Texture Classification based on Bidimensional Empirical Mode Decomposition and Local Binary Pattern

JianJia Pan*
*Department of Computer Science
Hong Kong Baptist University
Hong Kong

YuanYan Tang*†
†Department of Computer and Information Science
University of Macau
Macau

Abstract—This paper presents a new simple and robust texture analysis feature based on Bidimensional Empirical Mode Decomposition (BEMD) and Local Binary Pattern (LBP). BEMD is a locally adaptive decomposition method and suitable for the analysis of nonlinear or nonstationary signals. Texture images are decomposed to several Bidimensional Intrinsic Mode Functions (BIMFs) by BEMD, which present a new set multi-scale components of images. In our approach, firstly, saddle points are added as supporting points for interpolation to improve original BEMD, and then images are decomposed by the new BEMD to several components (BIMFs). After then, Local Binary Pattern (LBP) in different sizes is used to detect features from different BIMFs. At last, normalization and BIMFs selection method are adopted for features selection. The proposed feature presents invariant while preserving LBP's simplicity. Our method has also been evaluated in CuRet and KTH-TIPS2a texture image databases. It is experimentally demonstrated that the proposed feature achieves higher classification accuracy than other state-of-the-art texture representation methods, especially in small training samples condition.

Keyword: Texture classification, Empirical Mode Decomposition, Local Binary Pattern, Invariant feature

I. INTRODUCTION

Texture analysis is widely recognized as a difficult and challenging computer vision problem. It provides many applications such as remote sensing image, medical image diagnosis, document analysis, and target detection, etc. Recently, using texture methods to face image analysis and motion analysis have presented many applications, which indicate that texture methods can be adopted to many new fields of computer vision problems.

Over the last several decades, there have been many methods proposed for texture classification, such as co-occurrence matrix, Gabor wavelet, Local Binary Pattern [21], maximum response 8 (VZ-MR8) [1], Basic Image Features (BIF) [2] etc. The statistics describing the whole region is then computed form these transformed local descriptors. The Gabor-based filter representation has been shown to be optimal in the sense of minimizing the joint two-dimensional uncertainty in space and frequency, and is widely used in image analysis [24]. LBP [21] is an operator for image description that is based on signs of different of neighbouring pixels. Varma and Zisserman [1] proposed the statistics VZ-MR8 classification algorithm that uses a rotationally invariant filter bank and

clustering to estimate the full joint probability distribution of filter responses. Basic Image Features presented by Griffin and Lillholm [2] are defined by a partition of the filter-response space of a set of six Gaussian derivative filters and the set of filters describes an image locally up to second order at some scale. Those methods are all state-of-the-art statistics algorithm and present good classification results on many databases.

Among these descriptors, LBP is a widely used local descriptor. It is simplicity and present excellent performance in various texture and face recognition, which has gained increasing attention. However, LBP describes over-local structure in image, and many improvements of original LBP have been proposed [22], [23], [24]. LBP operator is extended to use neighbourhoods for different sizes [21]. The sign and the magnitude of LBP, and the binary code of the intensity of centre pixels were combined together in CLBP [28] to improve texture classification. However, the intensity information is very sensitive to illumination changes; thus, CLBP needs image normalization to remove global intensity effects before feature extraction. Based on local phase and local surface type extracted from Riesz transform, Zhang [35] proposed a rotation invariant LBP feature (M-LBP) for texture classification. By use of Gabor wavelet, the LBP encode the local information and compress the redundancy in the Gabor filtered images in multi-scale and multi-direction and achieve effectiveness in texture representation [4], [3]. For variational applications, many other descriptors based on LBP are proposed [25], [28], [34].

Recently, Empirical Mode Decomposition (EMD), developed by Huang [5], has attracted more and more attentions in past decade and has been used for texture analysis [9]. The EMD method is based on the direct extraction of the energy associated with various intrinsic time scales. Expressed in Intrinsic Mode Functions (IMFs), they are the expansion basis which can be linear or nonlinear as dictated by data. EMD has been used to analyse two-dimensional signals [7], for example, images, which is known as Bidimensional EMD (BEMD).

BEMD presents some better quality than Fourier, wavelet and other decomposition algorithms in extracting intrinsic components of textures because of its data driven property [6], [7]. In this paper, we proposed an efficiency application of saddle points added BEMD [32] combined with LBP in texture

classification, and present the effectiveness of BEMD/BIMFs invariant properties for texture images.

Local Binary Pattern (LBP) is used as texture descriptor to detect the features of texture images' BIMFs. BEMD decomposed the original image to new multi-scale components (Bidimensional Intrinsic Mode Functions). In those new components, LBP histograms can achieve better efficiency than in the original image and present more illumination invariant features to supplement LBP to improve classification accuracy while preserving its simplicity. Experiments show texture image recognition rate based on our method is better than other state-of-the-art texture representation methods.

This paper is an extension of our previous work [32]. In this paper, we further extend the LBP-BEMD feature to variance normalization and BIMFs selection for performance improvement. We also provide a more in-depth analysis and more extensive evaluation.

II. REVIEW OF BEMD

Empirical mode decomposition (EMD) is a data driven processing algorithm, which applies no predetermined filter. EMD is based on the local characteristics scale of data, which is able to perfectly analyse nonlinear and nonstationary signals [5].

Nonstationary signals have statistical properties that vary as a function of time and should be analysed differently than stationary data. Rather than assuming that a signal is a linear combination of predetermined basis functions, in EMD, the data are instead thought of as a superposition of fast oscillations onto slow oscillations. EMD identifies those oscillations that are intrinsically present in the signal and produces a decomposition using these modes as the expansion basis.

EMD decomposes signal into components called Intrinsic Mode Functions (IMFs) satisfying the following two conditions [5]: (a)The numbers of extrema and zero-crossings must be either equal or differ at most by one; (b)At any point, the mean value of the envelope defined by the local maxima and the envelope by the local minima is zero.

Huang [5] also proposed an algorithm called 'sifting' to extract IMFs $J_k(t)$ from the original signal $f(t)$:

$$f(t) = \sum_{k=1}^{K} J_k(t) + r_K(t) \qquad (1)$$

Where $J_k(t), \quad k = 1, \ldots, K$ is IMFs and r_K is the residue.

The EMD is originally developed for one-dimensional (1D) data. Nunes.[9] firstly extended it to two dimensional BEMD, and decomposed images to bidimensional IMFs (BIMFs). The Bidimensional EMD (BEMD) process is conceptually the same as the one dimension EMD. The main process of the BEMD can be described as:

Step 1 Identify the local extrema (both maxima and minima) of the image $I(x, y)$;

Step 2 Generate the 2D envelopes by connecting maximum points (respectively, minima points) using surface interpolation. The local mean m is the mean of the two extrema envelopes. Follow Damerval's work [11], Delaunay triangulation and then cubic interpolation on triangles is used in our work;

Step 3 Subtract out the mean from the image to get a proto-BIMF $r = I - m$, judge whether r is a BIMF, if it is, go to Step 4. Otherwise, repeat Step 1 and Step 2 using the proto-BIMF r, until the latest proto-BIMF turns to be a BIMF;

Step 4 Input the proto-BIMF r to the loop from Step 1 to Step 3 to get the next remained BIMFs until it cannot be decomposed further.

