



HAL
open science

Directional Beams of Dense Trajectories for Dynamic Texture Recognition

Thanh Tuan Nguyen, Thanh Phuong Nguyen, Frédéric Bouchara, Xuan Son Nguyen

► To cite this version:

Thanh Tuan Nguyen, Thanh Phuong Nguyen, Frédéric Bouchara, Xuan Son Nguyen. Directional Beams of Dense Trajectories for Dynamic Texture Recognition. ACIVS, Sep 2018, Poitier, France. ⟨hal-01860067⟩

HAL Id: hal-01860067

<https://hal.science/hal-01860067v1>

Submitted on 22 Aug 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.



HAL Authorization

Directional Beams of Dense Trajectories for Dynamic Texture Recognition

Thanh Tuan Nguyen^{1,2,3}, Thanh Phuong Nguyen^{1,2}, Frédéric Bouchara^{1,2}, and Xuan Son Nguyen⁴

¹ Université de Toulon, CNRS, LIS, UMR 7020, 83957 La Garde, France

² Aix-Marseille Université, CNRS, ENSAM, LIS, UMR 7020, 13397 Marseille, France

³ HCMC University of Technology and Education, Faculty of IT, HCM City, Vietnam

⁴ Université de Caen, CNRS, GREYC, UMR 6072, 14000 Caen, France

Abstract. An effective framework for dynamic texture recognition is introduced by exploiting local features and chaotic motions along beams of dense trajectories in which their motion points are encoded by using a new operator, named LVP_{full}-TOP, based on local vector patterns (LVP) in full-direction on three orthogonal planes. Furthermore, we also exploit motion information from dense trajectories to boost the discriminative power of the proposed descriptor. Experiments on various benchmarks validate the interest of our approach.

Keywords: Dynamic texture, local feature, dense trajectory, LBP, LVP

1 Introduction

Dynamic texture (DT) is a string of textures moving in the temporal domain such as fire, clouds, trees, waves, foliage, blowing flag, fountain, etc. Analysis for “*understanding*” DTs is one of fundamental issues in computer vision tasks. Various approaches have been proposed for DT description. In general, existing methods can be classified into six groups as follows. First, *optical-flow-based methods* [1] are natural approaches for DT recognition thanks to their efficient computation and describing videos in effective ways. Second, *model-based methods* [2,3] have been widely used for DT since the typical model Linear Dynamical System (LDS) [2] was introduced. Third, *filter-based methods* have been also utilized for handling DT recognition. Different filtering operations have been addressed for encoding dynamic features: Binarized Statistical Image Features on Three Orthogonal Planes (BSIF-TOP) [4], Directional Number Transitional Graph (DNG) [5]. Fourth, various *geometry-based methods* have been presented using fractal analysis techniques in which fractal dimension and other fractal characteristics are taken into account for DT representation: dynamic fractal spectrum (DFS) [6], Multi-fractal spectrum (MFS) [7], wavelet-based MFS descriptor [8]. Fifth, owing to outperforming results, *learning-based methods* have recently attracted researchers with promising techniques coming from recent advances in deep learning: Transferred ConvNet Features (TCoF) [9], PCA convolutional network (PCANet-TOP) [10], Dynamic Texture Convolutional Neural

Network (DT-CNN) [9]. Lately, dictionary-learning-based methods [11,12] have also become more popular in which local DT features are figured out by kernel sparse coding. Sixth, *local-feature-based methods* have been also considered with different LBP-based variants owing to their simplicity and efficiency since Zhao et al. [13] proposed two LBP-based variants for DT depiction: Volume LBP (VLBP) and LBP on three orthogonal planes (LBP-TOP). Lately, several efforts based on extensions of these typical operators are addressed to enhance the discriminative power of DT description [14,15,16,17,18].

This paper addresses a new efficient framework using directional beams of dense trajectories for DT representation, in which local feature patterns of motion points are encoded along their trajectories in conjunction with directional features of their neighbors structured by the proposed LVP_{full} -TOP operator with full-direction on three orthogonal planes of sequences. Furthermore, the motion information extracted from dense trajectories is conducted as a complement component to enhance the recognition power of DT descriptor. It could be seen that the advantages of both *optical-flow-based* and *local-feature-based* methods are consolidated into our approach to construct an effective DT descriptor.

2 Related works

LBP-based variants for dynamic texture: An efficient operator, called Local Binary Pattern (LBP), has been introduced in [19] to encode local features of a texture image as a binary chain by regarding relations between the center pixel and its surrounding neighbors interpolated on the neighboring circle centered at this pixel. In order to reduce effectively the dimensionality, different mappings have been proposed to select representative or important patterns: uniform patterns $u2$, $riu2$ [19], topological patterns [20], etc. Inherited by the benefits of LBP for still images, various LBP-based variants have been proposed to inspect DT recognition. At first, Zhao et al. [13] introduced VLBP considering three consecutive frames to form a $(3P + 2)$ -bit pattern for each voxel. An another variant, called LBP-TOP [13], has been also presented to overcome the curse of dimensionality of VLBP by addressing LBP on three orthogonal planes. Various extensions based on two above works have been then proposed to advance the discriminative power: CVLBC [21], CVLBP [15], CLSP-TOP [14], HLBP [16].

