



Completed Local Structure Patterns on Three Orthogonal Planes for Dynamic Texture Recognition

Thanh Tuan Nguyen, Thanh Phuong Nguyen, Frédéric Bouchara

► To cite this version:

Thanh Tuan Nguyen, Thanh Phuong Nguyen, Frédéric Bouchara. Completed Local Structure Patterns on Three Orthogonal Planes for Dynamic Texture Recognition. International Conference on Image Processing Theory, Tools and Applications (IPTA 2017), Nov 2017, Montréal, Canada. hal-01786823

HAL Id: hal-01786823

<https://hal.science/hal-01786823>

Submitted on 6 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Completed Local Structure Patterns on Three Orthogonal Planes for Dynamic Texture Recognition

Thanh Tuan Nguyen, Thanh Phuong Nguyen, and Frédéric Bouchara

Université de Toulon, CNRS, LSIS, UMR 7296, 83957 La Garde, France

Aix Marseille Université, CNRS, ENSAM, LSIS, UMR 7296, 13397 Marseille, France

thanh-tuan-nguyen2@etud.univ-tln.fr, tpnguyen@univ-tln.fr, bouchara@univ-tln.fr

Abstract—Dynamic texture (DT) is a challenging problem in computer vision because of the chaotic motion of textures. We address in this paper a new dynamic texture operator by considering local structure patterns (LSP) and completed local binary patterns (CLBP) for static images in three orthogonal planes to capture spatio-temporal texture structures. Since the typical operator of local binary patterns (LBP), which uses center pixel for thresholding, has some limitations such as sensitivity to noise and near uniform regions, the proposed approach can deal with these drawbacks by using global and local texture information for adaptive thresholding and CLBP for exploiting complementary texture information in three orthogonal planes. Evaluations on different datasets of dynamic textures (UCLA, DynTex, DynTex++) show that our proposal significantly outperforms recent results in the state-of-the-art approaches.

Index Terms—Dynamic texture, Dynamic texture recognition, Local binary patterns, Local structure patterns.

I. INTRODUCTION

Dynamic texture, which is a sequence of moving textures repeated spatially in varying over time [1] such as sea-wave, smoke, foliage, fire, blowing flag, fountain, etc, is an important topic in computer vision due to different important applications in this domain: facial expressions [2], [3]; fire detection; motion tracking and analysis [4], etc.

Many methods have been introduced for representation of DT. Generally speaking, they can be roughly grouped into several categories as follows: optical-flow-based, model-based, learning-based, filter-based and local-feature-based. First, optical-flow-based approaches [5] have received considerable attention owing to their efficient computation and characterizing DTs in natural way. Second, model-based methods such as Linear Dynamical System (LDS) [1] and its extension [6] or applications [7], [8] have been widely used for estimating the dynamic texture scenes. Third, learning-based methods recently become promising approaches thanks to their significant results. Inspired by the success of deep structures in image classification, Qi et al. [9] took them into feature depiction for DT recognition. Arashloo et al. [10] built a multilayer convolutional architecture (PCANetTOP) for spatio-temporal texture description and classification in which a PCA network (PCANet) is used on each of the three orthogonal planes of a DT sequence to learn filters. Other promising methods based on dictionary learning [11], [12] are utilized to extract local DT features via kernel sparse coding which exhibits strong ability

of discrimination for classification in computer vision. Fourth, filter-based approaches [2], [13] have been also utilized for handling DT recognition. Arashloo et al. [13] characterized DT sequences with Binarized Statistical Image Features on Three Orthogonal Planes (BSIF-TOP) and its multi-resolution scheme (MBSIF-TOP). In [2], Rivera et al. extracted spatio-temporal directional numbers for each frame and divided the sequence into a 3D grid to expose a novel descriptor called a Directional Number Transitional Graph (DNG). Finally, due to the simplicity and efficiency, LBP-based variants have been widely considered in local-feature-based approaches to deal with different problems of DT. Zhao et al. [3] introduced two LBP-based operators for DT representation: VLBP for capturing spatio-temporal relations from three consecutive frames; LBP-TOP for taking into account motions from three orthogonal planes. Lately, these typical operators are extended in different works [14], [15], [16] to improve the performance of descriptor. Ghanem et al. [17] also used LBP as one component in their DT descriptor.

We address in this paper a new operator for DT representation by considering local structure patterns (LSP) combining with completed schema of LBP for static images in three orthogonal planes to capture spatio-temporal textural structures.

II. RELATED WORK

As mentioned above, the LBP operator has been widely utilized in texture representation owing to its simple and efficient computation. In this section, we briefly inspect LBP [18] and several variants in still images and dynamic textures.