After the BEMD, the decomposition of the image can be rewritten as following form:

$$I(i, j) = \sum_{k=1}^{K} d_k(i, j) + r(i, j) \qquad (2)$$

The $d_k(i, j)$ is the BIMFs of the images, and $r(i, j)$ is the residual function.

Although the discussions about EMD/BEMD lack concrete theoretical foundation until now [14], numerous tests demonstrated empirically that EMD is a powerful tool for the analysis of nonlinear and nonstationary data, especially for time-frequency-energy representations [8], [14], [16]. For two dimensional image, there are also some successful applications [15], [19]. In this work, we fine down the saddle points added BEMD combined with LBP features proposed in our previous work [32], and provide a more in-depth analysis and more extensive evaluation.

III. BEMD BASED ON SADDLE POINTS

One practical implementations in BEMD is the local extrema points detection. Which points should be detected as supporting points for the interpolation is an open problem. Mathematical morphology is used to detect local maxima and minima points in Nunes method[9]. Further, the local neighbour location method is also proposed for extrema detection [13]. However, by use of these methods, saddle points may not be detected. Saddle points are local maximum and local minimum points evaluated in different directions, and they also give important supporting features about the local variation of the original function. We added saddle points as supporting points for interpolation, which provided more significant components for texture classification.

In discrete condition, a point $u(x, y)$ in a 2D matrix U is a *saddle point* $u_{saddle}(x, y)$, if the product of the eigenvalues of the Hessian matrix is negatives:

$$u_{saddle}(x, y) = \{u(x, y) | u_{xx}u_{yy} - u_{xy}^2 < 0\} \qquad (3)$$

After detecting the saddle points, neighbour location method [13], [31] is used to detect the local maxima or minima points. In ordinary BEMD methods [6], [7], mathematical morphology is used to detect local maxima and minima points, but we found the numbers of extrema points reducing fast. It

means that the component will be too smooth to detect any signification extrema points after one or two times 'sifting'. To improve local extrema points detection, neighbour location method is used to detect extrema points. In this method, a data point $u(x, y)$ is considered as a local maximum (or. local minimum) if its value is strictly larger (or. lower) than the value of u at the nearest neighbours of points (x, y).

Let the window size for local extrema determination be $(2w + 1) \times (2w + 1)$, then

$$u(x, y) = \begin{cases} u_{max} & \text{if } u(x, y) > u(i, j), \\ u_{min} & \text{if } u(x, y) < u(i, j). \end{cases} \quad (4)$$

Where $\forall (i, j) \in W(x, y)$ and $W(x, y) = \{(i, j)|(x-w) \leq i \leq (x + w), (y - w) \leq j \leq (y + w), i \neq x, j \neq y\}$. From experiment and following the method in [13], we use the 3×3 window ($w = 1$). We find that result is an optimum extrema map for given images. The larger windows can be used in some other applications, but it will lead to a smaller number of extrema points for given texture images.

After then, the detected saddle points are added to maxima or minima points sets. In saddle points set U_{saddle}, a saddle point location (x, y)'s neighbourhood window is $U(k, l) = \{u|(x - T) \leq k \leq (x + T), (y - T) \leq l \leq (y + T)\}$.

$$\begin{array}{ll} u_{saddle}(x, y) \in U_{max} & \text{if} \quad N_{max} > N_{min}, \\ u_{saddle}(x, y) \in U_{min} & \text{if} \quad N_{max} < N_{min}. \end{array} \quad (5)$$

where u_{saddle} is saddle points set element (point), U_{max} is maximum points set, U_{mix} is minimum points set, N_{max} is the number of maxima points in window $U(k, l)$, and N_{min} is the number of minima points in window $U(k, l)$. It means that if number of maxima points is more than minima points in window $U(k, l)$, saddle point is considered to be maxima point, and vice verse. In experiments, the window size is 5×5 ($T = 2$). The recognition performance is nonsensitive to this saddle point location windows size. Experimental result about relationship between recognition performance and the windows size is shown in Section V-B.

These three type points, saddle points, neighbour local maxima and neighbour local minima points, are detected from image and used as supporting points for BEMD's interpolation. As shown in Section IV-B, by use of saddle points added BEMD, texture images are decomposed into Bidimensional Intrinsic Mode Functions ($BIMF_s$), which represent images' multi-scale components. The saddle points added BEMD detected more details (high local frequencies of oscillation) of images and contributed the performance of texture images classification.

IV. TEXTURE DESCRIPTOR BASED ON BEMD AND LBP

To analyse and classify texture images, we propose using LBP descriptor to extract local features from decomposed BIMFs. And then, the variance normalization and BIMFs selection are employed for performance improvement.
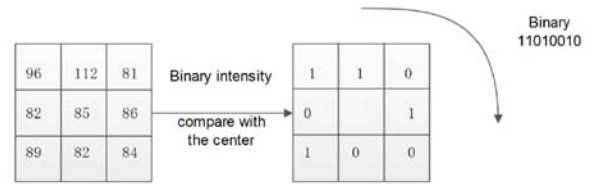

Fig. 1. The LBP operator

A. Local Binary Patterns (LBP)

LBP operator is originally developed for texture description. The operator assigns a label to every pixel of an image by thresholding the 3×3-neighbourhood of each pixel with the centre pixel value and considering the result as a binary number. Then histogram of the labels can be used as a texture descriptor [21].

The form of the resulting 8-bit LBP code can be defined as follows:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(u_n - u_c) 2^n \quad (6)$$

where u_c corresponds to the gray value of the centre pixel (x_c, y_c) into gray values of the 8 neighbourhood pixels, and function $s(m)$ is defined as:

$$s(m) = \begin{cases} 1 & \text{if } m \geq 0, \\ 0 & \text{if } m < 0. \end{cases} \quad (7)$$

LBP presents that it will be not affected by any monotonic gray-scale transformation which preserves the pixel intensity order in a local neighbourhood. Each bit of LBP code has the same significance level and that two successive bit values may have a totally different meaning.

To deal with textures at different scales, LBP operator is later extended to use neighbourhoods for different sizes [21]. The local neighbourhood is extended to as a set of sampling points evenly spaced on a circle centred at the pixel to be labelled allows any radius and number of sampling points [22]. If a sampling point is not in the centre of a pixel, it will be rebuilt by bilinear interpolation. The notation (P, R) is defined as the pixel neighbourhood which means P sampling points on a circle of radius of R. Figure.2 shows an example of circular neighbourhoods.

