Computer Vision and Image Understanding 150 (2016) 58–65

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

A novel scheme based on local binary pattern for dynamic texture recognition

Deepshikha Tiwari, Vipin Tyagi*

Jaypee University of Engineering and Technology, Raghogarh, Guna (MP), INDIA

ARTICLE INFO

Article history:
Received 6 October 2015
Revised 22 April 2016
Accepted 23 April 2016
Available online 25 April 2016

Keywords:
Dynamic texture
Local binary pattern
Michelson contrast

ABSTRACT

Dynamic textures (DTs) are moving sequences of natural scenes with some form of temporal regularity such as boiling water, a flag fluttering in the wind. The motion causes continuous changes in the geometry of dynamic textures thus it is difficult to apply traditional vision algorithms to recognize this class of textures. This paper proposes a scheme for modeling and classification of dynamic textures using a local image descriptor which efficiently encodes texture information in a space-time domain. The proposed descriptor extends the well-known spatial texture descriptor, local binary pattern (LBP), to spatio-temporal domain in order to represent the DT by combining appearance feature with the motion. Although, local binary patterns are used extensively in visual recognition applications due to their excellent performance and computational simplicity, but sometimes unable to differentiate different structures properly due to their dependency on center pixel as a threshold. In this paper, a new descriptor based on a global adaptive threshold is employed to compute the structure pattern of local image patch which differentiates various local image structures more efficiently. However, the LBP pattern defines the spatial structure of a local image patch but it does not give information about the contrast of local image patch. We have used Michelson contrast to compute the difference in luminance in the local texture and clubbed with local structure pattern computed using the proposed descriptor. Extensive experiments on dynamic texture databases (Dyntex, Dyntex++ and UCLA) prove the efficiency of the proposed method.

1. Introduction

Dynamic, or temporal, texture is a moving sequence of images that exhibit spatially repetitive and time-varying visual patterns. In dynamic texture (DT), the notion of self-similarity central to static image texture is extended to the spatio-temporal domain (Chetverikov and Péteri, 2005). DTs are typically videos of processes, such as waves, smoke, fire, a flag blowing in the wind or a moving escalator. DTs have gained much attention in the field of computer vision due to their usage in many areas such as surveillance applications like the detection of fire or smoke (Toreyin et al., 2006; Ye et al., 2015), environmental monitoring (Ali et al., 2008), traffic monitoring (Chan and Vasconcelos, 2005), crowd analysis and management (Chan et al., 2007), face recognition (Koelstra et al., 2010; Zhao and Pietikäinen, 2007).

Various approaches have been used in the past for the recognition and representation of dynamic texture. However, in early years of DT recognition research, most studies used optical flow as a tool to represent DTs (Chetverikov and Péteri, 2005). Later, Soatto and Doretto (Doretto et al., 2003; Soatto et al., 2001) published a work on dynamic texture representation based on linear dynamic systems (LDS) model which inspired various researchers to use LDS model to describe and process DTs. Saisan et al. (Saisan et al., 2001) used the LDS model (Soatto et al., 2001) to offer auto-regressive LDS (AP-LDS) approach to recognize temporal. Chan et al. (Chan and Vasconcelos, 2007) used kernel PCA to model a wide range of DTs with chaotic motion (e.g. turbulent water). Their framework used kernel dynamic texture with martin distance (KDT-MD) to classify temporal texture exhibits different motions. Ravichandran et al. (Ravichandran et al., 2009) used a bag of word (BoW) and LDS to represent DTs for view-invariant DT recognition. Other approaches (Wang and Hu, 2015; Wang and Hu) also used BoW approach along with chaotic vector to recognise DTs. Besides LDS model, fractal analysis is also employed by various studies such as dynamic fractal spectrum (DFS) (Xu et al., 2011), 3D oriented transform feature(3D-OTF) (Xu et al., 2012), wavelet domain multifractal spectrum (WMFS) (Ji et al., 2013) to represent the DTs. These methods used fractal dimension to compute the feature descriptor which shows robustness to environmental and view changes.

Recently, subspace analysis has also been used to model and classify DTs. Baktashmotlagh et al. (Baktashmotlagh et al., 2014; Baktashmotlagh et al., 2013) used the subspace analysis to extract

* Corresponding author.
E-mail address: dr.vipin.tyagi@gmail.com (V. Tyagi).

http://dx.doi.org/10.1016/j.cviu.2016.04.010
1077-3142/

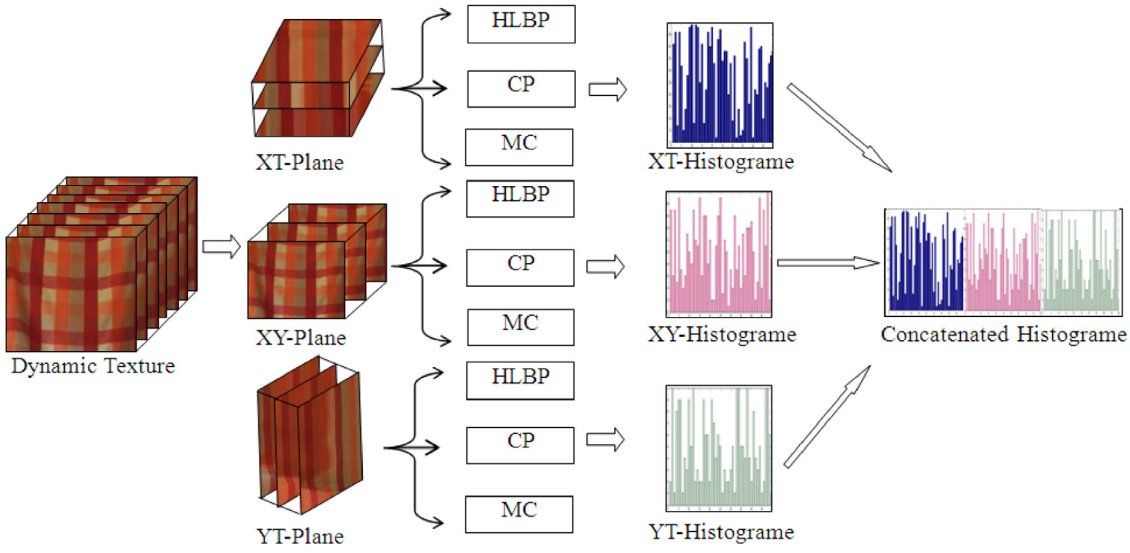

Fig. 1. Illustration of the computation of the proposed descriptor.

the stationary part of the video signal from its non-stationary part. The stationary part is used to offer low dimensional feature descriptors named kernel stationary subspace analysis (KSSA), nonlinear stationary subspace analysis (NLSSA) and discriminative kernel stationary subspace analysis (DKSSA), to describe the temporal texture. Author argued that the stationary signal is better representative of the class and thus can be used to devise a low dimensional feature descriptor. Other approaches such as filter based methods (Arashloo and Kittler, 2014; Rivera and Chae, 2015) are equally used to describe and recognize dynamic texture. Rivera et al. (Rivera and Chae, 2015) presented a new descriptor named Directional Number Transitional Graph (DNG) based on new 3D filter to create the signature for a dynamic texture. Arashloo et al. (Arashloo and Kittler, 2014) produced multiscale binary statistical feature descriptor (MBSIF) by binarizing the responses of a set of filters operating on local image patches at multiple resolutions.

