Dynamic Texture Representation Based on Hierarchical Local Patterns

Thanh Tuan Nguyen[1,2], Thanh Phuong Nguyen[1], and Frédéric Bouchara[1]

[1] Université de Toulon, Aix Marseille Université, CNRS, LIS, Marseille, France
[2] HCMC University of Technology and Education, Faculty of IT, HCM City, Vietnam

Abstract. A novel effective operator, named HIerarchical LOcal Pattern (HILOP), is proposed to efficiently exploit relationships of local neighbors at each adjacent pairwise of different regional hierarchies located surrounding a center pixel of a texture image. Instead of thresholding by the value of central pixel, the gray-scale of each local neighbor in a hierarchical area is compared to that of all of neighbors in the remain region. In order to capture shape and motion cues for dynamic texture (DT) representation, HILOP is taken into account for investigating hierarchical relationships in plane images of a DT sequence. The obtained histograms are then concatenated to form a robust descriptor with high performance for DT classification task. Experiments on various benchmark datasets (i.e., UCLA, DynTex, DynTex++) have validated the interest of our proposal.

Keywords: Dynamic texture · Hierarchical local pattern · Hierarchical encoding · LBP · Video representation.

1 Introduction

Efficiently encoding dynamic textures (i.e., textural structures repeated in a temporal domain) is a decisive task of various applications in computer vision, e.g., facial expressions [34, 44], tracking objects [18, 40, 11], fire and smoke detection [6], etc. To this end, many approaches have been proposed for DT representation in which the main problems (e.g., turbulent motions, noise, illumination, etc.) are addressed in order to improve the discrimination power in DT recognition. These approaches are roughly grouped into the following categories. First, *optical-flow-based methods* [25, 27, 10, 23] are mainly based on direction properties of normal flow to effectively capture the turbulent motion characteristics of DTs in sequences. In the meantime, *model-based methods* [7, 17, 41, 29] have addressed Linear Dynamical System (LDS) [35] and its variants in order to deal with the complication of chaotic motions (e.g., turbulent water) and camera moving features (e.g., panning, zooming, and rotations). On the other side, filter bank techniques are exploited in *filter-based methods* to reduce the negative impacts of illumination and noise on video representation [3, 34]. Motivated by geometry theory, *geometry-based methods* estimate self-similarity features using fractal analysis in order to be robust against the illumination and environmental changes, such as Dynamic Fractal Spectrum (DFS) [43], Multi-Fractal Spectrum (MFS) [42], Wavelet-based MFS [15], Spatio-Temporal Lacunarity Spectrum (STLS) [32]. Recently, *learning-based methods* are interested in, particularly deep learning techniques thanks to their high accuracy of DT recognition. Two trends of those are as follows: i) exploiting Convolutional Neural Network (CNN) for capturing deep features [28, 1, 2]; ii) using kernel sparse coding for learning featured dictionaries for DT description [31, 30]. In the meanwhile, *local-feature-based methods* have also achieved promising rates with simple and efficient computation, in which Local Binary Pattern (LBP) operator [24] and its variants have been taken into account DT representation. Two main techniques of those are mostly prompted for video description as follows: Volume LBP (VLBP) [44] for encoding dynamical features in consideration of spatio-temporal relationships on three consecutive frames; and LBP-TOP [44] for capturing motion and shape clues by using LBP on three orthogonal planes of sequences.

In consideration of gray-level differences between a center pixel and its local regions, LBP-based variants have acquired the promising rates on DT classification. However, they have remained several internal limitations, such as sensitivity to noise, near uniform regions [37, 21], and large dimension [44, 33, 36]. Addressing those obstacles, we introduce in this work a novel and effective operator HILOP to capture relationships between its local neighbors at a pairwise of different hierarchical regions. Accordingly, a center pixel is encoded by comparing the gray value of each neighbor in the first hierarchical region with all of those in the other. HILOP is then involved with analyzing plane images of a DT sequence in order to structure spatio-temporal features for DT representation. The obtained probability distributions are concatenated and normalized to form a descriptor with more discrimination power for DT classification. In short, it can be listed the major contributions of this work as follows.

- A novel, efficient local operator HILOP is able to efficiently capture textural features based on analyzing a pairwise of hierarchical regions where the local neighbors in a hierarchy are consecutively thresholded by all of those in the remain instead of by the center pixel as the existing LBP-based variants.
- Multi-hierarchy HILOP encoding allows to enrich appearance information by addressing more further supporting regions.
- An efficient framework for DT representation in which spatio-temporal features are hierarchically exploited thanks to utilizing benefits of HILOP.

2 Proposed Method

As mentioned above, using a simple computation, local-feature-based methods have achieved promising results on DT classification. However, limitations of their performance are often caused by problems of sensitivity to noise, illumination, and near uniform regions. In this section, we first take a look of LBP and its variants as well as their application in DT representation. We then propose a novel, simple operator HILOP in order to investigate local relationships in hierarchical supporting regions. Finally, an efficient framework for DT description is presented to take advantage of the beneficial properties of HILOP for addressing above restrictions of LBP-based variants.

2.1 A brief review of LBP and its operation for DT description

In consideration of relationships between a pixel and its surrounding regions, Ojala et al. [24] introduced a simple operator LBP in order to capture local characteristics for still image representation. Appropriately, let $\mathcal{I}$ denote a 2D gray-scale image. A LBP code of a pixel $\mathbf{q} \in \mathcal{I}$ is featured by comparing the gray-level differences between $\mathbf{q}$ and its local neighbors $\{\mathbf{p}_i\}_{i=1}^{P}$ as follows.