The number of patterns of original LBP grows with respect to the neighbourhood size, to address this problem, Ojala [21] observed that some patterns are more common than others, which is known as uniform patterns ('u2'). The number of transition between zero and one in uniform pattern at most two. For example, the patterns 01111000 and 11001111 are uniform whereas the patterns 11001001 and 01010011 are not. It is measured by:
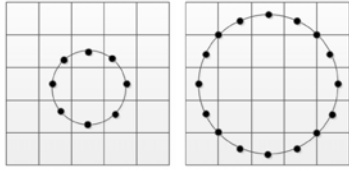
Fig. 2.   The circular (8,1) (16,2) neighbourhoods

Fig. 3.   Texture image and its BIMFs

$$U(LBP_{P,R}) = |s(u_{P-1} - u_c) - s(u_0 - u_c)| + \sum_{n=1}^{P-1} |s(u_n - u_c) - s(u_{n-1} - u_c)| \quad (8)$$

In uniform LBP, there is a separate output label for each uniform pattern and all the non-uniform patterns are assigned to a single label. Thus, the number of different output labels for mapping for patterns of P bits is $(P(P-1) + 3)$ [21]. For instance, the uniform mapping produces 59 output labels for neighbourhoods of 8 sampling points, and 243 labels for neighbourhoods of 16 sampling points. In the following, the mentioned LBP patterns are all uniform patterns.

B. LBP histograms of BIMFs

In this section, the proposed LBP via BIMFs feature frame is introduced. Figure.3 shows an example of texture image and its components BIMFs.

BEMD decomposes an image into its BIMFs basically on local frequency or oscillation information. The first BIMF contains the highest local frequencies of oscillation, the final BIMF contains the lowest local frequencies of oscillation and the residue contains the trend of the data. Corresponding high-frequency components are more robust to illumination changes [30]. BIMFs of image present a set of components of image from high-frequency to low-frequency. At the same time, the BEMD decomposition is an adaptive decomposition method. It is different from wavelet-based multi-scale analysis that characterizes the scale of a signal event using pre-specified basis functions. Moreover, corresponding BIMFs by saddle points added BEMD are able to capture more representative features of the original signal, especially more singular information in high frequency ones.

At the same time, LBP is a nonparametric method, which means that no prior knowledge about the distributions of images is needed.

We use the following procedure to extract texture features:

Firstly, the original image I is decomposed into its BIMF ($BIMF_s(i)$) by use of the saddle points added BEMD:
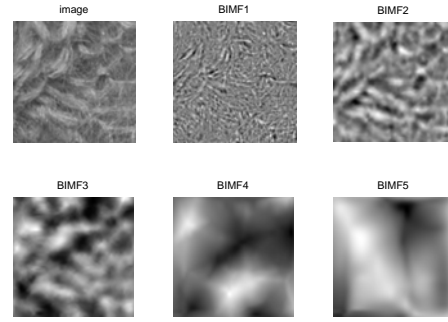
$$I = \sum_{i=1}^{K} BIMF_s(i) + r_N \quad (9)$$

Secondly, as we can find from texture images' BIMFs (Figure.3), the first and the second BIMF (higher BIMFs) remain the main detail of original image, and the last BIMFs (lower BIMFs) represent information in large scale. In our experiment, histograms of different size LBP ($LBP_{8,1}$ and $LBP_{16,2}$) for different BIMFs are mixed and the best combination is selected experimentally. All LBP patterns used in our algorithm are uniform patterns [21].

Thirdly, the LBP histograms of different BIMFs are normalized. Variance is a measure of how far a set of numbers are spread out from each other. Because we use variational size of LBPs to describe the BIMFs, the distributions of different BIMFs' LBP histograms are incongruous. To normalize LBP histograms, Variance-normalized LBP is defined as:

$$VLBP_{P,R}^{BIMF_s(i)} = LBP_{P,R}^{BIMF_s(i)} / VAR(LBP_{P,R}^{BIMF_s(i)}) \quad (10)$$

where $VAR(LBP_{P,R}^{BIMF_s(i)})$ is the variance of $LBP_{P,R}$ histogram of $BIMF_s(i)$:

$$VAR(x) = \sum_{i=1}^{n} p_i(x_i - \mu)^2 \quad (11)$$

μ is the expected value $\mu = \sum_{i=1}^{n} p_i x_i$ and p_i is the probability of x_i. $LBP_{P,R}$ describes the local feature of BIMFs and VAR describes the local variance.

Lastly, in EMD/BEMD literatures, there are findings that residual show trend of the whole signal/image. Figure.4 shows texture image samples and their BIMFs.

The higher BIMFs capture the detail information of original image, and the lower BIMFs capture the coarse contour information. Especially, the illumination and pose variety mainly appears in the residue. Therefore, this indicates that the lower BIMFs are sensitive to the variety. It is well understood that the variety effects can be reduced or eliminated by removing these lower BIMFs and residual. So we can just detecting the first two BIMFs for feature detection, and their LBP histograms are concatenated as the feature vector of image, which is Variance-normalized Saddle points added BEMD LBP:

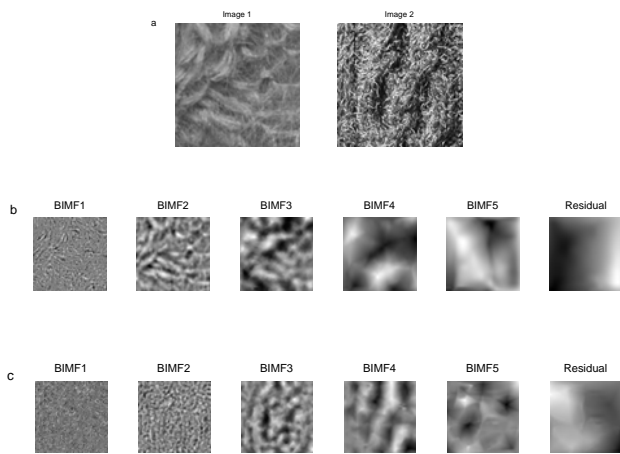$$VSBEMDLBP = \{VLBP_{P,R}^{BIMF_s(i)} | i = 1, 2\} \quad (12)$$

Fig. 4. Images and their BIMFs (a)Images from same class of KTH-TIPS2a (b)BIMFs of Image 1 (c)BIMFs of Image 2

Fig. 5. Some examples from KTH-TIPS2a database

TABLE I.   CLASSIFICATION ACCURACY OF PROPOSED METHODS WITH DIFFERENT SIZE WINDOWS IN THE DATABASES

windows size	KTH-TIPS2a	CUReT
3×3	96.36%	97.61%
5×5	96.96%	98.00%
7×7	96.30%	97.87%
9×9	96.26%	97.68%
11×11	96.42%	97.33%

There are many combination choices to combine different size $LBP(P,R)$ with different $BIMF_s(i)$. To find the optimum combination of $LBP(P,R)$ with $BIMF_s(i)$, different combination choices are used and their classification accuracy are compared. Some discussion and experimental result will be presented in Section V-B.

V. EXPERIMENT AND DISCUSSION

To validate the effectiveness of proposed VSEMDLBP feature, we carried out a series of experiment on two large databases compared with other methods: KTH-TIPS2a database [36] and CUReT database [37]. Nearest neighbourhood classifier (NN) is used for classification.