Many sparse and dense static texture descriptors like local binary pattern (LBP) (Ojala et al., 1996; Ojala et al., 2002) binarized statistical image features (BSIF) (Kannala and Rahtu, 2012), scale invariant feature transform (Lowe, 2004), Gabor (Manjunathi and Ma, 1996) have been extended to spatio-temporal domain for dynamic texture recognition (Zhao and Pietikäinen, 2007; Arashloo and Kittler, 2014; Xu et al., 2012; Gonçalves et al., 2012; Tiwari and Tyagi, 2015). Among static texture descriptors, LBP and its variants (Ojala et al., 2002; Ming et al., 2015; Qi et al., 2015) have gained much popularity in the field of image processing and computer vision due to its advantages like computational simplicity, gray scale invariance and no requirement of pre-training. Zhao et al. (Zhao and Pietikäinen, 2007) proposed volume local binary pattern (VLBP) and local binary pattern in three orthogonal planes (LBP-TOP) (Zhao and Pietikäinen, 2007) to extend LBP to spatio-temporal domain for the dynamic texture recognition. In (Ghanem and Ahuja, 2010), LBP is combined with pyramid of histograms of oriented gradients (PHOG) and linear dynamical system (LDS) to represent and classify dynamic texture. Ren et al. (Ren et al., 2013) used principal histogram analysis (PHA) with LBP to tackle the reliability issues of LBP histogram. To reduce the dimensionality of the VLBP feature vector, Ren et al. (Ren et al., 2014) offered data driven LBP (DDLBP) to learn the optimized LBP structure. DDLBP approach used the point selection method and maximal joint mutual information scheme (MJMI) to reduce the dimensionality of VLBP. Apart from long histogram, another shortcoming of VLBP approach is use of only sign information. VLBP does not include magnitude of local difference in feature descriptor which represents the contrast of local image patch. Tiwari and Tyagi (2016) argued that the inclusion of contrast information improves the performance of VLBP approach and suggested completed volume local binary pattern (CVLBP) descriptor to classify dynamic texture.

In VLBP, the size of the feature descriptor depends on P neighboring pixels as it uses $3P+2$ pixels along the helix to create the feature vector. As the value of P increases, the number of patterns for basic VLBP becomes very large (2^{3P+2}). Due to this, it is difficult to extend the VLBP to have large neighboring set. To overcome this, Zhao et al. (Zhao and Pietikäinen, 2007) proposed local binary pattern in three orthogonal planes (LBP-TOP). LBP-TOP used concatenated local binary patterns from three orthogonal planes to reduce the size of the feature vector. However, VLBP produces comparatively good results with small neighboring points as compared to LBP-TOP. LBP-TOP loses some important information because it does not include the center pixel in the feature vector, whereas VLBP uses the center pixels of neighboring frames in feature vector. Besides, both VLBP and LBP-TOP suffer with the limitation of the basic local binary pattern and are not able to discriminate coarse and smooth textures properly.

To address these problems, we propose a new method to extend the local binary pattern for the dynamic texture domain that contains the advantageous characteristics of both LBP-TOP and VLBP. Our contributions are threefold:

(a) The size of the VLBP feature vector is reduced as a new formulation is applied for feature vector computation.
(b) Center pixel is added to the feature vector to improve its discriminating power.
(c) Michelson contrast is used to compute the amount of local texture and clubbed with local structure pattern computed using the LBP.

2. Proposed technique

2.1. Overview

The computation procedure of proposed descriptor is illustrated in Fig. 1. It consists of two steps: local texture pattern computation and local texture contrast computation. To compute the local texture pattern, first the global mean absolute difference (GMAD) is computed for three parallel frames in a plane. GMAD is provided as

input for computation of helix local binary pattern (HLBP) feature vector and center pixel value. Along with the local texture pattern, Michelson Contrast (MC) is also computed to define local texture contrast. To characterize the feature descriptor, a histogram is created that uses the HLBP pattern along with center pixel (CP) as an index of the histogram and MC as the contribution into the bin corresponding to index. To add the flavor of LBP-TOP, feature descriptor is computed for all three planes, i.e. XY-plane, XT-plane, YT-plane and then concatenated.

2.2. The proposed HLBP

A local binary pattern is sometimes not able to differentiate different structure patterns as it only uses the sign of local difference between center pixel and neighborhood to compute the feature vector. For example, the difference vectors [3, 9, −13, −16, −15, 74, 39, 31] and [150, 1, −150, −1, −100, 150, 1, 150] have the same sign vector [−1, −1, −1, 1, 1, 1, 1, 1]. However, it is hard to say that they have similar local structures. LBP only uses sign information to compute gray scale invariant texture descriptor. In addition, LBP uses local reference value, i.e. center pixel as threshold and does not take advantage of global information from the whole image.

To mitigate this, we propose a new threshold that takes the advantage of both local and global information and helps to discriminate different textures properly. Global information part of the threshold is computed by taking the mean absolute deviation of local differences (*LD*) of the whole image. Center pixel works as local information part of the threshold. The threshold Th_t is defined as:

$$Th_t = g_{t_c,C} + GMAD_t \tag{1}$$

$$GMAD_t = \frac{1}{(H-2)\times W-2)\times P}\sum_{i=1}^{H-2}\sum_{j=1}^{W-2}\sum_{n=0}^{P-1} |LD_t(i,j,n) - mean(LD_t(i,j,n)| \tag{2}$$

$$LD_t(i,j,n) = (g_{t,n} - g_{M,C}) \tag{3}$$

where $g_{M,C}$ represents the gray value of the center pixel of the center frame, $g_{t,n}(t = F, M, B; n = 0 \ldots \ldots P-1)$ corresponds to the gray values of P equally spaced pixels in the previous *(F)*, current *(M)* and posterior *(B)* frames, *LD (I, j, n)* is the local difference at the ith row, jth column and nth pixel of the image of size $H \times W$ and *GMAD* is global mean absolute deviation. Mean absolute deviation is a robust statistical dispersion measure that shows more resilient to outliers in a data set than the mean. Hence, it is used in *GMAD* computation.