A. A brief review of LBP

Ojala et al. [18] proposed Local Binary Pattern (LBP) as a binary code to present the local structures of a texture image by considering the center pixel and its P neighbors sampled by interpolation on the centered circle of radius R .

Let I denote a 2D image, the encoding of LBP is defined as follows, for each pixel \mathbf{p} :

$$\text{LBP}_{P,R}(\mathbf{p}) = \sum_{i=0}^{P-1} s(I(\mathbf{q}_i) - I(\mathbf{p}))2^i, \quad (1)$$

where the $\{q_i\}$ represents the P points sampled on the circle of center \mathbf{p} and radius R , and

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Due to the high dimensionality of the basic LBP, a mapping process is often used in practice. The most popular mapping is based on uniform patterns (LBP^{u2}) having at most 2 bit-transitions (1-0 or 0-1) in its binary chain. Its use is based on the fact that almost patterns are uniform in natural images [18]. Other important mappings are: LBP^{ri} for handling rotation invariant; LBP^{riu2} for invariant rotation texture classification; TAP^A [19] for capturing topological information.

B. LBP-based variants in still images

The typical LBP remains several restrictions such as small region of support, lack of global textural information, and noise sensitivity. A lot of attempts have been made to treat these shortcomings by addressing different steps: preprocessing [20], [21], thresholding [22], pattern selecting [23], feature training [24], mapping [19], [25], etc. Complementary information [26] was also used by exploiting variation of magnitudes.

C. LBP-based variants for dynamic texture

VLBP: Zhao et al. [3] enlarged the encoding of basic LBP to videos for description of dynamic texture. They consider neighbors sampled on three circles from three consecutive frames centered at a same spatial coordinate together with the centers from the first and the last frames. By considering the relations between these $3P + 2$ neighbors and the second centers, they obtained a binary code of length $3P + 2$ to capture local motions around this voxel. This encoding asks a small value of P due to the high dimensionality of descriptor with 2^{3P+2} bins. Later, D. Tiwari et al. [15] extended this idea by combining with CLBP to introduce CVLBP operator.

LBP-TOP: To remedy the high dimensionality of VLBP, Zhao et al. [3] proposed another LBP-based encoding, called LBP-TOP. Their idea is to consider LBP operator on three orthogonal planes passing through the considered voxel. The histograms, obtained on each orthogonal plane, are then concatenated to form the descriptor of DT video.

III. PROPOSED METHOD

A. Overview

We propose a new operator for DT representation by considering local structure patterns (LSP) [22] combining with completed schema of LBP (CLBP) [26] for static images in three orthogonal planes to capture spatio-temporal texture structures. Fig. 1 illustrates the proposed operator, named Completed Local Structure Patterns on Three Orthogonal Planes (CLSP-TOP). This approach, which is introduced as an extension of LBP-TOP [3], is based on two main following improvements to enhance the discrimination power of descriptor.

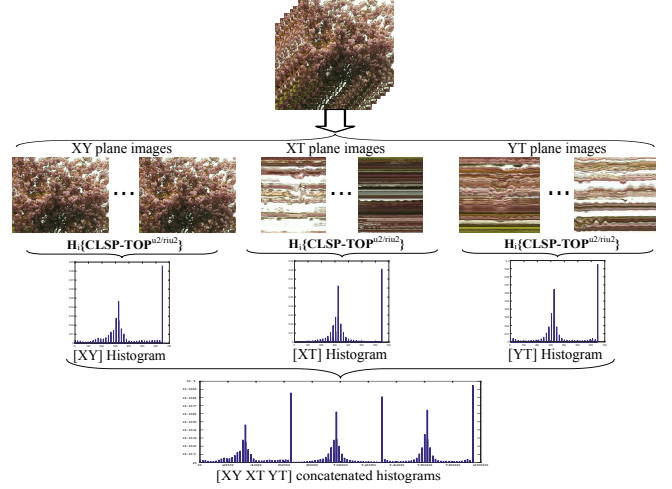


Fig. 1. Illustration of completed local structure patterns on three orthogonal planes

- LSP replaces LBP and allows to capture more stable spatial relations using adaptative global/local thresholds to remedy the problems of sensitivity to noise and near uniform images of the typical LBP.
- Completed schema (CLBP) is also inspired to take into account complementary texture information in the local variation of magnitudes.

We detail then hereafter the proposed CLSP-TOP operator.