Directional LBP-based patterns: The classical LBP captures only the first-order derivative variations. Thus, exploiting higher-order derivative variations is one important approach to develop LBP-based variants for different applications [22,23,24]. Zhang et al. [23] introduced Local Derivative Pattern (LDP), a directional extension of LBP, by taking into account local high-order derivative variations to encode directional patterns of voxels for capturing more robust features. To obtain potential information between derivative directions eliminated in the LDP, Fan et al. [25] proposed Local Vector Pattern (LVP) by regarding the pairwise of directional vectors to remedy the shortcomings remaining in local pattern representation. As adopting LVP as a component in our framework, we

recall LVP in detail hereafter. Let I denote a sub-region of a 2D image. The first-order LVP of the center pixel \mathbf{q}_c conducted by a direction α (in practice, 4 directions are considered, i.e. $\alpha = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$) is calculated as follows.

$$\text{LVP}_{P,R,\alpha}(\mathbf{q}_c) = \{h(V_{\alpha,D}(\mathbf{q}_c), V_{\alpha+45^\circ,D}(\mathbf{q}_c), V_{\alpha,D}(\mathbf{q}_i), V_{\alpha+45^\circ,D}(\mathbf{q}_i))\}_{1 \leq i \leq P} \quad (1)$$

where $V_{\alpha,D}(\mathbf{q}) = I(\mathbf{q}_{\alpha,D}) - I(\mathbf{q})$, known as the first-order LVP, means the directional value of a vector obtained by concerning the current pixel \mathbf{q} with its adjacent neighbor $(\mathbf{q}_{\alpha,D})$ in direction of α ; $D = \{1, 2, 3\}$ presents the distance of the considered pixel with its contiguous points, and the function $h(\cdot)$, called Comparative Space Transform (CST), with four parameters corresponding to four directional vectors in (1) is defined as

$$h(x, y, z, t) = \begin{cases} 1, & \text{if } t - \frac{(y * z)}{x} \geq 0 \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Other formulations of LVP and samples of encoding LVP-based patterns for texture images are clearly discussed in [25].

Dense trajectories: Wang et al. [26] extracted dense trajectories in videos by utilizing a dense optical flow field to sample and track the motion paths of points. For a point $\mathbf{q}_f = (x_f, y_f)$ at the f^{th} frame, its position is tracked into the $(f + 1)^{\text{th}}$ frame by interpolating with a median filter on an optical flow $\omega_f = (u_f, v_f)$, in which u_f and v_f refer to horizontal and vertical of the optical flow component. The new position of \mathbf{q}_f , i.e. \mathbf{q}_{f+1} , at the adjacent frame is inferred as

$$\mathbf{q}_{f+1} = (x_{f+1}, y_{f+1}) = (x_f, y_f) + (M * \omega_f)|_{(\bar{x}_f, \bar{y}_f)} \quad (3)$$

where (\bar{x}_f, \bar{y}_f) means the rounded position value of \mathbf{q}_f , M denotes a median filter kernel of 3×3 pixels. Finally, a trajectory $t = \{\mathbf{q}_f, \mathbf{q}_{f+1}, \dots, \mathbf{q}_{f+L-1}\}$ with length of L is formed by concatenating points of consecutive frames. In our framework, we use the latest version (1.2) of dense trajectories¹ as a tool to extract motion paths of dynamic textures for video representation.

3 Proposed method

3.1 Overview

We introduce an efficient framework taking into account the advantages of two well-known approaches: *optical-flow-based* and *local-feature-based* for an effective DT representation. The main idea is to exploit local features by using the proposed LVP_{full}-TOP operator along dense trajectories together with motion information extracted from them. Fig. 1 graphically illustrates the proposed

¹ http://lear.inrialpes.fr/people/wang/dense_trajectories

framework. Our main contribution is four-fold. First, dense trajectories are used to exploit motions from dynamic textures instead of typical optical-flow-based approaches. Second, an effective operator, called LVP_{full} -TOP, is introduced to capture more second-order derivative variations on three orthogonal planes. Third, a local-feature-based descriptor is presented by addressing LVP_{full} -TOP along a beam of directional trajectories. Fourth, motion angle patterns are utilized to capture more chaotic motions of DT from dense trajectories. In the following, we then detail the proposed method for DT representation.