To compute the HLBP, the threshold Th_t is subtracted from the gray values of the circularly symmetric neighborhood $g_{t,n}$ (n = 0,..., $P-1$) of previous *(F)*, current *(M)* and posterior*(B)* frames:

$$V_t = v(g_{t,0} - Th_t, g_{t,1} - Th_t, \ldots \ldots, g_{t,P-1} - Th_t) \tag{4}$$

where V_t represents the texture in the local neighborhood of a specific frame and v is the joint distribution of P local differences. The sign of local difference $g_{t,n} - Th_t$ is robust to change in mean luminance, therefore we consider the sign to make joint distribution invariant to gray scale shifts.

$$V_t = v(s(g_{t,0} - Th_t), s(g_{t,1} - Th_t), \ldots \ldots, s(g_{t,P-1} - Th_t)) \tag{5}$$

where

$$s(d) = \begin{cases} 1, & d \geq 0 \\ 0, & d < 0 \end{cases}$$

Each *s(d)* is multiplied by a binomial factor 2^n and then summed to define HLBP pattern that uniquely characterizes the spatial structure of the local image patch in a dynamic texture.

$$HLBP_{t,P} = \sum_{n=0}^{P-1} s(g_{t,n} - Th_t)2^n \tag{6}$$

HLBP also considers the sign of local differences to compute the feature vector, still it is able to differentiate different structures properly because the threshold Th_t consider both local and global information and does not solely depend on the single reference value as in LBP. Along with HLBP, center pixel information is also computed to boost the discriminative power of feature descriptor. It is defined as:

$$CP_t = s(g_{t,C} - Th_t) \tag{7}$$

where $g_{t,C}$ is the center pixel of local neighborhood of frame t and $s(x)$ defined as in Eq. (5). Both HLBP pattern and CP are used as an index in the histogram of specific plan.

2.3. The Michelson contrast

An image texture is made up of two things: spatial structure and the contrast. HLBP defines the spatial structure of the image texture. It does not provide any information about the contrast of the texture. Therefore, contrast is clubbed with HLBP to complete the definition of texture. To define the contrast measure, we have used the classic definition of Michelson contrast (Michelson, 1927). For simple periodic patterns (e.g., textures), there is no large area of uniform luminance that dominates the user's brightness adaptation. Michelson contrast, assumes that the viewer is adapted to the sum of the background and foreground, and therefore is well-suited to predicting the contrast of grids and other periodic patterns such as texture. It is also much more robust under the addition of ambient illumination. To compute the MC, maximum and minimum value of local neighborhood is computed and then a ratio is computed between the sum and difference of maximum and minimum values. It is defined as:

$$MC = \frac{\max(g_{t,n} - Th_t) - \min(g_{t,n} - Th_t)}{\max(g_{t,n} - Th_t) + min(g_{t,n} - Th_t)} \tag{8}$$

2.4. Deriving the feature descriptor

To define the dynamic texture feature descriptor; HLBP, CP and MC codes are computed for all the frames of an image sequence in each plane. The discrete occurrence histogram of the resulting code sequences for dynamic texture of size $H \times W \times L$; $x \in \{0, 1, \ldots \ldots H-1\}$, $y \in \{0, 1, \ldots \ldots W-1\}$, $f \in \{0, 1, \ldots \ldots L-1\}$ in each plane is computed by :

$$h = \left[h_{0,0}, h_{1,0}, \ldots \ldots \ldots \ldots \ldots, h_{b-1,0}, h_{0,1}, h_{1,1}, \ldots \ldots \ldots \ldots \ldots, h_{b-1,1}\right] \tag{9}$$

$$h_{i,j} = \sum_{x,y,f} MC(x,y,f) \times I_A\{HLPB(x,y,f) = i\} \times I_A\{CP(x,y,f) = j\} \quad i = 0, 1, \ldots \ldots \ldots b-1;\ j = 0, 1 \tag{10}$$

where b is the number of histogram bins produced by HLBP pattern corresponding to CP code, I_A Is the indicator function equal to one, when its argument is true and zero otherwise.

The size of the resulting feature descriptor is $2 \times b$, where $b = 2^P$. The larger the variation in pixels contrast in a local patch, the higher the MC value assigned to the pixels HLBP code. In this way, we can accumulate the contrast of all similar patterns in a single bin. This information helps to take into account whether

the pattern in the local area is of strong contrast or weak contrast. Thus, the resulting feature will contain both contrast and texture information of a local neighborhood in a single representation. When different temporal/spatial size dynamic textures are used for comparison, the histogram is normalized to get a coherent description.

$$NH_{i,j} = \frac{h_{i,j}}{\sum_{i=0,j=0}^{b-1,1} h_{i,j}} \tag{11}$$

Once the histograms are acquired on XY, XT and YT planes, three histograms are concatenated to form the final spatio-temporal dynamic texture descriptor for the image sequence as:

$$NH = \left[\mathrm{NH}_{\mathrm{i,j}}^{\mathrm{XT}},\ \mathrm{NH}_{\mathrm{i,j}}^{\mathrm{XY}},\ NH_{i,j}^{YT}\right] \tag{12}$$

Proposed approach is considerably different from LBP and its extensions to spatiotemporal domain, i.e., VLBP and LBP-TOP. LBP is a spatial texture descriptor. However, the proposed scheme offer a spatiotemporal texture descriptor composed of contrast (i.e. MC) and texture feature (i.e. HLBP). As in LBP, the HLBP feature vector is computed by subtracting the threshold value form local neighboring pixel. However, HLBP uses the statistics of whole frame to define an adaptive global threshold whereas the traditional LBP uses gray value of center as a threshold to compute the feature descriptor.