B. Local Structure Patterns

By thresholding by center pixel, the typical LBP captures effectively the local spatial relations around this pixel. In the meanwhile, it also leads to two of the main issues of LBP, the sensitivity to noise and near uniform images, since a small change of center pixel can largely modify the obtained binary code. We adopt in this work the adaptive thresholding proposed in [22]. The authors introduced two complemented components for thresholding. The first one, named Local Average Difference pattern (LAD), is defined as the mean of local variation of magnitudes around center pixel x_c as follows:

$$\text{LAD}(x_c) = \frac{1}{P} \sum_{p=1}^P |f(x_p) - f(x_c)| \quad (3)$$

where $f(\cdot)$ is grayscale image value of a pixel, x_c is the center point, x_p is the p^{th} neighbor of x_c ($p \in [1, P]$).

The second one, called Global Mean Difference pattern (GMD), is calculated as the mean of the absolute differences over the entire image. An adaptive threshold is then proposed as follows to calculate binary patterns, called Local Structure Patterns (LSP), at center pixel x_c .

$$T(x_c) = f(x_c) + \frac{a \cdot \text{LAD}(x_c) + b \cdot \text{GMD}}{a + b}, \quad a, b = \{0, 1\} \quad (4)$$

When $a = b = 0$, LSP is simply identical to LBP and this case is not considered.

C. Completed LBP

The typical LBP code also omits local variation of magnitude containing rich local textural information. Guo et al. [26] have overcome this issue by introducing completed LBP operator. They introduced two main components: $CLBP_S$ that is identical to LBP; $CLBP_M$ for capturing local variation of magnitudes. To construct $CLBP_M$, the differences of magnitudes between the center pixel and its neighbors are thresholded by its mean value calculated over the entire image. In addition, the third component $CLBP_C$ is also introduced to take into account global information of each center pixel. These components are complemented, thus they are often combined together to significantly improve the performance. The most popular combination, which is joined histograms of these components, is adopted in our proposed framework.

D. Dynamic texture representation with CLSP-TOP

As mentioned above, our proposed descriptor relies on using LSP in completed schema (CLBP), called CLSP, instead of using the typical operator LBP. To exploit spatio-temporal relations, we adopt the idea of LBP-TOP by considering CLSP on three orthogonal planes (XY, XT, YT) and then the descriptor is obtained by concatenating three histograms calculated from these planes. Two possible mappings can be used in our framework: *riu2* giving a descriptor of $6(P+2)^2$ dimensions, *u2* giving a descriptor of $6((P-1)P+3)^2$ dimensions, where P is the number of considered neighbors.

Furthermore, we take into account the advantage of multi-scale analysis to improve the recognition accuracy, in which a computation of multiple operators with various (P, R) outputs corresponding histograms which are normalized and concatenated to form multi-scale representation MCLSP-TOP.

E. Dissimilarity measure

In this paper, to concentrate on the performance of descriptor, we only use the simple nearest neighborhood classifier with the χ^2 similarity measure to estimate the dissimilarity D between two histograms. The estimation of dissimilar distance D is calculated as

$$D(t, m) = \sum_{b=1}^B \frac{(t_b - m_b)^2}{t_b + m_b} \quad (5)$$

where B is the total of bins, t_b and m_b are the values of the sample and the model image at the b^{th} bin respectively.

IV. EXPERIMENTS

We present a comprehensive evaluation of our method on different classic datasets by following specific experimental protocols and compare to the state-of-the-art results. Results of our method on DT datasets (UCLA, DynTex, and DynTex++) with *riu2* (multi-scale) and *u2* (multi-scale) configurations are presented in Table I (Table III) and Table II (Table IV) respectively. Results of the LBP-TOP, VLBP operators are referred to the evaluations of [16] and [9] while the remains

come from the original approaches. Bold rates in Tables V, VI, VII indicate the highest recognition accuracies.

A. Experimental settings

Using single scale also leads to good results (see Tables I, II) but multiscale is recommended since the performance is still improved (see Tables III, IV). In this case, LSP's parameters are complemented and give best results with $a = b = 1$. Concerning the neighborhood configuration, the best settings are chosen as follows to compare with existing methods: *riu2* mapping with multiscale $\{(P, R)\} = \{(6, 1), (6, 2), (6, 3)\}$ giving good compromise for almost test cases; *u2* mapping with multiscale $\{(P, R)\} = \{(4, 1), (6, 3)\}$ or $\{(P, R)\} = \{(4, 2), (6, 3)\}$ for particular test case.

B. Datasets and experimental protocols

UCLA dataset: UCLA dataset [1] originally comprises 50 classes (4 DTs per class) of various 200 DT sequences which illustrate fountain, fire, boiling water, waterfall, plant, and flower. Each sequence has 75 frames with 160×110 pixels for each frame. A small version of UCLA usually used for DT recognition is clipped by a 48×48 pixel window to capture the key statistical and dynamical features. Three following benchmarks are widely considered for this dataset.