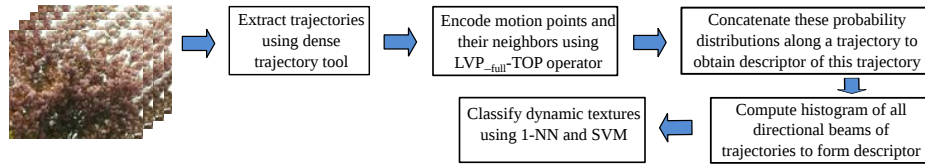


Fig. 1. Illustration of different steps of our proposed framework.

3.2 Components of the proposed descriptor

LVP_{full} -TOP operator: Because only 4 derivative directions of the center pixel are considered in LDP and LVP operators, the relations between the center and other neighbors are less exploited. To remedy this problem, we extend encoding of LVP on full-direction, called LVP_{full} with 8 directions of $\alpha = \{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ\}$, to capture full spatial relations between a pixel \mathbf{q}_c and its neighbors as follows.

$$LVP_{P,R,full}(\mathbf{q}_c) = \sum_{i=0}^{P-1} h(\cdot)2^i \Big|_{\alpha} \quad (4)$$

where $h(\cdot)$ is referred by Equations 1 and 2, P is neighbors of the considered pixel \mathbf{q}_c sampled on a circle with radius R . Due to full-directional encoding, each first-order LVP_{full} code of a pixel has a 64-bit binary pattern in total. It leads to be impossible in real implementation. Therefore, we apply the concept of $u2$ mapping of the typical LBP [19] for LVP^{u2} on each direction to reduce the dimension of descriptor, i.e. $8 \times (P(P-1) + 3)$ bins for full-direction where P is the sampled neighbors.

Furthermore, based on the idea of LBP-TOP [13], we investigate the proposed LVP_{full} on three orthogonal planes to form a new operator, LVP_{full} -TOP which is able to effectively obtain spatio-temporal structures of dynamic features.

Features of directional trajectory: Let $t = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_L, \mathbf{q}_{L+1}\}$ be a trajectory with length of L which is formed by $L + 1$ motion points corresponding

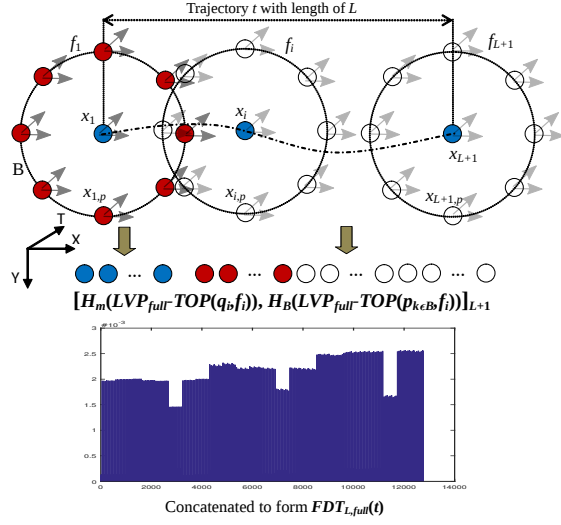


Fig. 2. (Best viewed in color) Encoding FDT with the proposed $LVP_{full-TOP}$ in direction $\alpha = 0^\circ$.

to $L + 1$ consecutive frames $F = \{f_1, f_2, \dots, f_L, f_{L+1}\}$. Fig. 2 graphically demonstrates the proposed method to encode features of directional trajectory (FDT) t in direction $\alpha = 0^\circ$. Accordingly, we consider movements of the motion point $\mathbf{q}_i \in t$ and its neighbors positioned by a vicinity of B in order to compute probability distributions for chaotic motions as well spatial features of \mathbf{q}_i along trajectory t using the proposed $LVP_{full-TOP}$ operator with full-direction. Finally, the gained histograms are concatenated to shape the FDT of t as follows.

$$FDT_{L,full}(t) = \sum_{i=1}^{L+1} \left[H_m(LVP_{full-TOP}(\mathbf{q}_i, f_i)), H_B(LVP_{full-TOP}(\mathbf{p}_{k \in B}, f_i)) \right] \quad (5)$$

where $LVP_{full-TOP}(\mathbf{q}_i, f_i)$ is local vector pattern of a pixel \mathbf{q}_i at the frame f_i based on full-direction in three orthogonal planes; \mathbf{p}_k refers to the k^{th} neighbor of \mathbf{q}_i in the vicinity B ; $H_m(\cdot)$ and $H_B(\cdot)$ are respectively the distributions of the motion point \mathbf{q}_i and its neighbors placed in B .

Motion angle patterns (MAP): To capture motion information from dense trajectories, we extend the idea of [27]. Accordingly, two angles of $\beta = \pi/4$ and $\beta = \pi/8$ can be considered to form 8 and 16 bins of directional motion angle pattern for each pair of motion points respectively. Fig. 3 graphically illustrates a sample of this calculation for a trajectory t , i.e. $MAP_\beta(t)$, using angle of $\beta = \pi/8$ with its length of L located by $L + 1$ motion points in corresponding consecutive frames. These obtained motion angle patterns at point level are concatenated to shape MAP histogram of the whole trajectory.