A. Databases and dissimilarity measurement

The KTH-TIPS2a database [36] is a database widely used for texture classification and material categorization. KTH-TIPS2a database contains 4 samples of 11 different materials. The sample images at 9 different scales, 12 lighting and pose setups. It contains 11 texture classes with 4572 images. Some examples from different classes are shown in Figure.5. It appears small inter class variations between textures and large intra class variations in the same class. In the top row, all images are of the same texture while in different scales and lighting/pose setups. In the bottom row, the images appear similar and yet they are belonging to different classes.

The images are 200×200 pixels in size (as in ref.[27], we did not include those images which are not of this size, so the experimental data is 10 class with 396 samples pre class, totally 3960), and all of images are transformed into 256 gray levels.

The CUReT database [37] contains images of 61 materials and includes many surfaces commonly seen in our environment [24]. Each of the materials in the database has been imaged under different viewing and illumination conditions. The effects of surface normal variations such as specularities, reflections and shadowing are evident. This database also includes some man-made textures, and is highlighted due to abundant imaging
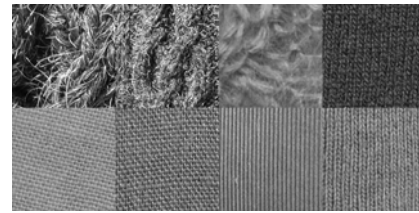
conditions. These make it far more challenging and become a benchmark widely used to assess classification performance.

There are 118 images which have been shot from a viewing angle of $< 60^o$. Follow ref.[24], in these 118 images, we selected 92 images, from which a sufficiently large region could be cropped (200×200) across all texture classes. And then they are converted all the cropped regions to gray level.

We use χ^2 statistic to measure the dissimilarity of sample and model histograms. Thus, a test sample x_t will be assigned to the class of model x_m that minimizes:

$$D(x_t, x_m) = \sum_{n=1}^{N} (x_t(n) - x_m(n))^2 / (x_t(n) + x_m(n)) \quad (13)$$

where N is the number of bins, and $x_t(n)$ and $x_m(n)$ are the values of the sample and model histogram at the nth bin, respectively.

B. Parameters selection and feature combination selection in experiment

In section III, we proposed to combine the saddle points to maximum points set or minimum points set based on the numbers of maximum points and minimum points in the saddle point location neighbourhood windows. In this section, we firstly gave some experiments to show the relationship between the size of saddle point location windows and the classification performance. The test data sets are KTH-TIPS2a (40 samples per class for training and 356 samples per class for testing) and CUReT (46 samples per class for training and 46 samples per class for testing). The used VSBEMDLBP feature is $VLBP_{8,1}^{BIMF_s(1)} VLBP_{8,1}^{BIMF_s(2)} VLBP_{16,2}^{BIMF_s(1)}$.

As Table.I shows, the classification performance is non-sensitive to the size of saddle point location windows. The value of T don't effect the total number of extrema points. T just effects the distribution of number of added saddle points between maximum set and minimum set. When T changes from 1 to 5 (windows size changes from 3 to 11), the different number of saddle points between maximum set and minimum set accounts for 5-10 percent of total number of saddle points,

TABLE II.   CLASSIFICATION ACCURACY OF PROPOSED METHODS WITH COMBINATIONS OF DIFFERENT $LBP_{P,R}$ WITH DIFFERENT $BIMF_s(i)$ IN THE DATABASES

Combination	KTH-TIPS2a	rank	CUReT	rank
$V_{8,1}^1 V_{8,1}^2 V_{16,2}^1$	96.96%	2	98.00%	1
$V_{8,1}^1 V_{8,1}^2$	98.27%	1	97.39%	3
$V_{8,1}^2 V_{16,2}^1$	96.38%	3	97.92%	2
$V_{8,1}^1 V_{16,2}^1$	93.19%	9	96.34%	5
$V_{8,1}^1$	93.17%	10	94.38%	8
$V_{16,2}^1 V_{16,2}^2$	92.37%	11	96.91%	4
$V_{8,1}^1 V_{8,1}^2 V_{16,2}^1 V_{16,2}^2$	95.62%	4	92.67%	10
$V_{8,1}^1 V_{8,1}^2 V_{16,2}^2$	94.41%	5	86.77%	14

thus T's influences on recognition result is small. For other combination features, the same conclusion can be found. In the following experiments, the window size is fixed to 5×5 ($T = 2$).

Further, in Section.IV-B, we proposed the VSBEMDLBP features combining the $LBP_{P,R}$ and $BIMF_s(i)$. The performance of different combination is different. Experiments to 'optimum' the choice of combining different $LBP_{P,R}$ with different $BIMF_s(i)$ is shown in Table.II. The test data sets are KTH-TIPS2a (40 samples per class for training and 356 samples per class for testing) and CUReT (46 samples per class for training and 46 samples per class for testing). Because there are many different combinations, we just reported the top five combinations result in the databases.

In order to simplify the expression, $V_{P,R}^i$ is short for the $VLBP_{P,R}^{BIMF_s(i)}$ and $V_{P,R}^1 V_{P,R}^2$ means to concatenate the $VLBP_{P,R}^{BIMF_s(1)}$ and $VLBP_{P,R}^{BIMF_s(2)}$ as the feature vector, etc.

Form TableII, we can find that the best performances are focus on three combinations ($V_{8,1}^1 V_{8,1}^2$, $V_{8,1}^1 V_{8,1}^2 V_{16,2}^1$ and $V_{8,1}^2 V_{16,2}^1$) for different databases. It indicated that the high frequency $BIMF_s(1)$ and $BIMF_s(2)$ present more representative features of the original image, and the high-frequency components are more robust to variant of texture image.

At the same time, other combinations' recognition performance varied in different database, but the top five combinations still can achieve good result. As we discussed before, the lower BIMFs are sensitive to the variety. The top 5 results all don't contain the lower BIMFs ($BIMF_s(4)$ and $BIMF_s(5)$), and even the $BIMF_s(3)$ (in fact, the combination $V_{8,1}^1 V_{8,1}^3 V_{16,2}^1$ achieved rank 7 in CUReT database (95.24%)). This indicated that these BIMFs can be removed to reduce variety effects.

In the following experiment, $V_{8,1}^1 V_{8,1}^2 V_{16,2}^1$ feature is used to compare with other methods in CUReT database. $V_{8,1}^1 V_{8,1}^2$ feature is used to compare with other methods in KTH-TIPS2a database. And we also reported the top three $V_{8,1}^1 V_{8,1}^2$, $V_{8,1}^1 V_{8,1}^2 V_{16,2}^1$ and $V_{8,1}^2 V_{16,2}^1$ features classification performance in the databases.

C. Classification result on texture database

After selecting the combination of $VLBP_{P,R}^{BIMF_s(i)}$, we evaluated the classification performance using our methods in the two texture databases: KTH-TIPS2a database and CUReT database.