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

in which P means quantity of $\mathbf{q}$'s neighbors that are sampled on a circle of radius R using an interpolated calculation, $\mathcal{I}(.)$ points out the gray-value of a pixel, and function $f(.)$ is defined as follows.

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

As the result of that, it takes a large dimension (i.e., 2^P distinct values) for describing a still image. Therefore, two conventional mappings should be applied in practice to deal with this restriction: $u2$ with $P(P-1)+3$ bins and $riu2$ with $P+2$ for structuring uniform patterns and rotation-invariant uniform patterns respectively. Furthermore, other mapping techniques can be also remarkable to enhance the encoding power, such as Local Binary Count [46] - a substitution for addressing uniform characteristics, topological mapping $TAP^{\mathcal{A}}$ [19].

Inspired by the simple and efficient properties of operator LBP in still image encoding, several efforts have taken it into account DT representation. First, Zhao et al. [44] structured a voxel by considering its P neighbors along with its two symmetrical voxels and $2P$ corresponding neighbors placed in the previous and posterior frames. All of these neighbors along with two symmetrical voxels are then thresholded by the concerning voxel to form a VLBP code of $3P+2$ binary bits. Due to the huge dimension of VLBP (i.e., 2^{3P+2} bins), it is limited in real applications. In order to handle this shortcoming, Zhao et al. [44] considered a voxel and its P neighbors on each orthogonal planes of a video to shape LBP-TOP patterns. The obtained probability distributions are concatenated and normalized to form the final descriptor with 3×2^P dimensions. After that, many proposals mainly based on these encoding models to improve the discriminative performance: CVLBC [45] - a combination of CLBC [46] and VLBP; CVLBP [36] - an integration of CLBP [13] and VLBP; CLSP-TOP [21], CSAP-TOP [22], and HLBP [37] - dealing with noise and illumination problems.

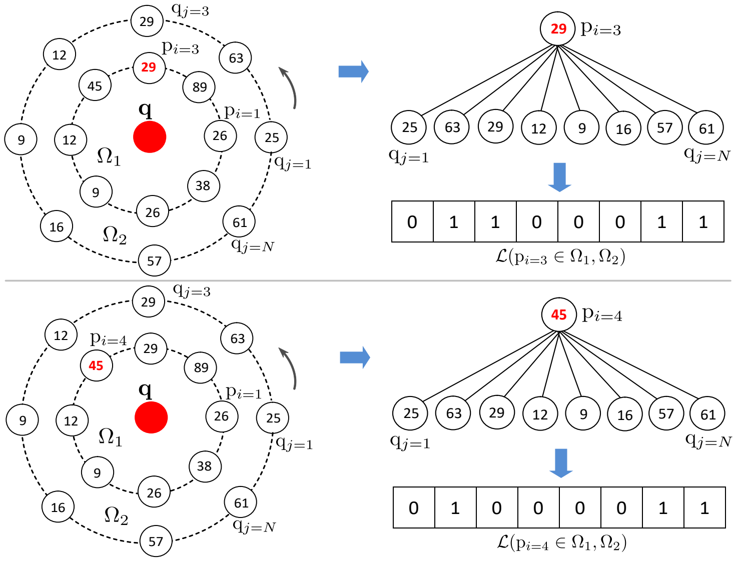

Fig. 1. An instance of structuring at $\mathbf{p}_{i=3}, \mathbf{p}_{i=4} \in \Omega_1$ based on $\{\mathbf{q}_j\}_{j=1}^{N=8}$ of Ω_2.

2.2 Hierarchical Local Patterns

Let $\Omega_1 = \{\mathbf{p}_i\}_{i=1}^N$ and $\Omega_2 = \{\mathbf{q}_j\}_{j=1}^N$ be two different hierarchies of supporting regions of a pixel $\mathbf{q}$ in a texture image $\mathcal{I}$, so that $\Omega_1 \bigcap \Omega_2 = \emptyset$. Each neighbor $\mathbf{p}_i$ in hierarchical region Ω_1 is encoded as a binary string of N bits by considering the difference of $\mathbf{p}_i$'s gray value with that of all $\mathbf{q}_j \in \Omega_2$ as follows.

$$\mathcal{L}(\mathbf{p}_i \in \Omega_1, \Omega_2) = \{g(\mathcal{I}(\mathbf{q}_j) - \mathcal{I}(\mathbf{p}_i))\}_{j=1}^N \tag{3}$$

in which $\mathcal{I}(.)$ returns the gray-value of a pixel. $g(.)$ is identical to Equation (2). Figure 1 graphically shows an instance of this computation using $N = 8$ neighbors for each circle-hierarchical region. Accordingly, two-hierarchical pattern of $\mathbf{q}$ based on a pairwise of supporting regions (Ω_1, Ω_2) is featured by addressing all neighbors $\mathbf{p}_i$ of supporting region Ω_1 as follows.

$$\Gamma_{\Omega_1,\Omega_2}(\mathbf{q}) = [\mathcal{L}(\mathbf{p}_i \in \Omega_1, \Omega_2)]_{i=1}^N \tag{4}$$

It should be noted that $\Gamma(.)$ is absolutely different from structuring difference-based patterns introduced in [16], i.e., RD-LBP and AD-LBP. More specifically, in this work, all of $\mathbf{q}_j \in \Omega_2$ are thresholded with each of $\mathbf{p}_i \in \Omega_1$ to be able to figure out N patterns. In contrast to that, RD-LBP [16] is formed by comparing a pairwise of $(\mathbf{q}_j, \mathbf{p}_j)$ in parallel to achieve only one pattern, while AD-LBP [16] is computed by addressing the differences of pixels in the same regions.