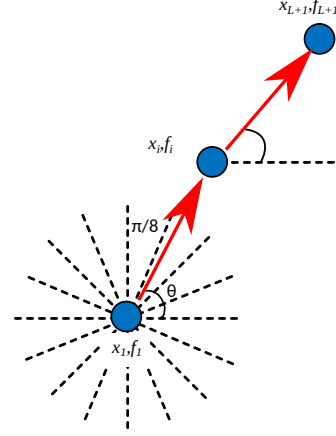


Fig. 3. (Best viewed in color) Computing motion angle patterns for a trajectory.

Proposed video descriptor: Let $T = \{t_1, t_2, \dots, t_n\}$ be a set of trajectories with the same length of L extracted from a video \mathcal{V} using a tool of dense trajectories introduced in [26]. To construct a robust descriptor for DT recognition, we combine and normalize the features of directional beams of T trajectories with the structures of their motion angle patterns, named FD-MAP and defined as follows.

$$\text{FD-MAP}_{L,\beta}(\mathcal{V}) = \sum_{i=1}^n \left[\text{FDT}_{L,\text{full}}(t_i), \text{MAP}_{\beta}(t_i) \right] \quad (6)$$

To investigate the advantage of the angle vector complement component, we also evaluate recognition of $\text{FDT}_{L,\text{full}}(T)$ descriptor on the benchmark DT datasets.

3.3 Classification

For DT recognition, we adopt two classifying algorithms as follows.

Support vector machines (SVMs): We utilize a linear SVM algorithm provided in the LIBLINEAR library [28] to implement a multi-class classifier. All our experiments in this work are inspected by the latest version (2.20) with the default parameters of this tool.

k-nearest neighbors (k-NN): To be comparable with results of existing methods [14,15,16], we employ the simple of k -nearest neighbors, i.e. $k = 1$ (1-NN), in which chi-square (χ^2) measure as dissimilarity measure.

4 Experiments

4.1 Experimental settings

Settings for extracting trajectories: Because most of chaotic motion points in DT sequences are usually “living” in a short period, we investigate dense trajectories with length of $L = \{2, 3\}$ using the tool presented in [26]. Since the default parameters of this tool are set for recognition of human actions in videos, to be suitable for the particular DT characteristics, we changed the original parameter of rejecting trajectory $\text{min_var} = 5 \times 10^{-5}$ in order to obtain “weak” directional trajectories of turbulent motion points. Fig. 4 graphically illustrates trajectories extracted from the corresponding sequences with the customized settings. Empirically, for datasets (like DynTex++) which are built by splitting from other original videos, some of cropped sequences have number of obtained trajectories that are not sufficient for DT representation (see Fig. 4(c)). In this case, a few tracking parameters should be decreased to boost the quantity of trajectories in our framework as $\text{min_distance} = 1$ and $\text{quality} = 10^{-8}$ ².

Parameter settings for descriptors: We use the first-order LVP_{full}-TOP with $D = 1, P = 8, R = 1$ to structure local vector patterns of dynamic features in full-direction on three orthogonal planes. To be compliant with LBP-based methods, encoding spatial patterns of each motion point is adopted with its

² Please see [26] for more details about these above parameters.

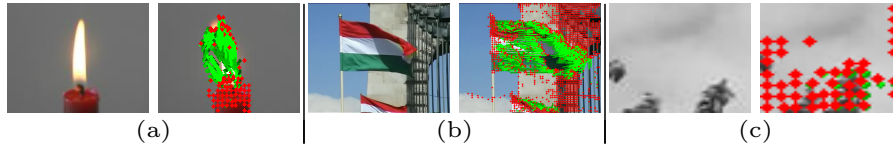


Fig. 4. (Best viewed in color) Samples (a), (b), (c) of directional trajectories extracted from the corresponding videos in UCLA, DynTex, and DynTex++ respectively in which green lines show paths of motion points through the consecutive frames.

neighbors $P_B = 8$ circled by radius $R_B = 1$, i.e. $B = \{P_B, R_B\} = \{8, 1\}$ (see Fig. 2). In this case, the $\text{FDT}_{L,full}$ descriptor has $3 \times 9 \times 8 \times (P(P-1) + 3)$ dimensions with $u2$ mapping utilized for the proposed operator. Furthermore, the angle vector complement with $L \times |\text{MAP}_\beta|$ bins is also employed to form the $\text{FD-MAP}_{L,\beta}$ descriptor with $|\text{FDT}_{L,full}| + L \times |\text{MAP}_\beta|$ bins, where L is length of the considered trajectory, $|\cdot|$ is the size of the descriptor. Regarding the configuration for the DT representation, the best parameter setting is selected as follows to compare with existing approaches: $\text{FD-MAP}_{L,\beta}$ with $L = 2$ and $\beta = \pi/4$, which will be further provided to the classifiers.