- *50-class breakdown:* 50 DT classes are used by considering 2 possible experimental protocols [1], [7], [13], [16], [27] : *leave-one-out* and *four cross-fold validation*.
- *9-class breakdown:* 50 DT classes are grouped into 9 semantic categories for DT classification. Due to [17], 50% of DT sequences in each class are randomly picked out for testing and the rest for training. The average result of 20 runtimes is considered as the final result.
- *8-class breakdown:* It is similar to *9-class breakdown* except 50 DT classes are now grouped into 8 semantic categories making the scheme more challenging [7].

DynTex dataset: DynTex dataset [28] originally consists of 656 videos captured under different environmental conditions and recorded in AVI format. In our experiments, we use "pr1" DynTex version¹ of 679 sequences with reasonable dimension of 352×288 and 250 color frames in 10 seconds. Following the protocol in [3], [13], [16], we use a version of the "pr1" dataset with 35 sequence categories, named as DynTex35. Each sequence is considered as a class and split into 8 non-overlapping sub-DTs with random cutting points along X, Y, T axes, but not half in these. For instance, partition points in the trial is selected as [3], i.e. $x = 170, y = 130, t = 100$. In addition, two sub-DTs are collected by randomly partitioning along T axis of the original sequence. As a result of that, 10 sub-DTs for each sequence have various spatio-temporal dimension and are more challenging for classification function.

Three following subsets of DynTex are often used as benchmarks for DT recognition using *leave-one-out cross validation* [10], [13], [29].

¹<http://dynTex.univ-lr.fr/download.html>

TABLE I
CLASSIFICATION RATES (%) ON DT DATASETS² USING CLSP-TOP^{riu2}

P,R,a,b	UCLA				DynTex				
	L50	4C	C9	C8	Dyn35	Alpha	Beta	Gamma	Dyn++
4,1,1,0	96.50	96.50	97.60	95.65	98.57	90.00	85.80	86.36	92.68
4,1,0,1	97.50	98.00	95.65	94.78	97.71	93.33	87.04	86.74	92.85
4,1,1,1	97.00	97.50	95.90	95.10	97.43	93.33	87.65	87.12	93.17
8,1,1,0	96.50	96.50	96.75	95.10	97.71	93.33	87.65	88.26	94.00
8,1,0,1	98.00	98.00	97.80	96.08	97.14	96.67	88.27	86.36	93.79
8,1,1,1	98.00	98.00	97.00	95.10	97.14	95.00	87.04	87.50	93.83

TABLE II
CLASSIFICATION RATES (%) OF CLSP-TOP^{u2} ON DT DATASETS²

P,R,a,b	UCLA				DynTex				
	L50	4C	C9	C8	Dyn35	Alpha	Beta	Gamma	Dyn++
4,2,1,0	95.50	95.50	97.25	95.22	96.57	91.67	88.27	87.12	94.35
4,2,0,1	97.00	97.00	98.05	96.41	96.86	91.67	91.36	88.64	94.79
4,2,1,1	97.00	97.00	97.70	95.54	96.86	91.67	91.36	89.77	95.24
7,1,1,0	96.50	97.00	97.75	97.50	96.57	95.00	88.27	88.26	94.96
7,1,0,1	97.00	97.00	98.40	95.76	96.57	95.00	90.12	88.26	96.07
7,1,1,1	97.00	97.00	97.25	95.98	96.57	95.00	90.74	89.39	95.51

- *Alpha*: 60 DT videos are grouped in three classes: *grass*, *sea*, and *trees* with 20 sequences per class.
- *Beta*: 162 DT videos are divided into 10 classes: *sea*, *vegetation*, *trees*, *flags*, *calm water*, *fountains*, *smoke*, *escalator*, *traffic*, and *rotation* with various numbers of sequences for each.
- *Gamma*: 264 DT videos are separated into 10 categories: *flowers*, *sea*, *naked trees*, *foliage*, *escalator*, *calm water*, *flags*, *grass*, *traffic*, and *fountains*. Each class contains a diverse collection of sequences.

DynTex++: Ghanem et al. [17] stated an extension of DynTex which was compiled by selecting 345 raw AVI videos from 656 sequences of DynTex. Each of which only includes one DT, not contain dynamic background, panning, and zooming. The selected sequences were filtered, preprocessed, and grouped into 36 classes with 100 sequences in fixed size of $50 \times 50 \times 50$ for each DT, i.e. 3600 DTs in total. Following [13], [17], a half of DTs in each class is randomly selected for testing and the remain for training. The test is repeated 10 times to take the average value as the final result.