In addition to texture feature, the proposed spatio-temporal descriptor uses contrast features of image sequences on three orthogonal planes, have different levels of locality. First, it views the dynamic texture as sets of volumes and capture feature on the basis of volume textons. This is achieved via computing the texture pattern and contrast value from three parallel frames. This information helps to get appearance information of DT. Distribution of codes in each plane is represented via plane-specific histograms. At the next level, dynamic texture is considered as a stack of XY planes in axis T, as a stack of XT planes in axis Y and as a stack of YT planes in axis X to capture the motion information. The histograms from different planes are concatenated to build a global description of DT with the spatial and temporal features. In contrast, VLBP used only parallel frames to compute the feature descriptor and LBP-TOP used only orthogonal frames to compute feature descriptor. Both VLBP and LBP-TOP have directly used center pixel value as threshold and do not include any global statistic in threshold thus suffered with the same LBP issues.

Proposed approach uses both local and global information in threshold computation thus building robust featuring descriptor. In addition, it uses a different weighting scheme as compared to uniform weighting scheme of VLBP and LBP-TOP.

3. Experimental evaluation

In this section, the proposed method is evaluated on various dynamic texture databases for classification.

3.1. UCLA database

UCLA database (Doretto et al., 2003; Chan and Vasconcelos, 2005; Saisan et al., 2001) is a benchmark in the field of dynamic texture recognition. It consists of 200 dynamic texture sequences categorized into 50 different classes. Each class contains four DT sequences and each DT sequence compromise 75 frames of 160 × 110 pixels. These include boiling water, fountains, fire, waterfalls, plants and flowers swaying in the wind. Each DT sequence is clipped to a 48 × 48 window to capture the key statistical and dynamical features. The results of VLBP, LBP-TOP and CVLBP are from our own implementation while the results of remaining approaches are from the original literatures.

Table 1
Results (%) on 50-class UCLA database in leave-one-out scheme.

Method	Classification rate (%)
AR-LDS (Saisan et al., 2001)	89.9
Spacetime orientation structure (Derpanis and Wildes, 2012)	81.0
Proposed method	95.0

(*Note:* all the results using the 1-NN classifier).

Table 2
Results (%) on 50-class UCLA database in four cross-fold scheme.

Method	Classification rate (%)
VLBP (Zhao and Pietikäinen, 2007)	89.5
KDT-MD (Chan and Vasconcelos, 2007)	89.5
LBP-TOP (Zhao and Pietikäinen, 2007)	94.5
DFS (Xu et al., 2011)	89.5
3D-OTF (Xu et al., 2012)	87.10
CVLBP (Tiwari and Tyagi, 2016)	93.00
Proposed method	95.00

(*Note:* All the results using the 1-NN classifier).

50-class breakdown: In this scenario, previous studies have applied two types of classification schemes that used different portions of the database as training and test data as follows:

Leave-one-out scheme: Similar to the experimental setup in (Saisan et al., 2001) and (Derpanis and Wildes, 2012), leave-one-out classification procedure is followed in the evaluation of the proposed method. A test sample is classified correctly, if one of the three remaining sequences of the same scene is its nearest neighbor. A comparison of the proposed descriptor to the methods of (Saisan et al., 2001) and (Derpanis and Wildes, 2012) using L-1 distance as dissimilarity measure is presented in Table 1.

It can be observed that the proposed method achieves the best recognition performance among other competitors by a large margin. In (Derpanis and Wildes, 2012), authors modeled DT by primarily using the dynamics of spatiotemporal texture whereas approach (Saisan et al., 2001) describes the DTs using the joint photometric-dynamic model. Results of proposed approach and (Arashloo and Kittler, 2014) show that the representation of DTs using both appearance and motion feature is more beneficial for view invariant dynamic texture recognition. However, the classification performance of proposed approach shows its efficiency in combining motion features with appearance.

In (Arashloo and Kittler, 2014), the authors have reported the classification rate of 99.5% using MBSIF feature descriptor. MBSIF is a multiresolution feature descriptor which incorporates the information from different resolutions into one representation to achieve better performance. However, with single resolution the performance of MBSIF has been dropped to 93%. The proposed scheme offers a single resolution descriptor thus we compared single resolution MBSIF descriptor with proposed descriptor and achieved batter classification performance with a margin of 2%.

Four cross-fold classification scheme: In studies (Ghanem and Ahuja, 2010; Chan and Vasconcelos, 2007; Xu et al., 2011), a different division of the database is used as training set and test set. In this scheme, one sequence from each class is used for testing and remaining three from the same class for training. The test is repeated four times, each time using a different sequence as the test sample. Finally, the average recognition rate of four trials is used as results. A comparison of the proposed approach with other approaches is shown in Table 2.

It is evident that the proposed method outperforms other techniques. Due to rapid growth in feature vector size, it is difficult to extend VLBP (Zhao and Pietikäinen, 2007) and CVLBP (Tiwari and Tyagi, 2016) to a large neighboring pixels such as $P > 4$. However,

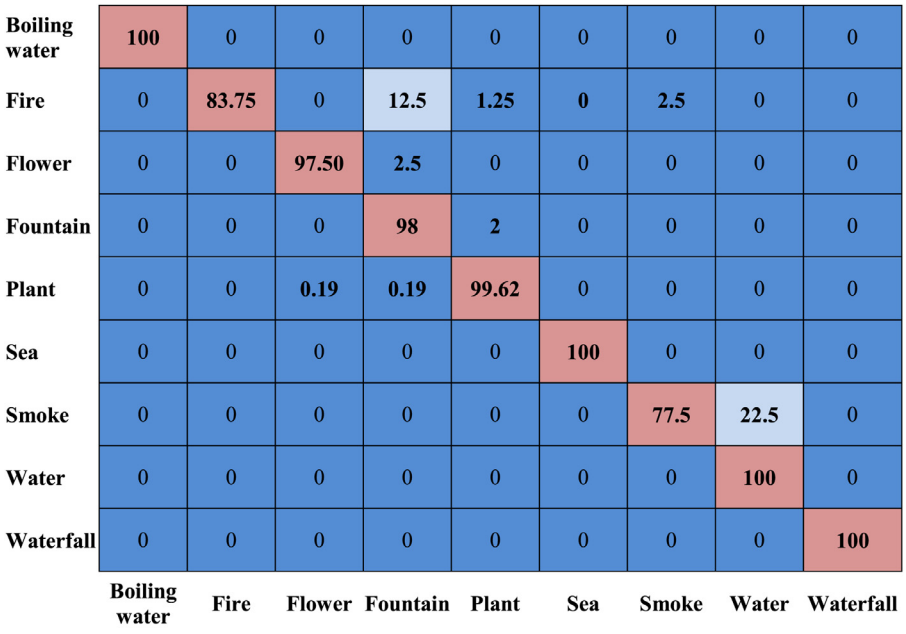

Fig. 2. Confusion matrix (%) of 9-class UCLA database.