4.2 Datasets and experimental protocols

UCLA dataset: UCLA dataset [2] consists of 50 categories with 200 different DT videos, corresponding to four sequences per class, which demonstrate the moving of dynamic textures such as fire, boiling water, fountain, waterfall, flower, and plant. Each original sequence is captured in 75 frames with dimension of 110×160 for each frame. The categories are arranged in varied ways to compose more challenging sub-datasets as follows.

- *50-class:* Original 50 classes are utilized using two experimental protocols: *leave-one-out* (50-LOO) [4,17,29] and *4-fold cross validation* (50-4fold) [14,16].
- *9-class and 8-class:* Original 50 classes of sequences are divided into 9 semantic categories [29,6] consisting of “boiling water” (8), “fire” (8), “flowers” (12), “fountains” (20), “plants” (108), “sea” (12), “smoke” (4), “water” (12), and “waterfall” (16), where the numbers in parentheses take account of sequences in each class. The “plants” category is eliminated from 9-class to form more challenging 8-class scheme [29,6]. For these two schemes, following [29,30,14], a half of DTs is randomly selected for training and the remaining for testing work. The final evaluation of recognition is estimated by the average of rates in 20 runtimes.

DynTex dataset: DynTex dataset [31] consists of more than 650 high-quality DT sequences recorded in various conditions of environment. Following [4,16,13], we use a common version of this dataset containing 35 categories (DynTex35) in which each sequence is randomly cropped into 8 non-overlapping sub-videos that splitting points are not in half of the X, Y, and T axes. In addition, two more sub-sequences are also obtained for the experiment by cutting along the

temporal axis of the original sequence. Consequently, 10 sub-DTs with different spatial-temporal dimension split from each sequence make classification tasks more challenging.

Other popular schemes stated as benchmark sub-datasets for DT recognition are compiled from [31] using leave-one-out as experimental protocol [10,4].

- *Alpha* consists of 60 DT videos equally divided into three categories, i.e. “sea”, “grass”, and “trees”, with 20 sequences in each of them.
- *Beta* contains 162 DT videos grouped into 10 classes with various numbers of sequences for each: “sea”, “vegetation”, “trees”, “flags”, “calm water”, “fountains”, “smoke”, “escalator”, “traffic”, and “rotation”.
- *Gamma* comprises 10 classes with 264 DT videos in total: “flowers”, “sea”, “naked trees”, “foliage”, “escalator”, “calm water”, “flags”, “grass”, “traffic”, and “fountains”. Each of which includes a sample of diverse sequences.

DynTex++ dataset: From more than 650 sequences of the original DynTex, Ghanem et al. [30] filtered 345 raw videos to build DynTex++ in which the filtered videos only contain the main dynamic texture, not consist of other DT features such as panning, zooming, and dynamic background. They are finally divided into 36 classes in which each class has 100 sequences with fixed dimension of $50 \times 50 \times 50$, i.e. 3600 dynamic textures in total. Similarity to [4,30], training set is formed by randomly selecting a half of DTs from each class and the rest for testing. This is repeated 10 times to obtain the average rate as the final result.

4.3 Experimental results

Estimations of our method on the datasets using the proposed descriptors are presented in Tables 1 and 2 respectively, in which the highest classification rates corresponding to the settings are in bold. It is clear from those tables that the combination between the descriptor of directional trajectories FDT and the complement component of motion angle patterns MAP outperforms significantly in comparison with only using FDT descriptor on these datasets. The obtained rates from the observations are then compared with the state-of-the-art approaches through Table 3, in which the highest rates are in bold and results of VLBP [13], LBP-TOP [13] operators are referred to the evaluations of [16,32] in the meanwhile the remains are taken from the original approaches. In general, our proposed FD-MAP descriptor achieves outstanding results in comparison with most of competitive LBP-based approaches on various DT datasets for recognition problems. For comparison with deep learning approaches, our method outperforms significantly on UCLA dataset (see Table 3), but not better on other datasets since deep-learning-based methods use sophisticated learning techniques with a gigantic cost of computation in the meanwhile our framework only concentrates on encoding directional trajectories of DT sequences.

UCLA dataset: It can be observed from Tables 1 and 3 that the proposed descriptor achieves promising results compared to competing methods (both LBP-based and deep-learning-based methods) on all subsets.