In order to forcefully enrich discriminative information, we address the function $\Gamma(.)$ on multi-region of adjacent hierarchies to capture more useful features

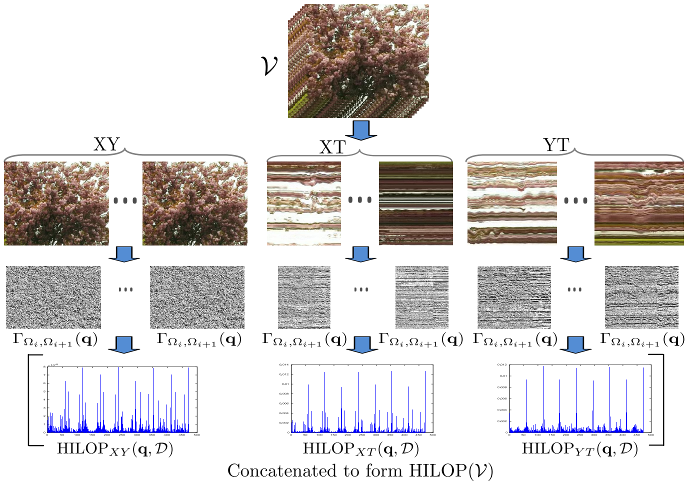
$\mathcal{V}$
XY
XT
YT
$\Gamma_{\Omega_i,\Omega_{i+1}}(\mathbf{q})$
$\Gamma_{\Omega_i,\Omega_{i+1}}(\mathbf{q})$
$\Gamma_{\Omega_i,\Omega_{i+1}}(\mathbf{q})$
$\Gamma_{\Omega_i,\Omega_{i+1}}(\mathbf{q})$
$\Gamma_{\Omega_i,\Omega_{i+1}}(\mathbf{q})$
$\Gamma_{\Omega_i,\Omega_{i+1}}(\mathbf{q})$
$\text{HILOP}_{XY}(\mathbf{q},\mathcal{D})$
$\text{HILOP}_{XT}(\mathbf{q},\mathcal{D})$
$\text{HILOP}_{YT}(\mathbf{q},\mathcal{D})$
Concatenated to form HILOP($\mathcal{V}$)

Fig. 2. Our proposed framework of encoding a video $\mathcal{V}$.

in further areas. According to that, let $\mathcal{D} = \{\Omega_1, \Omega_2, ..., \Omega_m\}$ be a set of hierarchical supporting regions of a pixel $\mathbf{q} \in \mathcal{I}$, so that $\Omega_k \bigcap \Omega_{k+1} = \emptyset$. HIerarchical LOcal Pattern (HILOP) of $\mathbf{q}$ is formed as follows.

$$\text{HILOP}_{\mathcal{I}}(\mathbf{q}, \mathcal{D}) = [\Gamma_{\Omega_k,\Omega_{k+1}}(\mathbf{q})]_{k=1}^{m} \tag{5}$$

2.3 DT Representation Based on HILOP Patterns

In this section, a simple framework to efficiently structure shape information and motion clues of DTs is proposed by exploiting the advantages of HILOP's properties for video representation. For an input sequence $\mathcal{V}$, the proposed framework takes three main stages as follows. Firstly, the video $\mathcal{V}$ is split into plane images subject to its three orthogonal planes $\{XY, XT, YT\}$ (see Fig. 2 for a graphical illustration). Secondly, the proposed operator HILOP is taken into account for analyzing each plane in order to capture hierarchical features based on a set of multi-layer supporting regions $\mathcal{D}$. Finally, the obtained histograms are concatenated and normalized to produce a robust descriptor HILOP($\mathcal{V}$) as

$$\text{HILOP}(\mathcal{V}) = [\text{HILOP}_{XY}(\mathbf{q}, \mathcal{D}), \text{HILOP}_{XT}(\mathbf{q}, \mathcal{D}), \text{HILOP}_{YT}(\mathbf{q}, \mathcal{D})] \tag{6}$$

Our proposed descriptor HILOP($\mathcal{V}$) based on the following beneficial properties in order to improve the discrimination power:

- The HILOP operator structures hierarchical patterns by considering relationships of a pairwise of regional hierarchies, instead of those between a center pixel and its local neighbors as conducted in LBP-based variants.
- The performance of proposed descriptor is enhanced thanks to taking HILOP into account analyzing plane images of a DT sequence in order to efficiently encode spatio-temporal properties of DTs.