Table 3
Results (%) on 9-class UCLA database.

Method	Classification rate (%)
VLBP (Zhao and Pietikäinen, 2007)	96.30
LBP-TOP (Zhao and Pietikäinen, 2007)	96.00
3D-OTF (Xu et al., 2012)	96.32
WMFS (Ji et al., 2013)	96.95
MBSIF (Arashloo and Kittler, 2014)	98.75
DNGP (Rivera and Chae, 2015)	98.1
CVLBP (Tiwari and Tyagi, 2016)	96.90
Chaotic vector (Wang and Hu)	85.1
High level feature (Wang and Hu, 2015)	92.67
Proposed method	98.35

(*Note:* All the results using the 1-NN classifier).

the proposed scheme can easily be extended to large neighboring pixel and provides batter classification accuracy as compared to the approaches in (Zhao and Pietikäinen, 2007; Tiwari and Tyagi, 2016). The proposed scheme also outperforms fractal based approaches (Xu et al., 2012, 2011), thus showing the high descriptive capabilities of LBP based approaches.

9-class breakdown: In (Ravichandran et al., 2009), UCLA database is reorganized to combine the sequences that are taken from different viewpoints. This results in reduced database with classes being boiling water (8), fire (8), flower (12), fountains (20), plants (108), sea (12), smoke (4), water (12) and waterfall (16). Here the numbers in parentheses represent the number of sequences in each class. We have used the same experimental setup as in (Ghanem and Ahuja, 2010). For each trial, 50 sequences are randomly selected from each class for training, and the rest of the 50 for testing. In this case, a correct classification is defined as assigning the test sequence into one of the training sequences from the same class based on a nearest neighbor rule.

The experiment is repeated 20 times and the average performance is provided in Table 3. As observed from the table, the proposed scheme outperforms the spatiotemporal LBP methods i.e., VLBP, LBP-TOP, CVLBP with approximately 2% margin. It demonstrates same difference in classification accuracy against fractal based approaches such as, 3D-OTF, WMFS. However, the filter based approaches such as DNGP and MBSIF perform nearly same as proposed scheme.

Although, the MBSIF approach shows an improvement of 0.40% over the proposed scheme, the size of MBSIF operator (i.e. 6144) is much higher as compared to the proposed scheme (i.e. 1536) which *offset* its improvement over proposed scheme. The size of feature vector is very important in real time applications like surveillance applications, where time is crucial. If the size of feature vector is high, then time taken by the algorithm for classification, will be high. The smaller size of feature vector will help in getting classification results in lesser time.

The proposed method achieves impressive 98.35% correct classification rate with a simple distance metric, i.e., L1 distance against other state-of-art methods. The confusion matrix of 9-class breakdown is shown in Fig. 2.

8-class breakdown: It is observed that in UCLA database, number of sequences of plant class far outnumbered the number of sequences for the other classes (Ravichandran et al., 2009). Therefore, the plant class is discarded and remaining eight classes are used for classification. For 8-class breakdown, same experimental setup as in (Ghanem and Ahuja, 2010; Xu et al., 2011) is adopted for classification. Each class is randomly bisected; half of the database is used for training and the remaining half for testing. As in 9-class breakdown, a correct classification is defined as assigning the test sequence into one of the training sequences from the same class based on a nearest neighbor rule.

Confusion matrix for this test is shown in Fig. 3. Confusion matrix shows the detailed performance of the proposed approach for each class. It can be observed from the confusion matrix that proposed method majorly confused smoke sequence with water sequence and fire sequence with fountain and sea. This happens because both the sequences may have very similar appearance/dynamic characteristics.

A comparison of the classification rate of the proposed method against other state-of-art methods is summarized in Table 4. It is evident from the table that proposed approach shows striking classification accuracy as compared to the baseline spatiotemporal approaches i.e., LBP-TOP and VLBP.

Our proposed method achieves impressive performance despite using a simple nearest neighbor classifier which emphasizes the discriminatory power of the proposed descriptor. The motivation behind our choice of simple NN-classifier is the desire to evaluate

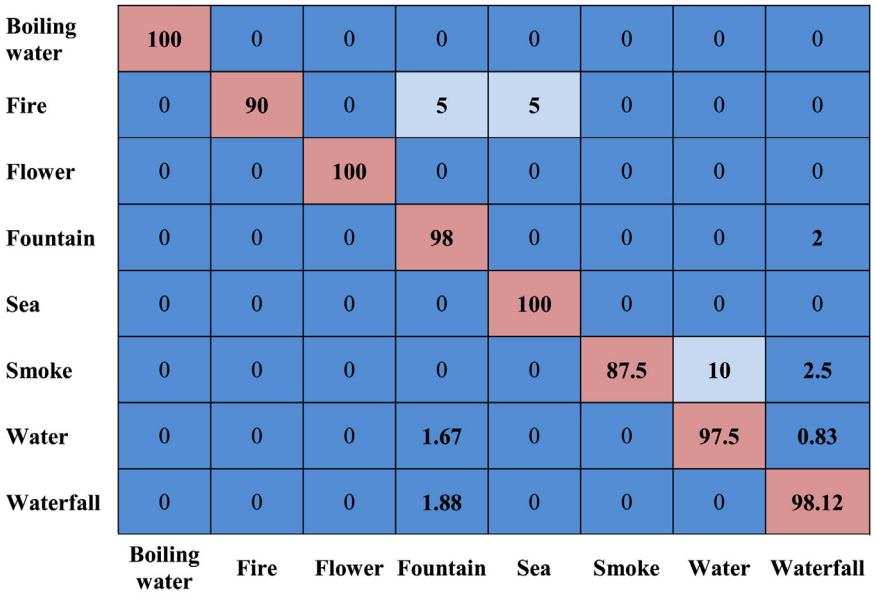

Fig. 3. Confusion matrix (%) of 8-class UCLA database.

Table 4
Results (%) on 8-class UCLA database.