The original LBP [21], improved LBP approach (CLBP [28]) and two statistical approaches (VZ-MR8 [1], BIF [2]) are used as comparison. At the same time, to validate the effectiveness of proposed saddle points added BEMD decomposition, we also carried out LBP combined with other transform methods: Gabor filters -LBP [4], [3], M-LBP (Riesz transform) [35], and original BEMD decomposition [11]. The VBEMD-LBP feature is to detect LBP features by our procedure from BIMFs by used of BEMD (no saddle points added BEMD) proposed in [11] (Ref.[11] just proposed BEMD method rather than the features).

1) Experiment result on KTH-TIPS2a database: For KTH-TIPS2a database, we firstly repeated the small training samples approach (1-7 samples per class for training, 389 samples per class for testing) with 100 random combinations as training and testing data and the results are reported as average value and shown in Table.III. It means that the training data is independent from physical, materials, illumination, pose, and scale. This small training samples approach only supports a few partial training samples and little knowledge about the data.

Secondly, we repeated the experiment with 100 different random selection of training and testing data (1-50 samples per class for training, 346 samples per class for testing) and reported the proposed and compared approaches results in a range of training set sizes (as in Crosier[2]), which is shown in Figure.6. In Figure.6 we just show the best result of proposed VSBEMDLBP ($V_{8,1}^1 V_{8,1}^2$) and VBEMD-LBP ($V_{8,1}^1 V_{8,1}^2$). The classifier is Nearest Neighbour classifier (NN).

The small training samples result classification accuracies are listed in Table.III. It can be seen that VBEMD-LBP and VSBEMDLBP have better performance than other methods. When the number of training samples pre class is small compared to the testing samples, in this case, 1-7 training samples ps. 389 testing samples, the recognition rates by other methods are dropped, especially for LBP. This is mainly because there are different scales, lighting and pose setups in KTH-TIPS2a and the number of samples per class is large. The proposed methods achieve the highest recognition rates among all the competing methods. Particularly, it is less sensitive to the small sample size problem.

Secondly, from Figure.6, we can note that the performance ranking of the eight representations tested remains the same regardless of the number of images in the training set. This can be seen as confirming the uncommitted nature of the nearest neighbour classifier used with each of the representations. VSBEMDLBP-columns score highest, followed by VBEMD-LBP, BIF and CLBP representations. When the number of training samples are relatively high such as 50 samples per class, the difference between the recognition rates of VSBE-MDLBP and other methods is getting smaller.

The performance of VZ-MR8 is significantly lower than other approaches. VZ-MR8 textons are trained from 40 samples of per class with 122 textons, the totally textons number is 1220. There are 396 samples per class in KTH-TIPS2a, textons from 40 samples per class maybe not representative enough. More training samples and more textons pre class may improve the result, but the training time and storage space will be very large and the cluster result may not be ideal for classification.

TABLE III. CLASSIFICATION ACCURACY OF VSBEMDLBP AND COMPARED APPROACHES IN THE KTH-TIPS2A DATABASE WITH DIFFERENT NUMBER OF TRAINING SAMPLES

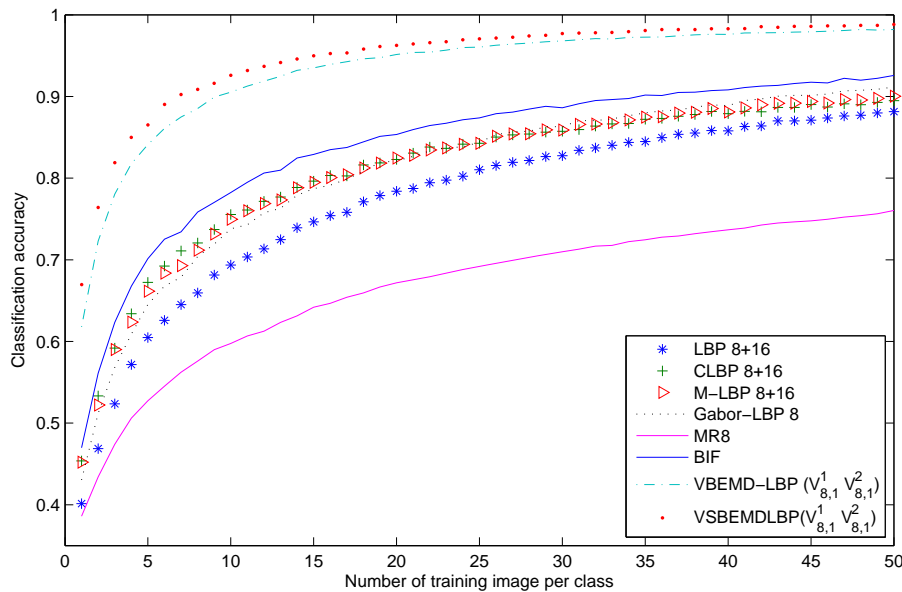| No.of training samples	1	2	3	4	5	6	7
LBP[21] (8,1)+(16,2)	40.14%	46.88%	52.36%	57.17%	60.47%	62.59%	64.50%
CLBP[28] (8,1)+(16,2)	45.37%	53.33%	59.19%	63.39%	67.23%	69.23%	71.08%
Gabor-LBP[4] (8,1)	42.04%	49.19%	54.81%	58.91%	62.52%	64.78%	65.88%
M-LBP [35] (8,1)+(16,2)	45.21%	52.22%	59.01%	62.38%	66.16%	68.37%	69.29%
VZ-MR8[1]	38.60%	43.41%	47.41%	50.60%	52.72%	54.52%	56.24%
BIF[2]	46.98%	56.11%	62.33%	66.77%	70.14%	72.53%	73.42%
VBEMD-LBP ($V_{8,1}^1 V_{8,1}^2$)	59.79%	70.28%	76.04%	79.77%	82.11%	84.09%	87.48%
VBEMD-LBP ($V_{8,1}^1 V_{8,1}^2 V_{16,2}^1$)	55.08%	66.33%	71.78%	75.90%	78.49%	80.92%	82.17%
VBEMD-LBP ($V_{8,1}^2 V_{16,2}^1$)	52.85%	61.93%	67.58%	71.36%	75.22%	77.15%	78.39%
VSBEMDLBP ($V_{8,1}^1 V_{8,1}^2$)	66.96%	76.39%	81.89%	84.99%	86.53%	89.04%	90.23%
VSBEMDLBP ($V_{8,1}^1 V_{8,1}^2 V_{16,2}^1$)	62.25%	71.80%	76.56%	80.76%	83.10%	85.02%	86.46%
VSBEMDLBP ($V_{8,1}^2 V_{16,2}^1$)	59.53%	68.36%	74.28%	78.34%	80.40%	82.03%	84.08%

Fig. 6. The mean proportion of correctly classified images over 100 random splits of the KTH-TIPS2a database into training/test data, for a range of training set sizes. The best result for $VSBEMDLBP(V_{8,1}^1 V_{8,1}^2)$-columns (with 50 training images per class) is 98.83%

Thirdly, there are 396 samples per class in KTH-TIPS2a at different scales, lighting and pose setups, which are high intra-class scatter. Hu [18] and some other researchers have pointed out that the BIMFs' residual is a trend of whole image. For KTH database texture images, the illumination and pose of images can be viewed as trend of images. The high oscillation information ($BIMF(1)$ and $BIMF(2)$) are not only more robust to illumination changes [30] but also more robust to pose and scales changes. In our method, residual and some lower BIMFs are removed in the VBEMD-LBP and VSBEMDLBP features, which means that the variable of samples are reduced by removing the lower BIMFs and residual and achieves a lower intra-class scatter.