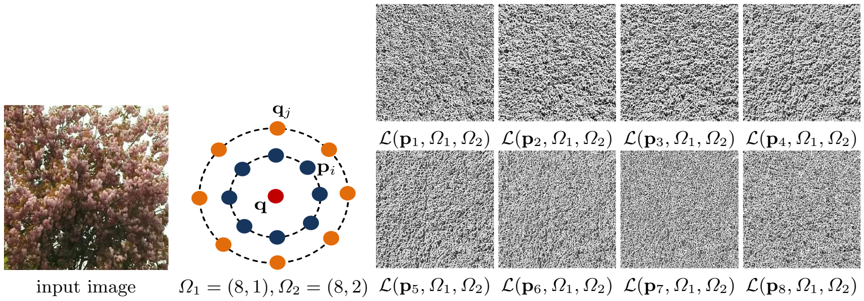
$\mathbf{q}_j$
$\mathbf{p}_i$
$\mathbf{q}$
input image
$\Omega_1 = (8,1), \Omega_2 = (8,2)$
$\mathcal{L}(\mathbf{p}_1, \Omega_1, \Omega_2)$
$\mathcal{L}(\mathbf{p}_2, \Omega_1, \Omega_2)$
$\mathcal{L}(\mathbf{p}_3, \Omega_1, \Omega_2)$
$\mathcal{L}(\mathbf{p}_4, \Omega_1, \Omega_2)$
$\mathcal{L}(\mathbf{p}_5, \Omega_1, \Omega_2)$
$\mathcal{L}(\mathbf{p}_6, \Omega_1, \Omega_2)$
$\mathcal{L}(\mathbf{p}_7, \Omega_1, \Omega_2)$
$\mathcal{L}(\mathbf{p}_8, \Omega_1, \Omega_2)$

Fig. 3. Several HILOP patterns structured using a hierarchy $\mathcal{D} = \{(8,1),(8,2)\}$.

– Incorporation of hierarchical features captured in multi-supporting hierarchies allows to enrich more forceful discriminative information.

3 Experiments

In order to verify execution of proposed descriptor HILOP, we address it for DT classification task on various benchmark datasets, i.e., UCLA [35], DynTex [26], and DynTex++ [12]. A linear multi-class SVM classifier which is enforced in the library of LIBLINEAR[3] [9] is employed with the default settings. The acquired results are then compared to those of the state-of-the-art approaches.

3.1 Experimental Settings

To be compliant with LBP encoding, each $\Omega_i \in \mathcal{D}$ should be structured by P neighbors which are interpolated on a circle of radius R_i at center pixel $\mathbf{q}$, i.e., $\Omega_i = (P, R_i)$. According to that, we address $\mathcal{D} = \{(8,1),(8,2),(8,3),(8,4),(8,5)\}$ in order to investigate further hierarchical local regions (i.e., $N = 8$). For computing a histogram, we use $u2$ mapping for each pattern $\mathcal{L}(.)$ to capture HILOP uniform features. As the result of that, the final descriptor HILOP has $(|\mathcal{D}|-1) * 3P(P(P-1)+3)$ dimensions (see Table 2 for specific instances), where $|\mathcal{D}|$ denotes the number of hierarchical regions involved in, i.e., $|\mathcal{D}| = 5$ in this case. Several HILOP samples of this encoding are shown in Fig. 3.

3.2 Datasets and Protocols

In order to verify the performance of our proposal, we detail in this section features of benchmark datasets along with their experimental protocols for DT recognition task. In addition, Table 1 shows their properties in brief for a look.

UCLA [35] includes 200 DT videos in 110 × 160 × 75 dimension that are categorized into 50 classes with four sequences for each of them (see Fig. 4(a) for some samples). With DT classification issue, a tiny version of 48 × 48 × 75 sequences is usually used and arranged into challenging sub-sets as follows.

[3] https://www.csie.ntu.edu.tw/∼cjlin/liblinear

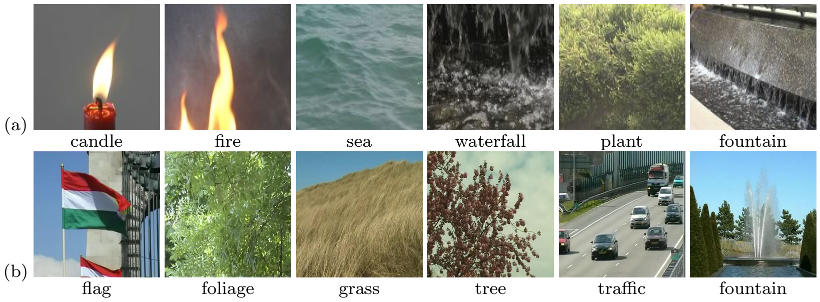

Fig. 4. Samples of DT videos in UCLA (a) and DynTex (b) datasets.

- *50-class:* Using the scheme of 50 original classes along with two following protocols: *leave-one-out* (LOO) [3, 38] and *4-fold cross validation* [21, 37].
- *9-class* and *8-class:* 200 DT sequences are grouped to form scheme *9-class* with the following labels and corresponding numbers of sequences: "boiling water"(8), "plants"(108), "flowers"(12), "fire"(8), "fountains"(20), "water"(12), "smoke"(4), "sea"(12), and "waterfall"(16). Because of the dominant quantity in "plants" category, it is eliminated to establish scheme *8-class* with more challenges in DT recognition [43]. Following the setting protocol in [12, 21], a half of sequences in each categories is randomly selected in order to train a classifying model, and the rest for the testing phase. The final rates on these schemes are reported as the average rates of 20 runtimes.