Method	Classification rate (%)
AR-LDS (Saisan et al., 2001)	54.12
VLBP (Zhao and Pietikäinen, 2007)	91.96
LBP-TOP (Zhao and Pietikäinen, 2007)	93.67
NLDR (Ravichandran et al., 2009)	70.00
3D-OTF (Xu et al., 2012)	95.80
WMFS (Ji et al., 2013)	97.18
MBSIF (Arashloo and Kittler, 2014)	97.8
DNGP (Rivera and Chae, 2015)	97.0
CVLBP (Tiwari and Tyagi, 2016)	95.65
Chaotic vector (Wang and Hu)	85.00
high level feature (Wang and Hu, 2015)	85.65
Proposed method	97.50

(*Note:* All the results using the 1-NN classifier).

Table 5
Results (%) on Dyntex++ database.

Method	Classification rate (%)
VLBP (Zhao and Pietikäinen, 2007)	94.98^{N}
LBP-TOP (Zhao and Pietikäinen, 2007)	94.05^{N}
DL-PEGASOS (Ghanem and Ahuja, 2010)	63.7^{S}
PCA-cLBP/PI-LBP/PD-LBP (Ren et al., 2013)	91.9^{S}
DFS (Xu et al., 2011)	89.9^{S}
3D-ODT (Xu et al., 2012)	89.17^{S}
DDLBP with MJMI (Ren et al., 2014)	95.8^{S}
NLSSA (Baktashmotlagh et al., 2014)	92.4^{S}
KSSA (Baktashmotlagh et al., 2014)	92.2^{S}
DKSSA (Baktashmotlagh et al., 2014)	91.1^{S}
MBSIF (Arashloo and Kittler, 2014)	97.12^{N}
DNGP (Rivera and Chae, 2015)	90.2^{N}
Chaotic vector (Wang and Hu)	69^{N}
high level feature (Wang and Hu, 2015)	64.22^{N}
Proposed method	96.28^{N}

(*Note:* Superscripts "S" and "N" are for results using the SVM and 1-NN classifier).

the utility of the proposed approach without confounding performance with classifier sophistication.

3.2. Dyntex++ database

Dyntex++ database (Ghanem and Ahuja, 2010) is a compiled version of Dyntex database (Péteri et al., 2010). In this database, DT sequences have been pre-processed (e.g. cropped) from their raw form to show its representative dynamics in the absence of any static or dynamic background. Only a single dynamic texture is present in each DT sequence and ground truth label is assigned to each sequence in order to make a benchmark database compatible with UCLA database.

This database contains only 345 video sequences out of 656 video sequences of Dyntex database. These 345 video sequences are processed and organized into 36 classes with 100 sequences in each class. Following the experimental setup as in (Ghanem and Ahuja, 2010), each class of the database is randomly divided into two halves, training is done by using one half and the other half is used for the testing purpose. A test sample is assumed to be correctly classified if it is the nearest neighbor of one of the 50 training samples of the same class. The experiments are repeated 10 times. The results of VLBP, LBP-TOP methods are from our own implementation while the results of remaining methods are from the original literatures.

Table 5 shows the result of classification. It is evident from the table that proposed approach performs very well compared to other approaches except the MBSIF approach. MBSIF shows high classification accuracy (i.e. 97.12%) as compared to proposed approach (i.e. 96.28%). However, the proposed approach is still competitive against MBSIF approach since it produces the recognition result of 96.28 % using single scale representation and 1536 long feature descriptor, whereas the MBSIF approach incorporates the information from seven scales into one feature vector and produces a long feature descriptor of size 6144. The long feature descriptor increases the classification time complexity of MBSIF approach as compared to proposed descriptor.

Dynamic texture recognition has been used in various potential real time applications such as surveillance, environmental monitoring, face recognition and activity recognition, in such application time is very crucial factor. A high dimensional feature vector slows down the calculation of distance matrix which shows the dissimilarity of an unknown dynamic texture from known dynamic textures and thus increases the time required to classify a DT. Various studies (Zhao and Pietikäinen, 2007; Wang and Hu, 2015; Baktashmotlagh et al., 2014; Ojala et al., 2002; Ren et al., 2014) point towards the need of low dimensional feature descriptor for the

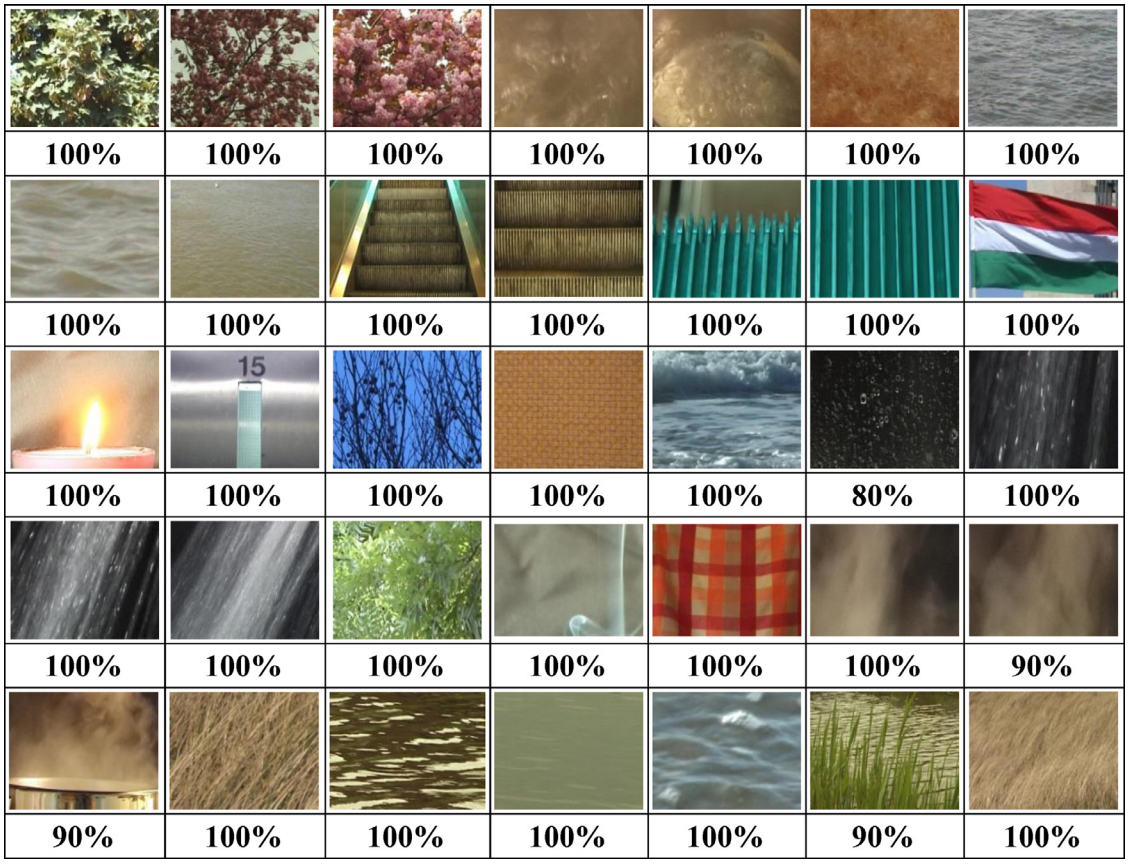

Fig. 4. Classification rate (%) of proposed approach on each class of the Dyntex database.

