

[48] H. Wendt, S. G. Roux, S. Jaffard and P. Abry, "Wavelet Leaders and Bootstrap for Multifractal Analysis of Images", *Signal Processing*, vol. 89, no. 6, pp. 1100-1114, 2009.

[49] R. P. Wildes and S. R. Bergen, "Qualitative Spatiotemporal Analysis Using an Oriented Energy Representation", in *Proc. ECCV*, Dublin, 2000, pp. 768-784.

[50] S. Wu and M. S. Chantler, "Combining Gradient and Albedo for Rotation Invariant Classification of 2D Surface Texture", in *Proc. ICCV*, Nice, 2003, pp. 848-855.

[51] Y. Xu, S. B. Huang, H. Ji and C. Fermuller, "Combining Powerful Local and Global Statistics for Texture Description", in *Proc. IEEE Conf. on CVPR*, Miami, USA, 2009, pp. 573-580.

[52] Y. Xu, S. B. Huang, H. Ji and C. Fermuller, "Scale-space Texture Description on SIFT-like Textons", in *Computer Vision and Image Understanding*, vol. 116, no. 6, pp. 999-1013, 2012.

[53] Y. Xu, H. Ji and C. Fermuller, "Viewpoint Invariant Texture Description Using Fractal Analysis", *Int'l J. Comput. Vision*, vol. 83, no. 1, pp. 85-100, 2009.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

14

[54] Y. Xu, X. Yang, H. Ling and H. Ji, "A New Texture Descriptor Using Multifractal Analysis in Multi-orientation Wavelet Pyramid", in *Proc. IEEE Conf. on CVPR*, San Francisco, USA, 2010, pp. 161-168.

[55] J. Zhang, M. Marszalek, S. Lazebnik and C. Schmid, "Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study", *Int'l J. Comput. Vision*, vol. 73, no. 2, pp. 213-238, 2007.

[56] G. Zhao and M. Pietikäinen, "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expression", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol, 29, no. 6, pp. 915-928, 2007.

[57] S. C. Zhu, C. Guo, Y. Wang and Z. Xu, "What are Textons?", *Int'l J. Comput. Vision*, vol. 62, no. 1, pp. 121-143, 2005.

[58] S. C. Zhu, Y. Wu and D. Mumford, "Filters, Random Fields and Maximum Entropy (FRAME): Towards a Unified Theory for Texture Modeling", *Int'l J. Comput. Vision*, vol. 27, no. 2, pp. 107-126, 1998.

Yong Xu (M'96–SM'99) received the B.S., the M.S. and Ph.D. degrees in Mathematics from Nanjing University, China, in 1993, 1996 and 1999 respectively. From 1999 to 2001, he was a postdoc research fellow in computer science at South China University of Technology and became a faculty member afterward. He is currently a Professor with the School of Computer Science & Engineering at South China University of Technology, China. His research interests include image analysis, image/video recognition and image quality assessment. He is the member of IEEE Computer Society and of ACM.

Hui Ji received the B.Sc. degree in Mathematics from Nanjing University in China, the MSc degree in Mathematics from National University of Singapore and the PhD degree in Computer Science from University of Maryland, College Park. In 2006, He joined National University of Singapore as an Assistant Professor in Mathematics. Currently, he is an Associate Professor in Mathematics at National University of Singapore. His research interests include computational harmonic analysis, optimization, computational vision, image processing and biological imaging.

Xiong Yang received the B.S. and M.S. degrees in computer science from South China University of Technology, China, in 2009 and 2012 respectively. His research interests include image processing, medical image analysis, computer vision, machine learning and pattern recognition. He is currently working in private sector as a software engineer.

Haibin Ling received the B.S. degree in mathematics and the MS degree in computer science from Peking University, China, in 1997 and 2000, respectively, and the PhD degree from the University of Maryland, College Park, in Computer Science in 2006. From 2000 to 2001, he was an assistant researcher at Microsoft Research Asia. From 2006 to 2007, he worked as a postdoctoral scientist at the University of California Los Angeles. After that, he joined Siemens Corporate Research as a research scientist. Since fall 2008, he has been an Assistant Professor at Temple University. Dr. Ling's research interests include computer vision, medical image analysis, human computer interaction, and machine learning. He received the Best Student Paper Award at the ACM Symposium on User Interface Software and Technology (UIST) in 2003.