DynTex [26] consists of more than 650 high-quality DT sequences which are recorded in different conditions of environment (see Fig. 4(b) for some particular instances of them). Similar to the setting in [2, 3, 8], rates of DT classification are obtained by using LOO protocol for all of the following DynTex variants:

- *DynTex35* is composed by taking out 35 sequences from DynTex and splitting them into sub-videos in the following ways: randomly clipping each of them at partition points of X, Y, and T axes but not at the center of them to achieve 8 non-overlapping sub-sequences; 2 more obtained by splitting along its T axis. Finally, these outputs are arranged into 10 categories [3, 37, 44].
- *Alpha* consists of three categories of 20 sequences labeled as follows: "Sea", "Grass", and "Trees".
- *Beta* contains 162 sequences categorized into 10 groups with various numbers of samples: "sea", "vegetation", "trees", "flags", "calm water", "fountains", "smoke", "escalator", "traffic", and "rotation".
- *Gamma* also includes 10 groups of 264 DT sequences with different quantities: "flowers", "sea", "naked trees", "foliage", "escalator", "calm water", "flags", "grass", "traffic", and "fountains".

DynTex++ is constructed by 345 original sequences of DynTex which are pre-processed in order to retain dominant chaotic motions [12]. The outputs are then fixed in dimension of 50 × 50 × 50 and arranged into 36 groups with 100

Table 1. A summary of main properties of DT datasets.

Dataset	Sub-dataset	#Videos	Resolution	#Classes	Protocol
UCLA	50-class	200	$48 \times 48 \times 75$	50	LOO and 4fold
	9-class	200	$48 \times 48 \times 75$	9	50%/50%
	8-class	92	$48 \times 48 \times 75$	8	50%/50%
DynTex	DynTex35	350	different dimensions	10	LOO
	Alpha	60	$352 \times 288 \times 250$	3	LOO
	Beta	162	$352 \times 288 \times 250$	10	LOO
	Gamma	264	$352 \times 288 \times 250$	10	LOO
DynTex++		3600	$50 \times 50 \times 50$	36	50%/50%

Note: LOO and 4fold are leave-one-out and four cross-fold validation respectively. 50%/50% denotes a protocol of taking randomly 50% samples for training and the remain (50%) for testing.

sub-videos for each, i.e., 3600 DTs in total. Following the setting set in [3, 12, 23], a half of samples in each group is randomly selected for training phase and the rest for testing. The final rate is calculated by averaging 10 trials.

3.3 Experimental Results

The specific executions of our proposed descriptor HILOP on different DT datasets are presented in Table 2, in which the highest rates are in bold. It can be validated from this table that encoding DT features in consideration of local relationships on hierarchical regions has pointed out a robust descriptor with promising power, as expected in Sections 2.2 and 2.3. Furthermore, it is also verified that taking into account multi-supporting regions makes the proposed descriptor more discriminative. Specifically, the settings of $|\mathcal{D}| = 4$ and $|\mathcal{D}| = 5$ have reported the best DT recognition rates (see Table 2). Due to the more "stable" performance, $\mathcal{D} = \{(8, 1), (8, 2), (8, 3), (8, 4), (8, 5)\}$ is addressed for a comparison with those of state of the art. In general, the performance of our proposal is more efficient than most of methods (see Table 3), except deep learning techniques using a complex framework for learning DT features. Hereafter, we express in detail the effectiveness of HILOP on the particular DT datasets.

Table 2. Classification rates (%) on DT benchmark datasets.

Descriptor		UCLA				DynTex				Dyn++
$\mathcal{D} = \{(P, \{R\})\}$	#bins	50-LOO	50-4fold	9-class	8-class	Dyn35	Alpha	Beta	Gamma	
$\{(8, \{1, 2\})\}$	1416	98.00	98.50	96.40	93.15	98.00	**98.33**	88.89	92.42	96.05
$\{(8, \{1, 2, 3\})\}$	2832	98.50	98.50	96.95	95.22	98.57	**98.33**	89.51	92.42	96.19
$\{(8, \{1, 2, 3, 4\})\}$	4248	99.00	**99.50**	97.55	**96.41**	99.43	96.67	90.12	**92.80**	96.06
$\{(8, \{1, 2, 3, 4, 5\})\}$	5664	**99.50**	**99.50**	**97.80**	96.30	**99.71**	96.67	**91.36**	92.05	**96.21**

Note: *50-LOO* and *50-4fold* denote results on *50-class breakdown* using leave-one-out and four cross-fold validation. Dyn35 and Dyn++ are shortened for *DynTex35* sub-set and DynTex++ respectively.

UCLA dataset: Thanks to exploiting hierarchical features, the proposed HILOP descriptor obtain promising results on this scenario. More specifically,