Table 6
Comparison of various methods with different neighborhood size.

Method	Bin Size	Classification rate (%)
$VLBP_{1,4,1}$	16,384	94.80
$LBP\text{-}TOP_{1,4,1}$	3×16	91.49
$LBP\text{-}TOP_{1,8,1}$	3×256	94.05
Proposed method $_{1,4,1}$	3×32	94.45
Proposed method $_{1,8,1}$	3×512	96.28

success of many video classification algorithms. Hence a minor improvement in classification accuracy is compromised with low dimensional feature descriptor.

Proposed approach also shows impressive performance using a simple NN-classifier against DDLBP method (Ren et al., 2014) which attempted to reduce the size of VLBP feature descriptor using point selection approach and used more complex SVM classifier for classification purpose. In addition to this, proposed approach has also performed better then the low dimensional descriptors based on subspace analysis such as NLSSA, KSSA and DKSSA. It shows that both appearance and motion aspects of DT contribute useful information in DT representation as proposed scheme used both features of DTs in contrast to subspace analysis approaches which are primarily based on stationary part of DTs.

We have also evaluated the performance of the proposed approach on different neighborhood size as shown in Table 6. It also includes the performance comparison of the proposed approach against two state-of-the-art LBP extensions, i.e. VLBP and LBP-TOP, to dynamic texture.

It can be observed from the table that with only four neighbors, VLBP shows considerably better results than LBP-TOP when the histogram from individual planes are linearly combined. The proposed approach is performing comparatively same as VLBP with reduced feature descriptor size and better than LBP-TOP on the same neighborhood size. It is also interesting to note that only with four neighbors proposed descriptor is working well than eight neighbors LBP-TOP.

3.3. Dyntex database

Dyntex database (Péteri et al., 2010) is a large and diverse collection of dynamic texture sequences. It contains the dynamic texture videos from various fields ranging from struggling flames to whelming waves, from calm water to boiling water, from swaying branches to moving escalators. The sequences in the database are captured under different environmental condition involving cropping, deinterlacing with spatio-temporal median filter. Each sequence consists of 250 color frames of 400 $\times$ 300 pixels.

For the evaluation of the proposed approach, we have used a version of Dyntex database that contains 35 DTs. Similar to the experimental setup in (Arashloo and Kittler, 2014), each DT sequence is first divided into eight non-overlapping subsequences, but not half in X, Y and T plane. The cutting position in volume is selected randomly. In addition, the sequences with the original size, but cut only in time direction are also included in the experiment. Finally, each DT sequence creates 10 samples of different dimensions with same class label. Each DT is treated as a class and all samples used in the classification task. This sampling method makes the recognition task more challenging.

The leave-one-group-out scheme is used to conduct the experiment, i.e., one subsequence from each class is used to form the test sample and rest of the subsequences are used as training data. In our experiments, each class is represented by all the feature

Table 7
Results (%) on Dyntex database.

Method	Classification rate (%)
VLBP (Zhao and Pietikäinen, 2007)	81.14
LBP-TOP (Zhao and Pietikäinen, 2007)	92.45
DFS (Xu et al., 2011)	97.63
MBSIF (Arashloo and Kittler, 2014)	98.61
CVLBP (Tiwari and Tyagi, 2016)	85.14
Proposed method	98.57

vectors of the samples in the training set. Each test sample is assumed to be classified correctly if it has the smallest distance to training sample of the same class in the feature space.

Table 7 summarizes the classification performance of the proposed approach along with other LBP-based approaches.

It can be observed from Table 7 that the proposed method performs very well compared to other approaches. Proposed descriptor achieved an impressive performance of 98.57% correct classification rate on this database, improving the previous results of state-of-the-art LBP extensions to spatio-temporal domain, i.e. VLBP and LBP-TOP by more than 17% and 7% respectively. We have also reported the classification rate within the each class in Fig. 4. We have computed the results of LBP - TOP, VLBP and CVLBP from our own implementation while the results of DFS and MBSIF are taken from the original literature.

4. Conclusion

In this paper, we have proposed a new method for the representation and recognition of dynamic texture using discriminative LBP structures along with Michelson contrast. Existing algorithms such as VLBP and LBP-TOP suffer with the shortcomings of LBP and also discard contrast information. This is not desired as texture and contrast both contain discriminative information.

In addition, VLBP produces a long histogram as the number neighboring pixels increases. This leads to the high computational cost of the classification procedure and makes distance matrix calculation slower. Thus, a new formulation of LBP in spatio-temporal domain is proposed to reduce the size of VLBP feature descriptor and boost the discriminative power of feature vector. The proposed approach used a new threshold to compute texture pattern and Michelson contrast to determine the contrast of local image patch. Furthermore, both texture and contrast information fused in a single one-dimensional histogram to complete DT representation. The experimental evaluations of the proposed approach on different DT databases clearly demonstrate the merits of the proposed descriptor for dynamic texture recognition compared to other alternatives. In future, the suggested framework can be tested on noisy DTs and its multiresolution extension can also be explored.

References

Ali, W., Georgsson, F., Hellstrom, T., 2008. Visual tree detection for autonomous navigation in forest environment. In: Proceedings of IEEE Intelligent Vehicles Symposium, pp. 560–565.

Arashloo, S.R., Kittler, J., 2014. Dynamic texture recognition using multiscale binarized statistical image features. IEEE Trans. Multimed. 16 (8), 2099–2109.

Baktashmotlagh, M., Harandi, M., Bigdeli, A., Lovell, B., Salzmann, M., 2013. Non–linear stationary subspace analysis with application to video classification. In: Proceedings of the 30th International Conference on Machine Learning, pp. 450–458.