Table 1. Classification rates (%) on UCLA using $FDT_{L,full}$ and $FD-MAP_{L,\beta}$

Descriptor	50-LOO		50-4fold		9-class		8-class	
	1-NN	SVM	1-NN	SVM	1-NN	SVM	1-NN	SVM
FDT_2	97.00	99.00	98.50	99.00	97.50	98.75	98.48	98.59
FDT_3	97.00	98.50	98.50	99.00	97.60	97.70	98.91	99.35
$FD-MAP_{2,\pi/4}$	97.00	99.50	98.50	99.00	97.30	99.35	99.02	99.57
$FD-MAP_{2,\pi/8}$	97.00	99.50	98.50	99.50	97.25	99.15	98.70	99.13
$FD-MAP_{3,\pi/4}$	97.50	99.00	99.00	99.00	98.00	99.00	98.59	99.35
$FD-MAP_{3,\pi/8}$	97.50	98.50	99.00	99.00	97.35	99.00	99.13	99.13

Table 2. Rates (%) on DynTex and Dyntax++ with $FDT_{L,full}$ and $FD-MAP_{L,\beta}$

Descriptor	DynTex35		Alpha		Beta		Gamma		DynTex++	
	1-NN	SVM	1-NN	SVM	1-NN	SVM	1-NN	SVM	1-NN	SVM
FDT_2	95.71	98.86	93.33	98.33	84.57	92.59	80.30	91.67	92.63	95.66
FDT_3	96.00	98.86	91.67	98.33	83.95	93.21	80.68	91.67	92.62	95.31
$FD-MAP_{2,\pi/4}$	95.71	98.86	91.67	98.33	84.57	92.59	80.30	91.67	92.87	95.69
$FD-MAP_{2,\pi/8}$	95.71	98.57	91.67	98.33	84.57	92.59	80.30	91.67	93.07	95.56
$FD-MAP_{3,\pi/4}$	96.00	98.57	91.67	98.33	83.95	93.21	81.06	91.29	92.81	95.26
$FD-MAP_{3,\pi/8}$	96.00	98.57	91.67	98.33	83.95	92.59	81.06	91.67	92.55	95.47

50-class: It can be verified that our obtained results (99.5% and 99%) are similar to methods' using deep learning techniques, i.e. PCANet-TOP [10] and DT-CNN [9]. Although our descriptor takes dimension of 12,760 bins, larger than MBSIF-TOP's [4] with 7-scale (5,376 bins), to obtain the same accuracy (99.5%), it is clear that our method is more efficient in other DT datasets (DynTex, DynTex++) in which MBSIF-TOP [4] achieves those with different multi-scale settings. DFS [6], a geometry-based method, gains rate of 100% in 4-fold cross validation scenario, but not perform better ours in other DT datasets.

9-class: Our proposed framework gains the best result of 99.35% compared to all existing approaches including deep-learning-based methods, DT-CNN [9] with rates of 98.05% and 98.35% using AlexNet and GoogleNet architectures respectively, (except DNGP [5], just 0.25% higher than ours).

8-class: On the more challenging 8-class scheme, our method demonstrates outstanding performance with 99.57%, the highest evaluation in comparison with the state-of-the-art approaches while the best rate of DT-CNN [9] just achieves 99.02% by the GoogleNet framework. A geometry-based method, 3D-OTF [7], also obtains nearly same ours but it failed in testing on other DT datasets.

DynTex dataset: It could be seen from Table 3 that our descriptor performs better than all competing LBP-based methods. Furthermore, it is also comparative to deep-learning-based methods such as PCANet-TOP [10], st-TCof [32].

DynTex DynTex35: Our method obtains the best performance on this dataset with 98.86% rate of classification with the selected settings for comparison (see Table 2) compared to most of LBP-based approaches except MEWLSP [18] with

Table 3. Comparison on UCLA, Dyntex, and Dyntex++ datasets.