Table 3. Comparison of recognition rates (%) on benchmark DT datasets

Group	Dataset	UCLA				DynTex				Dyn++
	Encoding method	50-LOO	50-4fold	9-class	8-class	Dyn35	Alpha	Beta	Gamma	
A	FDT [23]	98.50	99.00	97.70	99.35	98.86	98.33	93.21	91.67	95.31
	FD-MAP [23]	99.50	99.00	99.35	**99.57**	98.86	98.33	92.59	91.67	95.69
B	AR-LDS [35]	89.90^{N}	-	-	-	-	-	-	-	-
	KDT-MD [4]	-	97.50	-	-	-	-	-	-	-
	Chaotic vector [41]	-	-	85.10^{N}	85.00^{N}	-	-	-	-	-
C	3D-OTF [42]	-	87.10	97.23	99.50	96.70	83.61	73.22	72.53	89.17
	WMFS [15]	-	-	97.11	96.96	-	-	-	-	-
	NLSSA [5]	-	-	-	-	-	-	-	-	92.40
	DFS [43]	-	**100**	97.50	99.20	97.16	85.24	76.93	74.82	91.70
	2D+T [8]	-	-	-	-	-	85.00	67.00	63.00	-
	STLS [32]	-	99.50	97.40	99.50	98.20	89.40	80.80	79.80	94.50
D	MBSIF-TOP [3]	99.50^{N}	-	-	-	98.61^{N}	90.00^{N}	90.70^{N}	91.30^{N}	97.12^{N}
	DNGP [34]	-	-	**99.60**	99.40	-	-	-	-	93.80
E	VLBP [44]	-	89.50^{N}	96.30^{N}	91.96^{N}	81.14^{N}	-	-	-	94.98^{N}
	LBP-TOP [44]	-	94.50^{N}	96.00^{N}	93.67^{N}	92.45^{N}	98.33	88.89	84.85^{N}	94.05^{N}
	DDLBP with MJMI [33]	-	-	-	-	-	-	-	-	95.80
	CVLBP [36]	-	93.00^{N}	96.90^{N}	95.65^{N}	85.14^{N}	-	-	-	-
	HLBP [37]	95.00^{N}	95.00^{N}	98.35^{N}	97.50^{N}	98.57^{N}	-	-	-	96.28^{N}
	CLSP-TOP [21]	99.00^{N}	99.00^{N}	98.60^{N}	97.72^{N}	98.29^{N}	95.00^{N}	91.98^{N}	91.29^{N}	95.50^{N}
	MEWLSP [39]	96.50^{N}	96.50^{N}	98.55^{N}	98.04^{N}	99.71^{N}	-	-	-	98.48^{N}
	WLBPC [38]	-	96.50^{N}	97.17^{N}	97.61^{N}	-	-	-	-	95.01^{N}
	CVLBC [45]	98.50^{N}	99.00^{N}	99.20^{N}	99.02^{N}	98.86^{N}	-	-	-	91.31^{N}
	CSAP-TOP [22]	**99.50**	99.50	96.80	95.98	**100**	96.67	92.59	90.53	-
	Our HILOP	**99.50**	99.50	97.80	96.30	99.71	96.67	91.36	92.05	96.21
F	DL-PEGASOS [12]	-	97.50	95.60	-	-	-	-	-	63.70
	Orthogonal Tensor DL [31]	-	99.80	98.20	99.50	-	87.80	76.70	74.80	94.70
	Equiangular Kernel DL [30]	-	-	-	-	-	88.80	77.40	75.60	93.40
	st-TCoF [28]	-	-	-	-	-	**100***	**100***	98.11*	-
	PCANet-TOP [2]	99.50*	-	-	-	-	96.67*	90.74*	89.39*	-
	D3 [14]	-	-	-	-	-	**100***	**100***	98.11*	-
	DT-CNN-AlexNet [1]	-	99.50*	98.05*	98.48*	-	**100***	99.38*	**99.62***	98.18*
	DT-CNN-GoogleNet [1]	-	99.50*	98.35*	99.02*	-	**100***	**100***	**99.62***	**98.58***

Note: "-" means "not available". Superscript "*" indicates results using deep learning algorithms. "N" indicates rates with 1-NN classifier. *50-LOO* and *50-4fold* denote results on *50-class breakdown* using leave-one-out and four cross-fold validation respectively. Dyn35 and Dyn++ are abbreviated for *DynTex35* sub-set and DynTex++ respectively. Evaluations of VLBP and LBP-TOP operators are referred to the evaluations of implementations in [37, 28]. Group A denotes *optical-flow-based approaches*, B: *model-based*, C: *geometry-based*, D: *filter-based*, E: *local-feature-based*, F: *learning-based.*

its performance with the comparing parameters is at 99.5% for both *50-LOO* and *50-4fold*, the highest rates compared to all existing methods, including deep learning approaches, i.e., PCANet-TOP [2] and DT-CNN[1] (see Table 3). In terms of DT recognition on *9-class* and *8-class* schemes, our descriptor gains a promising rate of 97.8% on *9-class*, but just 96.3% on *8-class.* In comparison with the typical LBP-based approaches, its ability is only better than those of VLBP [44] (96.3%,91.96%), LBP-TOP [44] (96%, 93.67%), and CVLBP [36] (96.9%, 95.65%) respectively (see Group E in Table 3). It may be caused by the similarity of turbulent motion properties on regional hierarchies in DT sequences of two schemes. In the meanwhile, other LBP-based variants have better recognition rates, such as CVLBC [45] (99.2%, 99.02%), CLSP-TOP [21] (98.6%, 97.72%), MEWLSP [39] (98.55%, 98.04%), and WLBPC [38] (97.17%, 97.61%) respectively, but most of them have not been validated on the challenging DynTex dataset, except CLSP-TOP. However, the CLSP-TOP's performance in this case is also not better than ours on DynTex (see Group E in Table 3).