2) Experiment result on CUReT database: For CURet database, there are three training approaches used: 46 training samples C_{46}, 23 training samples C_{23} and 3 training samples C_3. C_{46} means 46 samples from each class are training data, and other 46 samples are testing data. It is the normal method in previous works [23], [24]. C_{23} means 23 samples from each class be training data, and other 69 samples are testing

data[35]. C_3 means only 3 samples from each class be training data, and other 89 samples are testing data, which is a small training size data. Using the above three settings rather than just one made it possible to better investigate the properties of different operators [24]. The C_{46} data is to simulate the condition that there are enough training samples. C_{23} data is used to simulate the situation of small but comprehensive training set. C_3 data is to simulate the condition that there is only a few partial training samples.

We firstly repeated the three training methods (C_{46}, C_{23}, C_3) with 100 random combinations as training and testing data and the results are reported as average value and shown in Table.IV. Secondly, we repeated the experiment with 100 different random selection of training and testing data (1-46 samples per class for training, 46 samples per class for testing, C_{46}) and reported the combined features results in a range of training set sizes (as in Crosier[2]), which is shown in Figure.7. The small training samples part (1-7 samples per class for training, 46 samples per class for testing, C_{46}) classification accuracies are listed in Table.V. The classifier is

TABLE IV.   CLASSIFICATION ACCURACY OF VSBEMDLBP AND COMPARED APPROACHES IN THE CUReT DATA

Data setting	C_3	C_{23}	C_{46}
LBP[21] (8,1)+(16,2)	60.14%	88.10%	92.99%
CLBP[28] (8,1)+(16,2)	61.00%	89.91%	94.33%
Gabor-LBP[4] (8,1)	64.82%	90.22%	94.37%
M-LBP [35] (8,1)+(16,2)	67.76%	94.39%	97.59%
VZ-MR8[1]	65.02%	92.48%	96.35%
BIF[2]	73.51%	96.49%	98.56%
VBEMD-LBP ($V_{8,1}^1 V_{8,1}^2 V_{16,2}^1$)	73.25%	94.76%	97.28%
VBEMD-LBP ($V_{8,1}^2 V_{16,2}^1$)	71.03%	93.70%	96.82%
VBEMD-LBP ($V_{8,1}^1 V_{8,1}^2$)	70.55%	93.14%	96.25%
VSBEMDLBP ($V_{8,1}^1 V_{8,1}^2 V_{16,2}^1$)	76.29%	95.90%	98.00%
VSBEMDLBP ($V_{8,1}^2 V_{16,2}^1$)	75.18%	95.69%	97.92%
VSBEMDLBP ($V_{8,1}^1 V_{8,1}^2$)	74.63%	94.98%	97.39%

Nearest Neighbour classifier (NN).

Table.IV shows classification accuracy of proposed methods and other methods in CURet database. By comparing classification rates, we can find that VBEMD-LBP achieve better accuracy that other LBP-based methods and VSBEMDLBP accuracy is higher than other methods except BIF. In the few training samples condition C_3, accuracies of all methods are dropped, only the proposed features and BIF achieve classification accuracy more than 70% and VSBEMDLBP is even better than BIF. It indicates that the proposed VSBEMDLBP is more robust for real applications where training samples are limited and not comprehensive. On the other hand, Crosier [2] has proposed that the BIF feature achieves $5 - 10\%$ higher classification rate than LBP-based methods in CURet database, the main superiority of BIF over our method comes from the local feature detector. At the same time, BIF feature is detected by a series of multi-scale filters' responses, which verifies the multi-scale representations will be more efficiency.

We repeated the experiment with 100 different random selection of training and testing data and reported the results in a range of training set sizes, which is shown in Figure.7. We can note that the performance ranking of the eight representations tested remains the same regardless of the number of images in the training set. BIF-columns score highest, followed by VS-BEMDLBP, VBEMD-LBP and VZ-MR8 representations. In the small training samples conditions, VSBEMDLBP presents better result than BIF, and their performances are similar after training samples are more than 10. The detail result of the small training samples conditions can be found in Table.V.

In CURet database, VZ-MR8 textons are trained following the approaches in [1] (10 textons training from 13 samples per class, totally 610 textons). The performance of VZ-MR8 is significantly higher than in another database. VZ-MR8 feature is sensitive to the choice of different number of textons and the textons cluster result in different database.

As Table.III to Table.V show, transform based LBP features present better results compared with LBP and CLBP feature. It shows multi-scale or frequency domain representations extract features in character level, which is proved to be more discriminative feature level. Further, as we discussed before, unlike other priori transform methods (Gabor wavelet, etc.), BIMFs depend on an adaptive decomposition and present a different time-frequency space and more meaningful components.

D. Discussion

In the two databases, the proposed VSBEMDLBP features achieve higher recognition result. Especially, it is less sensitive to the small training sample problem. When the number of samples are relatively high such as 396 samples per class and the number of training samples are relatively low such as 1-7 samples per class in KTH-TIPS2a database, the performance difference between VSBEMDLBP and other methods is higher.

As shown in the Section IV-B, BIMFs of image present adaptive multi-scale components. The higher BIMFs contain the higher local frequencies of oscillation and are more robust to illumination, pose, scale changes. The variety effects can be reduced or eliminated by removing lower BIMFs and residual as in our method. The saddle points added BEMD achieves a better classification result than other transform-based method included the original BEMD. To our best knowledge, the local descriptor based on BIMFs is a new framework to validate the BEMD's powerful compared with other transforms in two dimensional. At the same time, as a decomposition-based method, this framework could be applied to different LBP variants, such as CLBP, Dominant LBP [27], LBP histogram fourier features [23], and other descriptors, for example, BIF.

Although the local descriptor based on BIMFs framework is a powerful method, there are some challenges that should be addressed in the future. The first challenge is choosing the optimum combination of $LBP_{P,R}$ and $BIMF_s(i)$. Though we have validated the performances of the three combinations ($V_{8,1}^1 V_{8,1}^2$, $V_{8,1}^1 V_{8,1}^2 V_{16,2}^1$ and $V_{8,1}^2 V_{16,2}^1$) are better than other combinations in the experiment databases as in section.V-B, the performances of other combinations varied and depended on the texture database. As a future work, more theoretically research is needed.