Dataset Method	UCLA				Dyntex				Dyntex++
	50-LOO	50-4fold	9-class	8-class	Dyntex35	Alpha	Beta	Gamma	
VLBP [13]	-	89.50	96.30	91.96	81.14	-	-	-	94.98
LBP-TOP [13]	-	94.50	96.00	93.67	92.45	96.67	85.80	84.85	94.05
DFS [6]	-	100^s	97.50 ^s	99.00 ^s	97.16 ^s	85.24 ^s	76.93 ^s	74.82 ^s	91.70 ^s
3D-OTF [7]	-	87.10 ^s	97.23 ^s	99.50 ^s	-	82.80 ^s	75.40 ^s	73.50 ^s	89.17 ^s
CLSP-TOP [14]	99.00	99.00	98.60	97.72	98.29	95.00	91.98	91.29	95.5
MBSIF-TOP [4]	99.50	99.50	98.75	97.80	98.61	90.00	90.70	91.30	97.12
MEWLSP [18]	96.50	96.50	98.55	98.04	99.71	-	-	-	98.48
CVLBP [15]	-	93.00	96.90	95.65	85.14	-	-	-	-
HLBP [16]	95.00	95.00	98.35	97.50	98.57	-	-	-	96.28
DNGP [5]	-	-	99.60^s	99.40 ^s	-	-	-	-	93.80 ^s
DDLBP with MJMI [33]	-	-	-	-	-	-	-	-	95.80 ^s
WLBPC [17]	-	96.50	97.17	97.61	-	-	-	-	95.01
CVLBC [21]	98.50	99.00	99.20	99.02	98.86	-	-	-	91.31
Chaotic vector [3]	-	-	85.10	85.00	-	-	-	-	69.00
Orthogonal Tensor DL [12]	-	-	-	-	99.00	87.80 ^s	76.70 ^s	74.80 ^s	94.70 ^s
Equiangular Kernel DL [11]	-	-	-	-	-	88.80 ^s	77.40 ^s	75.60 ^s	93.40 ^s
PCANet-TOP [10]	-	99.50 ^d	-	-	-	96.67 ^d	90.74 ^d	89.39 ^d	-
st-TCoF [32]	-	-	-	-	-	98.33 ^d	98.15 ^d	98.11 ^d	-
DT-CNN-AlexNet [9]	-	99.50 ^d	98.05 ^d	98.48 ^d	-	100^d	99.38 ^d	99.62^d	98.18 ^d
DT-CNN-GoogleNet [9]	-	99.50 ^d	98.35 ^d	99.02 ^d	-	100^d	100^d	99.62^d	98.58^d
Ours	99.50^s	99.00 ^s	99.35 ^s	99.57^s	98.86 ^s	98.33 ^s	92.59 ^s	91.67 ^s	95.69 ^s

Note: Superscript “d” indicates deep-learning methods, “s” is for results using SVM, otherwise using 1-NN; “-” means “not available”.

99.71% but it is not efficient on UCLA as well not verified on other challenging DynTex variants (i.e. *Alpha*, *Beta*, *Gamma*).

Dyntex alpha: Our proposal achieves the highest rate of 98.33% compared to all LBP-based methods; and even outperforms deep learning methods (i. e. PCANet-TOP [10], st-TCoF [32] with rate of 96.67%, 98.33% respectively).

Dyntex beta and gamma: In the best configurations formed for comparison, our result of 92.59% and 91.67% on DynTex beta and gamma respectively shows that our method performs the best compared to all existing approaches except deep learning methods, i.e. st-TCoF [32] and DT-CNN [9] utilizing a complicated computation in training step.

Dyntex++ dataset: It is evident from Table 3 that our proposal gained comparative results in comparison with LBP-based approaches. More specifically, our best result of recognition rate on this scheme is 95.69% (see Table 2), nearly same DDLBP with MJMI [33] (95.8%), the highest rate of recognition on this scheme using SVM classifier. With 92.87% using 1-NN, our descriptor outperforms DNGP [5] and CVLBC [21] over 2% and 1% respectively in the meanwhile it is not better than other LBP-based methods. This may be because DT sequences in DynTex++ dataset, which includes sub-videos split from the original DynTex dataset, comprise lack of directional trajectories (see Fig. 4(c)) for encoding although we reduced the *min.distance* parameter of the tool to minimum value for extracting more trajectories (see more detail in 4.1). In LBP-based approaches, MEWLSP [18] points out the highest rate of 98.48% in this scheme,

even higher than the rate of DT-CNN [9] (98.18%) using AlexNet for learning patterns. However, it is not better than ours on UCLA dataset as well has not been tested on other challenging DynTex variants (i.e. *Alpha*, *Beta*, *Gamma*). Deep learning methods [9] have outstanding results but they take a long time to learn features with a huge complex computation.

5 Conclusions

An effective framework for DT representation has been proposed in this paper in which directional trajectories extracted from DT sequences are encoded by the proposed operator LVP_{full} -TOP, an extension of local vector pattern operator to full-direction on three orthogonal planes in order to exploit the entire reactions between motion points and their neighbors. In addition, we also introduced motion angle patterns of directional trajectories to capture the angle vector of pairs of their motion points as a complementary feature for texture representation in order to make the proposed descriptor more robust. Evaluations on different DT datasets have demonstrated that the proposed framework significantly outperforms recent state-of-the-art results. A combination between LVP and a filtering technique [34] will be the subject of a future work.