TABLE III
RESULTS (%) OF MCLSP-TOP^{riu2} ON DT DATASETS²

{P,[R]},a,b	UCLA				DynTex				
	L50	4C	C9	C8	Dyn35	Alpha	Beta	Gamma	Dyn++
{6,[1,2]},1,0	96.00	96.00	96.70	94.02	97.71	95.00	89.51	88.64	94.18
{6,[1,2]},0,1	98.00	98.50	97.45	96.84	97.71	95.00	90.12	87.88	94.13
{6,[1,2]},1,1	98.00	98.00	97.10	93.91	97.71	95.00	90.74	88.64	94.00
{6,[1,2,3]},1,0	97.50	97.50	96.95	94.89	98.29	93.33	87.65	86.74	93.78
{6,[1,2,3]},0,1	99.00	99.00	96.75	96.63	97.71	95.00	88.89	87.88	93.60
{6,[1,2,3]},1,1	99.00	99.00	98.30	97.06	97.71	95.00	90.12	89.39	93.73

Note: {P,[R]} means multi-scales of P with various R.

²L50: leave-one-out 50 classes; 4C: four cross-fold scheme; C9: 9-class breakdown; C8: 8-class breakdown; Dyn35: DynTex with 35 categories; Dyn++: DynTex++ dataset.

TABLE IV
CLASSIFICATION RATES (%) OF MCLSP-TOP^{u2} ON DT DATASETS²

{(P,R)},a,b	UCLA				DynTex				
	L50	4C	C9	C8	Dyn35	Alpha	Beta	Gamma	Dyn++
{(4,1),(6,3)},1,0	97.50	97.50	97.25	96.30	96.86	91.67	90.12	88.64	95.25
{(4,1),(6,3)},0,1	98.00	98.00	97.35	96.63	96.57	91.67	88.89	90.15	95.50
{(4,1),(6,3)},1,1	97.50	97.50	98.60	97.72	97.14	91.67	89.51	90.53	95.50
{(4,2),(6,3)},1,0	96.00	96.00	97.40	92.83	97.14	91.67	90.12	87.50	95.00
{(4,2),(6,3)},0,1	98.50	98.50	97.00	96.41	96.86	91.67	91.98	89.02	94.85
{(4,2),(6,3)},1,1	98.00	98.50	96.95	94.89	97.14	91.67	91.36	91.29	95.36

TABLE V
COMPARISON OF CLASSIFICATION RATE (%) ON UCLA DATASET²

Method	L50	4C	C9	C8
VLBP [3]	-	89.50	96.30	91.96
CVLBP [15]	-	93.00	96.90	95.65
LBP-TOP [3]	-	94.50	96.00	93.67
AR-LDS [1]	89.90	-	-	54.12
KDT-MD [6]	-	89.50	-	-
Space-time oriented [27]	81.00	-	-	-
NLDR [7]	-	-	-	70.00
MBSIF-TOP [13]	99.50	-	98.75	97.80
DFS [30]	-	89.50	-	-
3D-OTF [31]	-	87.10	96.32	95.80
WMFS [32]	-	-	96.95	97.18
DNGP [2]	-	-	98.10	97.00
Chaotic vector [8]	-	-	85.10	85.00
High level feature [33]	-	-	92.67	85.65
HLBP [16]	95.00	95.00	98.35	97.50
PCANet-TOP [10]	99.50	-	-	-
Ours	99.00	99.00	98.60	97.72

Note: All the results using the 1-NN classifier; "-" means "not available".

C. Recognition on UCLA dataset

1) *50-class breakdown*: Table V presents the result of our method compared to recent existing approaches by using two popular experimental protocols: *Leave-one-out* and *Four cross-fold validation*. It can be observed in Tables I, II that our method achieves the best performance among other competitors with recognition rates from 97% to 98% using single resolution descriptor. With 3-multi-scale descriptor of 1,152 bins, our rate of 99% (see Table III) obtains 3% margin better classification compared to 96% of MBSIF-TOP [13] using 3-scale with 2,304 bins. The highest rate of 99.5% on this scheme (see Table V) is achieved by filter-based methods as MBSIF-TOP with 7-scale and PCANet-TOP [10] using PCA and deep multi-scale convolutional network. However, they take complex computation and need much time to operate.

2) *9-class breakdown*: Table V presents our result on this scheme compared to other methods. It can be realized that our method achieves the highest recognition accuracy of 98.60% using 2-scale representations of {(4,1),(6,3)} compared to the spatio-temporal LBP results. Only MBSIF-TOP [13] with descriptor dimension of 6,144 bins performs 0.15% slightly higher than our method using 7,884 bins. The confusion matrix (see Fig. 2) indicates that our approach mostly confused *smoke* with *water*, *fountain* with *waterfall*, *fire* with *smoke* sequence because of the similar features of these sequences.