The next challenge in BEMD's applications is its time complexity. The main time consumption of BEMD is from the many times two dimensional interpolations (the step 2 of Bidimensional EMD (BEMD) process), which is still known to be a time-consuming problem. We compared the time complexity experimentally. The experimental computer is Inter, Core(TM)2, CPU, Q6600 @2.40GHz. The platform is MATLAB R2011a. For a 200×200 pixel image, average BEMD decomposition time was 3.7 seconds, which is more than wavelet (0.06 seconds) and Riesz transform (0.53 seconds). The method with higher time consumption than our method is VZ-MR8 (154 seconds per image for KTH-TIPS2a database and 32 seconds per image for CUReT database), which took a lot of time to cluster textons. BIF took average 3.6 seconds per image to detect feature. LBP and CLBP consumed smaller than 0.05 seconds per image.

Based on the discussion and experimental result of Section IV-B, the best combination of VSBEMDLBP just detected LBP histograms of the first two BIMFs, so in the practice, we can only decompose image to the first two BIMFs and then stop the BEMD processing. Its average decomposition time was 2.5 seconds and faster than BIF. To further reduce computation complex of EMD/BEMD, we have developed a fast decomposition method for one dimensional EMD [33], which only takes 32% time consumption of original EMD and presents more meaningful IMFs. But for two dimensional BEMD, more research is needed to reduce the time while

TABLE V.   CLASSIFICATION ACCURACY OF VSBEMDLBP AND COMPARED APPROACHES WITH DIFFERENT NUMBER OF TRAINING SAMPLES (PART) IN C_{46}

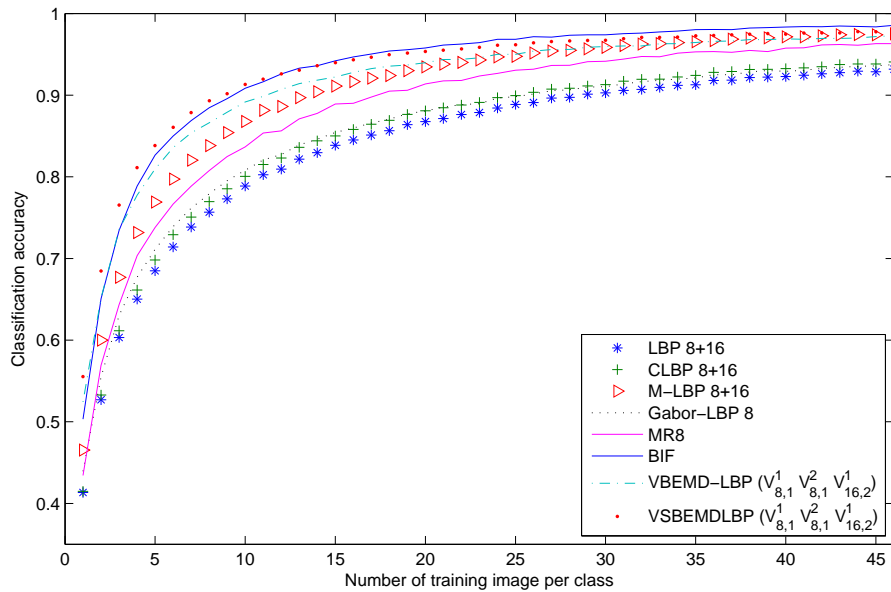| No.of training samples	1	2	3	4	5	6	7
LBP[21] (8,1)+(16,2)	41.34%	52.69%	60.31%	65.02%	68.48%	71.40%	73.85%
CLBP[28] (8,1)+(16,2)	41.48%	53.27%	61.15%	66.13%	69.81%	72.91%	75.06%
Gabor-LBP[4] (8,1)	43.98%	55.49%	63.01%	67.71%	71.12%	73.91%	76.17%
M-LBP [35] (8,1)+(16,2)	46.55%	59.99%	67.70%	73.18%	76.90%	79.72%	82.03%
VZ-MR8[1]	43.42%	56.93%	64.35%	70.34%	73.80%	76.66%	78.85%
BIF[2]	50.32%	65.07%	74.44%	78.87%	82.65%	84.99%	86.87%
VBEMD-LBP ($V_{8,1}^1 V_{8,1}^2 V_{16,2}^1$)	52.43%	65.19%	73.56%	77.75%	80.86%	83.58%	85.38%
VBEMD-LBP ($V_{8,1}^2 V_{16,2}^1$)	52.31%	64.24%	71.16%	76.12%	73.35%	81.56%	83.58%
VBEMD-LBP ($V_{8,1}^1 V_{8,1}^2$)	51.76%	63.46%	70.81%	75.42%	78.38%	80.96%	82.91%
VSBEMDLBP ($V_{8,1}^1 V_{8,1}^2 V_{16,2}^1$)	55.54%	68.46%	76.54%	81.14%	83.83%	86.08%	87.86%
VSBEMDLBP ($V_{8,1}^2 V_{16,2}^1$)	53.96%	67.14%	74.18%	78.90%	82.05%	84.47%	86.14%
VSBEMDLBP ($V_{8,1}^1 V_{8,1}^2$)	54.82%	67.04%	74.24%	78.87%	82.13%	84.43%	86.33%

Fig. 7.   The mean proportion of correctly classified images over 100 random splits of the CURet database into training/test data, for a range of training set sizes. The best result for BIF-columns (with 46 training images per class) is 98.56%

preserving the decomposition characteristics.

VI.   CONCLUSION

Texture analysis is a difficult and challenging computer vision problem. In this paper, a new powerful method (VSBE-MDLBP) is proposed for texture classification. An adaptive decomposition method (saddle points added BEMD) is used to supply new components ($BIMF_s$). The saddle points added BEMD detected more details (high local frequencies of oscillation) of images and contributed the performance of texture images classification. At the same time, the higher BIMFs capture the detail information of original image, and the lower BIMFs capture the coarse contour information. Especially, the illumination and pose variety mainly appears in the residue. The higher frequency $BIMF_s$ present more invariant properties for texture classification. In these new adaptive multi-scale components (higher frequency $BIMF_s$), LBP descriptor can achieve better performance than in original images and other transform-based methods. Experiments show the texture image recognition rate based on our method is better than other state-of-the-art texture representation methods. Especially, it is less sensitive to the small training sample size problem.

REFERENCES

[1] M. Varma, A. Zisserman, "A statistical approach to texture classification from single image", International Journal of Computer Vision, vol.62(1),pp.61-81, 2005 .

[2] M. Crosier, LD. Griffin, "Using basic image features for texture classification", International Journal of Computer Vision, vol.88(3), pp.447-460, 2010.

[3] M. Yang, L. Zhang, "Gabor Feature based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary", Lecture Notes in Computer Science, 2010 European Conference on Computer Vision pp.448-461, 2010.