Baktashmotlagh, M., Harandi, M., Lovell, B.C., Salzmann, M., 2014. Discriminative non-linear stationary subspace analysis for video classification. IEEE Trans. Pattern Anal. Machine Intell. 36 (12), 2353–2366.

Chan, A.B., Vasconcelos, N., 2005. Classification and retrieval of traffic video using auto-regressive stochastic processes. In: Proceedings of IEEE Intelligent Vehicles Symposium, pp. 771–776.

Chan, A.B., Vasconcelos, N., 2005. Probabilistic kernels for the classification of auto-regressive visual processes. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR 2005), pp. 846–851.

Chan, A., Vasconcelos, N., 2007. Classifying video with kernel dynamic textures. In: Proceedings of IEEE Computer Vision Pattern Recog., pp. 1–6.

Chan, A.B., Liang, Z.S.J., Vasconcelos, N., 2007. Privacy preserving crowd monitoring: counting people without people models or tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–7.

Chetverikov, D., Péteri, R., 2005. A brief survey of dynamic texture description and recognition. In: 4th International Conference on Computer Recognition Systems, pp. 17–26.

Derpanis, K.G., Wildes, R.P., 2012. Spacetime texture representation and recognition based on a spatiotemporal orientation analysis. IEEE Trans. Pattern Anal. Mach. Intell. 34 (6), 1193–1205.

Doretto, G., Chiuso, A., Soatto, S., Wu, Y.N., 2003. Dynamic Textures. Int. J. Comput. Vis. 51 (2), 91–109.

Ghanem, B., Ahuja, N., 2010. Maximum margin distance learning for dynamic texture recognition. In: Proceedings of European Conference on Computer Vision, pp. 223–236.

Gonçalves, W.N., Machado, B.B., Bruno, O.M., 2012. Spatiotemporal Gabor filters: a new method for dynamic texture recognition. CoRR arXiv preprint arXiv:1201.3612.

Ji, H., Yang, X., Ling, H., Xu, Y., 2013. Wavelet domain multifractal analysis for static and dynamic texture classification. IEEE Trans. Image Process. 22 (1), 286–299.

Kannala, J., Rahtu, E., 2012. BSIF: Binarized statistical image features. In: Proceedings of International Conference on Image Process., pp. 1363–1366.

Koelstra, S., Pantic, M., Patras, I., 2010. A dynamic texture-based approach to recognition of facial actions and their temporal models. IEEE Trans. Pattern Anal. Machine Intell. 32 (11), 1940–1954.

Lowe, D., 2004. Distinctive image features from scale invariant key points. Int. J. Comput. Vis. 60 (2), 91–110.

Manjunathi, B.S., Ma, W.Y., 1996. Texture features for browsing and retrieval of image data. IEEE Trans. Pattern Anal. Mach. Intell. 18 (8), 837–842.

Michelson, A.A., 1927. Studies in Optics. Univ. Chicago Press, Chicago, IL.

Ming, Y., Wang, G., Fan, C., 2015. Uniform local binary pattern based texture-edge feature for 3D human behavior recognition. Plos One 10 (5), 1–15.

Ojala, T., Pietik¨ainen, M., Harwood, D., 1996. A comparative study of texture measures with classification based on featured distributions. Pattern Recognit. 29 (1), 51–59.

Ojala, T., Pietikäinen, M., Mäenpää, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24 (7), 971–987.

Péteri, R., Fazekas, S., Huiskes, M.J., 2010. DynTex: A comprehensive database of dynamic textures. Pattern Recognit. Lett. 31 (12), 1627–1632.

Qi, X., Shen, L., Zhao, G., Li, Q., Pietikäinen, M., 2015. Globally rotation invariant multi-scale co-occurrence local binary pattern. Image Vis. Comput. 43, 16–26.

Ravichandran, A., Chaudhry, R., Vidal, R., 2009. View-invariant dynamic texture recognition using a bag of dynamical systems. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1651–1657.

Ren, J., Jiang, X.D., Yuan, J., 2013. Dynamic texture recognition using enhanced LBP features. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 2400–2404.

Ren, J., Jiang, X., Yuan, J., Wang, G., 2014. Optimizing LBP structure for visual recognition using binary quadratic programming. IEEE Signal Process. Lett. 21 (11), 1346–1350.

Rivera, A.R., Chae, O., 2015. Spatiotemporal directional number transitional graph for dynamic texture recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37 (10), 2146–2152.

Saisan, P., Doretto, G., Wu, Y., Soatto, S., 2001. Dynamic texture recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 58–63.

Soatto, S., Doretto, G., Wu, Y.N., 2001. Dynamic textures. In: IEEE International Conference on Computer Vision, pp. 439–446.

Tiwari, D., Tyagi, V., 2015. Dynamic texture recognition: a review. Adv. Intell. Syst. Comput. 434, 365–374. doi:10.1007/978-81-322-2752-6_36.

Tiwari, D., Tyagi, V., 2016. Dynamic texture recognition based on completed volume local binary pattern. Multidimensional Syst. Signal Process. 27 (2), 563–575.

Toreyin, B.U., Dedeoglu, Y., Gudukbay, U., Cetin, A.E., 2006. Computer vision based method for real-time fire and flame detection. Pattern Recognit. Lett. 27, 149–158.

Wang, Y., Hu, S., 2015. Exploiting high level feature for dynamic textures recognition. Neurocomputing 154, 217–224.

Wang, Y., Hu, S., Chaotic features for dynamic textures recognition. Soft Comput., 20 (5), 1977–1989, doi:10.1007/s00500-015-1618-4.

Xu, Y., Quan, Y., Ling, H., Ji, H., 2011. Dynamic texture classification using dynamic fractal analysis. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), pp. 1219–1226.

Xu, Y., Huang, S., Ji, H., Fermüller, C., 2012. Scale-space texture description on SIFT–like textons. Comput. Vis. Image Understand. 116, 999–1013.

Ye, W., Zhao, J., Wang, S., Wang, Y., Zhang, D., Yuan, Z., 2015. Dynamic texture based smoke detection using surfacelet transform and HMT model. Fire Safety J. 73, 91–101.

Zhao, G., Pietikäinen, M., 2007. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Machine Intell. 29 (6), 915–928.

Zhao, G., Pietikäinen, M., 2007. Dynamic texture recognition using volume local binary patterns. In: Proceedings of Workshop on Dynamical Vision WDV 2005/2006, LNCS 4358, pp. 165–177.