	Boiling water	Fire	Flowers	Fountains	Plants	Sea	Smoke	Water	Waterfall
Boiling water	100	0	0	0	0	0	0	0	0
Fire	0	93.75	0	2.5	0	0	3.75	0	0
Flowers	0	0	100	0	0	0	0	0	0
Fountains	0	0	2	94	0	0	0	0	4
Plants	0	0	0	0	100	0	0	0	0
Sea	0	0	0	0	0	100	0	0	0
Smoke	0	0	0	0	0	0	80	20	0
Water	0	0	0	0	0	0	0	97.5	2.5
Waterfall	0	0	0	0	0	0	0	0	100

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

	Boiling water	Fire	Flower	Fountains	Sea	Smoke	Water	Waterfall
Boiling water	100	0	0	0	0	0	0	0
Fire	0	95	0	1.25	0	3.75	0	0
Flower	0	0	100	0	0	0	0	0
Fountains	0	0	0	98.5	0	0	0	0.5
Sea	0	0	0	0	100	0	0	0
Smoke	0	0	0	0	0	70	30	0
Water	0	0	0	0	0	3.33	96.67	0
Waterfall	0	0	0	0	0	0	0	100

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

3) *8-class breakdown*: Our result on this scheme is 97.72% of classification rate (see Tables IV, V) which significantly outperforms in comparison to other approaches, just slightly 0.08% lower than MBSIF-TOP [13]. The confusion matrix of 8-class breakdown shows the detailed performance of the proposed method for each class (see Fig. 3). It can be obtained from the confusion matrix that our method mainly confused *smoke* with *water* sequence and *fire* sequence with *fountain* and *smoke* due to the very similar characteristics between these sequences.

D. Recognition on DynTex dataset

Our best performance on DynTex35 is 98.57%, which is the same as HLBP [16], but only using 216 bins (see Table I). The detail of classification is specifically shown in Fig. 4. In the best configurations formed for comparison, the rate is 98.29% (see Table III). The dictionary learning approach (Orthogonal Tensor DL) [12] collected 0.43% slightly higher than ours but it ineffectively reacted with DynTex variants (i.e. *Alpha*, *Beta*, *Gamma*) which our method mostly outperforms

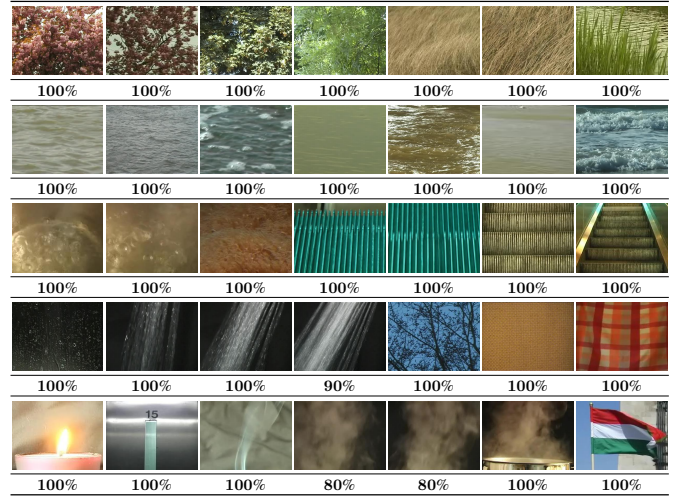


Fig. 4. Classification rates of the proposed method on DynTex35

TABLE VI
RESULTS OF RECOGNITION RATE (%) ON DYNTEX DATASET

Method	Dyn35	Alpha	Beta	Gamma
VLBP [3]	81.14	-	-	-
CVLBP [15]	85.14	-	-	-
HLBP [16]	98.57	-	-	-
DFS [30]	97.63	84.90 ^s	76.50 ^s	74.50 ^s
DFS+ [34]	-	85.20 ^s	76.90 ^s	74.80 ^s
3D-OTF [31]	-	82.80 ^s	75.40 ^s	73.50 ^s
Orthogonal Tensor DL [12]	99.00	87.80 ^s	76.70 ^s	74.80 ^s
Equiangular Kernel DL [11]	-	88.80 ^s	77.40 ^s	75.60 ^s
2D+T [29]	-	85.00	67.00	63.00
MBSIF-TOP [13]	98.61	90.00	90.70	91.30
PCANet-TOP [10]	-	93.33	90.12	89.39
LBP-TOP [3]	92.45	96.67	85.80	84.85
st-TCoF [9]	-	98.33^d	98.15^d	98.11^d
Ours	98.29	95.00	91.98	91.29

Note: Superscript "d" indicates the approach using deep structural TCoF; "s" is for results using the SVM; "-" means "not available".

compared to the state-of-the-art approaches with recognition rates of 95.00%, 91.98%, 91.29% for *Alpha*, *Beta*, *Gamma* respectively (see Table VI). The deep structure approach extracting Transferred ConvNet Feature (TCoF) scheme [9] obtained impressive results (over 98% for DynTex variants). However, it used a huge dimension of covariance feature along with a complicated learning algorithm.