DynTex dataset: It can be verified from Table 2 that the performance of HILOP descriptor has been steadily increased along with more hierarchical supporting regions taken into account. For DT classification on *DynTex35*, our obtained rate is 99.71%, the highest in comparison with those of all existing methods, except CSAP-TOP [22] with 100% (see Table 3). However, its dimension is up to over double, 13200 bins compared to 5664 of our (see Table 2), as well as not better than our performance on other schemes, e.g., *9-class* and *8-class* of UCLA. MEWLSP [39] also has the same our ability, but not working well on 50 categories of UCLA and not been validated on the challenging sub-sets of DynTex (i.e., *Alpha*, *Beta*, *Gamma*). In terms of DT recognition on DynTex variants, our proposed framework achieves rates of 96.67%, 91.36%, 92.05% on *Alpha*, *Beta*, and *Gamma* respectively. It can be seen from Table 3 that our results are mostly better than most of state of the art, except deep learning approaches, i.e., st-TCoF [28], D3 [14], and DT-CNN [1] in which DT characteristics are captured by utilizing complicated algorithms in many layers of learning process. It should be noted that this shortcoming has restricted taking them into account mobile applications due to the limited resources of mobile devices.

DynTex++ dataset: It can be observed from Table 2 that our descriptor achieves over 96% on this scheme for all settings of regional hierarchies. With the highest rate of 96.21% for the comparing setting, ours is better than most of shallow methods (see Table 3), except MEWLSP [39], HLBP [37], and MBSIF-TOP [3]. However, as mentioned above, those have either not been verified on DynTex variants (MEWLSP) or not outperformed ours on DynTex (MBSIF-TOP), and on *50-class* of UCLA (MEWLSP, HLBP). Furthermore, learning methods with complex computation also obtain lower performances compared to ours (see Group F in Table 3), except those of using deep learning techniques, i.e., DT-CNN [1] with the best rate of 98.58% involving with GoogleNet architecture.

4 Conclusions

A simple and efficient operator HILOP have been proposed to capture local features in consideration of hierarchical regions surrounding an image pixel. For DT representation, HILOP is involved with video analysis for addressing shape information and motion cues through plane images of a DT sequence. Concatenation of the obtained outputs forms the discriminative descriptor, which has been proved in above experiments for DT classification on different datasets. In the further contexts, it can be exploited HILOP for moment images [20] to make the description more robust against negative impacts of illumination.

References

1. Andrearczyk, V., Whelan, P.F.: Convolutional neural network on three orthogonal planes for dynamic texture classification. Pattern Recognition **76**, 36 – 49 (2018)
2. Arashloo, S.R., Amirani, M.C., Noroozi, A.: Dynamic texture representation using a deep multi-scale convolutional network. JVCIR **43**, 89 – 97 (2017)
3. Arashloo, S.R., Kittler, J.: Dynamic texture recognition using multiscale binarized statistical image features. IEEE Trans. Multimedia **16**(8), 2099–2109 (2014)

4. B. Chan, A.B., Vasconcelos, N.: Classifying video with kernel dynamic textures. In: CVPR. pp. 1–6 (2007)
5. Baktashmotlagh, M., Harandi, M.T., , A., C. Lovell, B.C., Salzmann, M.: Discriminative non-linear stationary subspace analysis for video classification. IEEE Trans. PAMI **36**(12), 2353–2366 (2014)
6. Barmpoutis, P., Dimitropoulos, K., Grammalidis, N.: Smoke detection using spatio-temporal analysis, motion modeling and dynamic texture recognition. In: EUSIPCO. pp. 1078–1082 (2014)
7. Chan, A.B., Vasconcelos, N.: Modeling, clustering, and segmenting video with mixtures of dynamic textures. IEEE Trans. PAMI **30**(5), 909–926 (2008)
8. Dubois, S., Péteri, R., Ménard, M.: Characterization and recognition of dynamic textures based on the 2d+t curvelet transform. Signal, Image and Video Processing **9**(4), 819–830 (2015)
9. Fan, R., Chang, K., Hsieh, C., Wang, X., Lin, C.: LIBLINEAR: A library for large linear classification. JMLR **9**, 1871–1874 (2008)
10. Fazekas, S., Chetverikov, D.: Analysis and performance evaluation of optical flow features for dynamic texture recognition. Sig. Proc.: Image Comm. **22**(7-8), 680–691 (2007)
11. Garrigues, M., Manzanera, A., Bernard, T.M.: Video extruder: a semi-dense point tracker for extracting beams of trajectories in real time. J. Real-Time IP **11**(4), 785–798 (2016)
12. Ghanem, B., Ahuja, N.: Maximum margin distance learning for dynamic texture recognition. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV. LNCS, vol. 6312, pp. 223–236 (2010)
13. Guo, Z., Zhang, L., Zhang, D.: A completed modeling of local binary pattern operator for texture classification. IEEE Trans. IP **19**(6), 1657–1663 (2010)
14. Hong, S., Ryu, J., Im, W., Yang, H.S.: D3: recognizing dynamic scenes with deep dual descriptor based on key frames and key segments. Neurocomputing **273**, 611–621 (2018)
15. Ji, H., Yang, X., Ling, H., Xu, Y.: Wavelet domain multifractal analysis for static and dynamic texture classification. IEEE Trans. IP **22**(1), 286–299 (2013)
16. Liu, L., Zhao, L., Long, Y., Kuang, G., Fieguth, P.W.: Extended local binary patterns for texture classification. Image Vision Comput. **30**(2), 86–99 (2012)
17. Mumtaz, A., Coviello, E., Lanckriet, G.R.G., Chan, A.B.: Clustering dynamic textures with the hierarchical EM algorithm for modeling video. IEEE Trans. PAMI **35**(7), 1606–1621 (2013)
18. Nguyen, T.P., Manzanera, A., Garrigues, M., Vu, N.: Spatial motion patterns: Action models from semi-dense trajectories. IJPRAI **28**(7) (2014)
19. Nguyen, T.P., Manzanera, A., Kropatsch, W.G., N'Guyen, X.S.: Topological attribute patterns for texture recognition. Pattern Recog. Letters **80**, 91–97 (2016)
20. Nguyen, T.P., Vu, N., Manzanera, A.: Statistical binary patterns for rotational invariant texture classification. Neurocomputing **173**, 1565–1577 (2016)
21. Nguyen, T.T., Nguyen, T.P., Bouchara, F.: Completed local structure patterns on three orthogonal planes for dynamic texture recognition. In: IPTA. pp. 1–6 (2017)
22. Nguyen, T.T., Nguyen, T.P., Bouchara, F.: Completed statistical adaptive patterns on three orthogonal planes for recognition of dynamic textures and scenes. J. Electronic Imaging **27**(05), 053044 (2018)
23. Nguyen, T.T., Nguyen, T.P., Bouchara, F., Nguyen, X.S.: Directional beams of dense trajectories for dynamic texture recognition. In: Blanc-Talon, J., Helbert, D., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS. pp. 74–86 (2018)

24. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. PAMI **24**(7), 971–987 (2002)
25. Peh, C., Cheong, L.F.: Synergizing spatial and temporal texture. IEEE Trans. IP **11**(10), 1179–1191 (2002)
26. Péteri, R., Fazekas, S., Huiskes, M.J.: Dyntex: A comprehensive database of dynamic textures. Pattern Recognition Letters **31**(12), 1627–1632 (2010)
27. Péteri, R., Chetverikov, D.: Dynamic texture recognition using normal flow and texture regularity. In: Marques, J.S., de la Blanca, N.P., Pina, P. (eds.) IbPRIA. LNCS, vol. 3523, pp. 223–230 (2005)
28. Qi, X., Li, C.G., Zhao, G., Hong, X., Pietikainen, M.: Dynamic texture and scene classification by transferring deep image features. Neurocomputing **171**, 1230 – 1241 (2016)
29. Qiao, Y., Xing, Z.: Dynamic texture classification using multivariate hidden markov model. IEICE Transactions **101-A**(1), 302–305 (2018)
30. Quan, Y., Bao, C., Ji, H.: Equiangular kernel dictionary learning with applications to dynamic texture analysis. In: CVPR. pp. 308–316 (2016)
31. Quan, Y., Huang, Y., Ji, H.: Dynamic texture recognition via orthogonal tensor dictionary learning. In: ICCV. pp. 73–81 (2015)
32. Quan, Y., Sun, Y., Xu, Y.: Spatiotemporal lacunarity spectrum for dynamic texture classification. CVIU **165**, 85–96 (2017)
33. Ren, J., Jiang, X., Yuan, J., Wang, G.: Optimizing LBP structure for visual recognition using binary quadratic programming. SPL **21**(11), 1346–1350 (2014)
34. Rivera, A.R., Chae, O.: Spatiotemporal directional number transitional graph for dynamic texture recognition. IEEE Trans. PAMI **37**(10), 2146–2152 (2015)
35. Saisan, P., Doretto, G., Wu, Y.N., Soatto, S.: Dynamic texture recognition. In: CVPR. pp. 58–63 (2001)
36. Tiwari, D., Tyagi, V.: Dynamic texture recognition based on completed volume local binary pattern. MSSP **27**(2), 563–575 (2016)
37. Tiwari, D., Tyagi, V.: A novel scheme based on local binary pattern for dynamic texture recognition. CVIU **150**, 58–65 (2016)
38. Tiwari, D., Tyagi, V.: Improved weber's law based local binary pattern for dynamic texture recognition. Multimedia Tools Appl. **76**(5), 6623–6640 (2017)
39. Tiwari, D., Tyagi, V.: Dynamic texture recognition using multiresolution edge-weighted local structure pattern. Computers & Electrical Engineering **62**, 485–498 (2017)
40. Wang, H., Kläser, A., Schmid, C., Liu, C.: Dense trajectories and motion boundary descriptors for action recognition. IJCV **103**(1), 60–79 (2013)
41. Wang, Y., Hu, S.: Chaotic features for dynamic textures recognition. Soft Computing **20**(5), 1977–1989 (2016)
42. Xu, Y., Huang, S.B., Ji, H., Fermüller, C.: Scale-space texture description on sift-like textons. CVIU **116**(9), 999–1013 (2012)
43. Xu, Y., Quan, Y., Zhang, Z., Ling, H., Ji, H.: Classifying dynamic textures via spatiotemporal fractal analysis. Pattern Recognition **48**(10), 3239–3248 (2015)
44. Zhao, G., Pietikäinen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. PAMI **29**(6), 915–928 (2007)
45. Zhao, X., Lin, Y., Heikkilä, J.: Dynamic texture recognition using volume local binary count patterns with an application to 2d face spoofing detection. IEEE Trans. Multimedia **20**(3), 552–566 (2018)
46. Zhao, Y., Huang, D.S., Jia, W.: Completed Local Binary Count for Rotation Invariant Texture Classification. IEEE Trans. IP **21**(10), 4492–4497 (2012)