[4] Y. Su, D. Tao, X.L. Li, X. Gao, "Texture Representation in AAM using Gabor Wavelet and Local Binary Patterns" 2009 IEEE International Conference on Systems, Man, and Cybernetics, San Antonio, TX, USA, pp.3274-3279, 2009

[5] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", Royal Society.Lond, vol.454, pp.903-1005, 1998.

[6] N.E. Huang, M.L.C. Wu, S.R. Long, "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis", Royal Society A, vol.459, pp.2317-2345, 2003.

[7] J.C. Nunes, Y. Bouaoune, E. Delechelle, O.Niang, Ph. Bunel, "Image analysis by bidimensional empirical mode decomposition", Image and Vision Computing, vol.21(12), pp.1019-1026, 2003 .

[8] P. Flandrin, G. Rilling, P. Goncalves, "Empirical mode decomposition as a filter bank", IEEE Signal Processing Letters, vol.11(2), pp.112-114, 2004.

[9] J.C. Nunes, S. Guyot , E. Deechelle, "Texture analysis based on local analysis of the Bidimensional Empirical Mode Decomposition", Machine Vision and Applications, vol.16(3), pp.177-188, 2005.

[10] A. Linderhed, "Variable sampling of the empirical mode decomposition of two-dimensional signals", International Journal of Wavelets, Multiresolution and Information Processing., vol.3(3), pp.435-452 , 2005.

[11] C, Damerval, S. Meignen, V. Perrier, "A fast algorithm for bidimensional EMD", IEEE Signal Processing Letter, vol.12(10), pp.701-704, 2005.

[12] R. Bhagavatula, S. Marios, M. Acoustics, "Analyzing Facial Images using Empirical Mode Decomposition for Illumination Artifact Removal and Improved Face Recognition", 2007 IEEE International Conference on Speech and Signal Processing, pp.505-508, April 2007.

[13] M.A.B. Sharif, R.A. Reza, F.K. Jesmin, "Fast and Adaptive Bidimensional Empirical Mode Decomposition Using Order-Statistics Filter Based Envelope Estimation", Journal on Advances in Signal Processing, Article ID 728356, 18 pages, 2008.

[14] N.E. Huang, Z. Wu, "A review on Hilbert-Huang transform Method and its applications to geophysical studies", Reviews of Geophysics, vol. 46, RG2006, 23 PP., 2008.

[15] Q.W. Xie, B. Xuan, S.L. Peng, J.P. Li, W.X. Xu, H. Han, "Bandwidth empirical mode decomposition and its application", International Journal of Wavelets, Multiresolution and Information Processing, Vol.6(6), pp.777-798, 2008.

[16] Z. Wu, N.E. Huang, "On the filtering properties of the Empirical Mode Decomposition", Advances in Adaptive Data Analysis, vol. 2(4), pp.397-414, 2010.

[17] Z. Wu and N.E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method",. Advances in Adaptive Data Analysis, vol.1(1), pp.1-41, 2009.

[18] X. Hu, W. Xia, S.L. Peng, W.L. Hwang, "Multiple component predictive coding framework of still images", 2011 IEEE International Conference on Multimedia and Expo (ICME), Barcelona, Spain , pp.1-6, July 2011.

[19] O. Niang, A. Thioune, M.C.E Gueirea, E. Delchelle, J. Lemoine, "Partial Differential Equation-Based Approach for Empirical Mode Decomposition: Application on Image Analysis", IEEE Transactions on Image Processing, Vol.21(9), pp.3991-4001, 2012.

[20] S. Lazebnik, C. Schmid, J. Ponce, "A Sparse Texture Representation Using Local Affine Regions", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.27(8), pp.1265-1278, 2005.

[21] T. Ojala, M. Pietikainen, T.T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with Local Binary Pattern", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24(7), pp.971-987, 2002.

[22] G. Zhao, M. Pietikainen, "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.29(6), pp.915-928, 2007.

[23] T. Ahonen, J. Matas, C. He, M. Pietikainen, "Rotation Invariant Image Description with Local Binary Pattern Histogram Fourier Features", Lecture Notes in Computer Science, Image Analysis, pp.61-70, 2009.

[24] Y. Guo, G. Zhao, M. Pietikainen, Z. Xu, "Descriptor Learning Based on Fisher Separation Criterion for Texture Classifcation", Lecture Notes in Computer Science, 2010 Asian Conference on Computer Vision, pp.185-198, 2010.

[25] Heikkilä, Marko and Pietikäinen, Matti and Heikkilä, Janne, "A texture-based method for detecting moving objects", BMVC, pp.1-10, 2004

[26] G. Zhao, T. Ahonen, J. Matas, M. Pietikinen, "Rotation-invariant image and video description with local binary pattern features". IEEE Transactions on Image Processing, vol.21(4),pp.1465-1467, 2012.

[27] S. Liao , M.W.K. Law, A.C.S. Chung, "Dominant Local Binary Patterns for Texture Classification", IEEE Transactions on Image Processing, vol.18(5), pp.1107-1118, 2009.

[28] Z. Guo, L. Zhang , D. Zhang, "A Completed Modeling of Local Binary Pattern Operator for Texture Classification", IEEE Transactions on Image Processing. vol.19(6), pp.1657-1663, 2010.

[29] Z. Guo, L. Zhang, D. Zhang, "Rotation invariant texture classification using LBP variance(LBPV) with global matching", Pattern Recognition, vol.43(3), pp.706-719, 2010 .

[30] Z.B. Zhang , S.L. Ma, D.Y. Wu, "The application of neural network and wavelet in human face illumination compensation", Proc. Advances in Neural Networks, pp.828-835, 2005.

[31] J.J. Pan, D. Zhang, Y.Y. Tang, "A fractal-based BEMD method for image texture analysis", 2010 IEEE International Conference on Systems, Man, and Cybernetics, Turkey, October, pp.3817-3820, 2010 .

[32] J.J. Pan, Y.Y. Tang, "Texture analysis based on saddle points-based BEMD and LBP", Lecture Notes in Computer Science, Computer Analysis of Images and Patterns, pp.506-514, 2011.

[33] J.J. Pan, Y.Y. Tang, "Extremum Mean Empirical Mode Decomposition", 2012 International Congress on Image and Signal Processing, pp.1888-1893 , Chongqing, China, 2012

[34] F.M. Khellah, "Texture Classification Using Dominant Neighborhood Structure", IEEE Transactions on Image Processing, vol.20(11), pp.3270-3279, 2011.

[35] L. Zhang, L. Zhang, Z. Guo, D. Zhang, "Monogenic-LBP: A new approach for rotation invariant texture classification", 2010 IEEE International Conference on Image Processing, pp.2677-2680, 2010.

[36] B. Caputo, E. Hayman, P. Mallikarjuna, "Class-Specific Material Categorisation", 2005 IEEE International Conference on Computer Vision, pp.1597-1604, 2005.

[37] K.J. Dana, S.K. Nayar, B. van Ginneken, J.J. Koenderink, "Reflectance and texture of real-world surfaces", ACM Transactions on Graphics, Vol.18(1), pp.295-306,, Jan. 1999.