E. Recognition on DynTex++ dataset

Our best result of recognition rate on this scheme is 96.07% (see Table II), nearly same HLBP [16]. In the best configurations formed for comparison, the classification rate of 95.50% (see Table VII) demonstrates our method performs well compared to existing approaches except methods: MBSIF-TOP [13] using a large feature vector (8-scale); HLBP [16] in which local pixels get involved with their surrounding neighbors in consecutive frames; DDLBP with MJMI [14] using SVM classifier. It should note that we only used simple nearest neighborhood criterion for classification. On the other hand, it may be that our method, based on local structure patterns

TABLE VII
RESULTS OF RECOGNITION RATE (%) ON DYNTEX++ DATASET

Encoding method	Recognition rate (%)
DL-PEGASOS [17]	63.70 ^s
NLSSA [35]	92.40 ^s
PCA-cLBP/PI-LBP/PD-LBP [36]	91.90 ^s
DFS [30]	89.90 ^s
DFS+ [34]	91.70 ^s
3D-OTF [31]	89.17 ^s
DDLBP with MJMI [14]	95.80 ^s
Orthogonal Tensor DL [12]	94.70 ^s
Equiangular Kernel DL [11]	93.40 ^s
MBSIF-TOP [13]	97.12
DNGP [2]	90.20
Chaotic vector [8]	69.00
High level feature [33]	64.22
VLBP [3]	94.98
LBP-TOP [3]	94.05
HLBP [16]	96.28
Ours	95.50

Note: Superscript "s" indicates results using the SVM classifier.

of static images, is more sensitive to similar characteristics of DTs in DynTex++ whose sequences include only a main motion captured from the original DynTex.

V. CONCLUSION

We have proposed an effective spatio-temporal DT descriptor based on LSP along with using a robust adaptive thresholding and CLBP for complementary information on three orthogonal planes. Experiments on different DT datasets have demonstrated that the proposed operator significantly outperforms almost recent approaches of DT classification. For the perspective, utilizing TAP^A [19] to capture topological information and dissimilarity distances for classifiers should be addressed in the future work.

REFERENCES

- [1] P. Saisan, G. Doretto, Y. N. Wu, and S. Soatto, "Dynamic texture recognition," in *CVPR*, 2001, pp. 58–63.
- [2] A. R. Rivera and O. Chae, "Spatiotemporal directional number transitional graph for dynamic texture recognition," *IEEE Trans. PAMI*, vol. 37, no. 10, pp. 2146–2152, 2015.
- [3] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. PAMI*, vol. 29, no. 6, pp. 915–928, 2007.
- [4] T. P. Nguyen, A. Manzanera, M. Garrigues, and N. Vu, "Spatial motion patterns: Action models from semi-dense trajectories," *IJPRAI*, vol. 28, no. 7, 2014.
- [5] C. Peh and L. F. Cheong, "Synergizing spatial and temporal texture," *IEEE Trans. Image Processing*, vol. 11, no. 10, pp. 1179–1191, 2002.
- [6] A. B. B. Chan and N. Vasconcelos, "Classifying video with kernel dynamic textures," in *CVPR*, 2007, pp. 1–6.
- [7] A. Ravichandran, R. Chaudhry, and R. Vidal, "View-invariant dynamic texture recognition using a bag of dynamical systems," in *CVPR*, 2009, pp. 1651–1657.
- [8] Y. Wang and S. Hu, "Chaotic features for dynamic textures recognition," *Soft Computing*, vol. 20, no. 5, pp. 1977–1989, 2016.
- [9] X. Qi, C.-G. Li, G. Zhao, X. Hong, and M. Pietikäinen, "Dynamic texture and scene classification by transferring deep image features," *Neurocomputing*, vol. 171, pp. 1230 – 1241, 2016.
- [10] S. R. Arashloo, M. C. Amirani, and A. Noroozi, "Dynamic texture representation using a deep multi-scale convolutional network," *JVCIR*, vol. 43, pp. 89 – 97, 2017.

- [11] Y. Quan, C. Bao, and H. Ji, "Equiangular kernel dictionary learning with applications to dynamic texture analysis," in *CVPR*, 2016, pp. 308–316.
- [12] Y. Quan, Y. Huang, and H. Ji, "Dynamic texture recognition via orthogonal tensor dictionary learning," in *ICCV*, 2015, pp. 73–81.
- [13] S. R. Arashloo and J. Kittler, "Dynamic texture recognition using multiscale binarized statistical image features," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2099–2109, 2014.
- [14] J. Ren, X. Jiang, J. Yuan, and G. Wang, "Optimizing LBP structure for visual recognition using binary quadratic programming," *IEEE Signal Processing Letters*, vol. 21, no. 11, pp. 1346–1350, 2014.
- [15] D. Tiwari and V. Tyagi, "Dynamic texture recognition based on completed volume local binary pattern," *MSSP*, vol. 27, no. 2, pp. 563–575, 2016.
- [16] D. Tiwari and V. Tyagi, "A novel scheme based on local binary pattern for dynamic texture recognition," *CVIU*, vol. 150, pp. 58–65, 2016.
- [17] B. Ghanem and N. Ahuja, "Maximum margin distance learning for dynamic texture recognition," in *ECCV*, ser. LNCS, K. Daniilidis, P. Maragos, and N. Paragios, Eds., vol. 6312, 2010, pp. 223–236.
- [18] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. PAMI*, vol. 24, no. 7, pp. 971–987, 2002.
- [19] T. P. Nguyen, A. Manzanera, W. G. Kropatsch, and X. S. N'Guyen, "Topological attribute patterns for texture recognition," *Pattern Recognition Letters*, vol. 80, pp. 91–97, 2016.
- [20] T. P. Nguyen, N. Vu, and A. Manzanera, "Statistical binary patterns for rotational invariant texture classification," *Neurocomputing*, vol. 173, pp. 1565–1577, 2016.
- [21] N. Vu, T. P. Nguyen, and C. Garcia, "Improving texture categorization with biologically-inspired filtering," *Image and Vision Computing*, vol. 32, no. 6-7, pp. 424–436, 2014.
- [22] N. Shrivastava and V. Tyagi, "An effective scheme for image texture classification based on binary local structure pattern," *The Visual Computer*, vol. 30, no. 11, pp. 1223–1232, 2014.
- [23] S. Liao, M. W. K. Law, and A. C. S. Chung, "Dominant local binary patterns for texture classification," *IEEE Trans. Image Processing*, vol. 18, no. 5, pp. 1107–1118, 2009.
- [24] L. Nanni, S. Brahnam, and A. Lumini, "A Simple Method for Improving Local Binary Patterns by considering Non-uniform Patterns," *Pattern Recognition*, vol. 45, no. 10, pp. 3844–3852, 2012.
- [25] A. Fathi and A. R. Naghsh-Nilchi, "Noise Tolerant Local Binary Pattern Operator for Efficient Texture Analysis," *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1093–1100, 2012.
- [26] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010.
- [27] K. G. Derpanis and R. P. Wildes, "Spacetime texture representation and recognition based on a spatiotemporal orientation analysis," *IEEE Trans. PAMI*, vol. 34, no. 6, pp. 1193–1205, 2012.
- [28] R. Péteri, S. Fazekas, and M. J. Huiskes, "Dyntex: A comprehensive database of dynamic textures," *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1627–1632, 2010.
- [29] S. Dubois, R. Péteri, and M. Ménard, "Characterization and recognition of dynamic textures based on the 2d+t curvelet transform," *Signal, Image and Video Processing*, vol. 9, no. 4, pp. 819–830, 2015.
- [30] Y. Xu, Y. Quan, H. Ling, and H. Ji, "Dynamic texture classification using dynamic fractal analysis," in *ICCV*, 2011, pp. 1219–1226.
- [31] Y. Xu, S. B. Huang, H. Ji, and C. Fermüller, "Scale-space texture description on sift-like textons," *CVIU*, vol. 116, no. 9, pp. 999–1013, 2012.
- [32] H. Ji, X. Yang, H. Ling, and Y. Xu, "Wavelet domain multifractal analysis for static and dynamic texture classification," *IEEE Trans. Image Processing*, vol. 22, no. 1, pp. 286–299, 2013.
- [33] Y. Wang and S. Hu, "Exploiting high level feature for dynamic textures recognition," *Neurocomputing*, vol. 154, pp. 217–224, 2015.
- [34] Y. Xu, Y. Quan, Z. Zhang, H. Ling, and H. Ji, "Classifying dynamic textures via spatiotemporal fractal analysis," *Pattern Recognition*, vol. 48, no. 10, pp. 3239–3248, 2015.
- [35] M. Baktashmotlagh, M. T. Harandi, A. B. C. C. Lovell, and M. Salzmann, "Discriminative non-linear stationary subspace analysis for video classification," *IEEE Trans. PAMI*, vol. 36, no. 12, pp. 2353–2366, 2014.
- [36] J. Ren, X. Jiang, and J. Yuan, "Dynamic texture recognition using enhanced LBP features," in *ICASSP*, 2013, pp. 2400–2404.