References

1. Peh, C., Cheong, L.F.: Synergizing spatial and temporal texture. *IEEE Trans. IP* **11**(10) (2002) 1179–1191
2. Saisan, P., Doretto, G., Wu, Y.N., Soatto, S.: Dynamic texture recognition. In: *CVPR*. (2001) 58–63
3. Wang, Y., Hu, S.: Chaotic features for dynamic textures recognition. *Soft Computing* **20**(5) (2016) 1977–1989
4. Arashloo, S.R., Kittler, J.: Dynamic texture recognition using multiscale binarized statistical image features. *IEEE Trans. Multimedia* **16**(8) (2014) 2099–2109
5. Rivera, A.R., Chae, O.: Spatiotemporal directional number transitional graph for dynamic texture recognition. *IEEE Trans. PAMI* **37**(10) (2015) 2146–2152
6. Xu, Y., Quan, Y., Zhang, Z., Ling, H., Ji, H.: Classifying dynamic textures via spatiotemporal fractal analysis. *Pattern Recognition* **48**(10) (2015) 3239–3248
7. Xu, Y., Huang, S.B., Ji, H., Fermüller, C.: Scale-space texture description on sift-like textons. *CVIU* **116**(9) (2012) 999–1013
8. Ji, H., Yang, X., Ling, H., Xu, Y.: Wavelet domain multifractal analysis for static and dynamic texture classification. *IEEE Trans. IP* **22**(1) (2013) 286–299
9. Andrearczyk, V., Whelan, P.F.: Convolutional neural network on three orthogonal planes for dynamic texture classification. *Pattern Recognition* **76** (2018) 36 – 49
10. Arashloo, S.R., Amirani, M.C., Noroozi, A.: Dynamic texture representation using a deep multi-scale convolutional network. *JVCIR* **43** (2017) 89 – 97
11. Quan, Y., Bao, C., Ji, H.: Equiangular kernel dictionary learning with applications to dynamic texture analysis. In: *CVPR*. (2016) 308–316
12. Quan, Y., Huang, Y., Ji, H.: Dynamic texture recognition via orthogonal tensor dictionary learning. In: *ICCV*. (2015) 73–81

13. Zhao, G., Pietikäinen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. PAMI* **29**(6) (2007) 915–928
14. Nguyen, T.T., Nguyen, T.P., Bouchara, F.: Completed local structure patterns on three orthogonal planes for dynamic texture recognition. In: *IPTA*. (2017) 1–6
15. Tiwari, D., Tyagi, V.: Dynamic texture recognition based on completed volume local binary pattern. *MSSP* **27**(2) (2016) 563–575
16. Tiwari, D., Tyagi, V.: A novel scheme based on local binary pattern for dynamic texture recognition. *CVIU* **150** (2016) 58–65
17. Tiwari, D., Tyagi, V.: Improved weber’s law based local binary pattern for dynamic texture recognition. *MTA* **76**(5) (2017) 6623–6640
18. Tiwari, D., Tyagi, V.: Dynamic texture recognition using multiresolution edge-weighted local structure pattern. *Computers & Electrical Engineering* **62** (2017) 485–498
19. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. PAMI* **24**(7) (2002) 971–987
20. Nguyen, T.P., Manzanera, A., Kropatsch, W.G., N’Guyen, X.S.: Topological attribute patterns for texture recognition. *Pattern Recog. Letters* **80** (2016) 91–97
21. 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) (2018) 552–566
22. Nguyen, X.S., Nguyen, T.P., Charpillat, F., Vu, N.S.: Local derivative pattern for action recognition in depth images. *MTA* **77**(7) (2018) 8531–8549
23. Zhang, B., Gao, Y., Zhao, S., Liu, J.: Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor. *IEEE Trans. IP* **19**(2) (2010) 533–544
24. Nguyen, X.S., Mouaddib, A.I., Nguyen, T.P., Jeanpierre, L.: Action recognition in depth videos using hierarchical gaussian descriptor. *MTA* (2018)
25. Fan, K., Hung, T.: A novel local pattern descriptor - local vector pattern in high-order derivative space for face recognition. *IEEE Trans. IP* **23**(7) (2014) 2877–2891
26. Wang, H., Kläser, A., Schmid, C., Liu, C.: Dense trajectories and motion boundary descriptors for action recognition. *IJCV* **103**(1) (2013) 60–79
27. Nguyen, T.P., Manzanera, A., Garrigues, M., Vu, N.: Spatial motion patterns: Action models from semi-dense trajectories. *IJPRAI* **28**(7) (2014)
28. Fan, R., Chang, K., Hsieh, C., Wang, X., Lin, C.: LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research* **9** (2008) 1871–1874
29. Ravichandran, A., Chaudhry, R., Vidal, R.: View-invariant dynamic texture recognition using a bag of dynamical systems. In: *CVPR*. (2009) 1651–1657
30. Ghanem, B., Ahuja, N.: Maximum margin distance learning for dynamic texture recognition. In: *ECCV*. Volume 6312 of LNCS. (2010) 223–236
31. Péteri, R., Fazekas, S., Huiskes, M.J.: Dyntex: A comprehensive database of dynamic textures. *Pattern Recognition Letters* **31**(12) (2010) 1627–1632
32. Qi, X., Li, C.G., Zhao, G., Hong, X., Pietikainen, M.: Dynamic texture and scene classification by transferring deep image features. *Neurocomputing* **171** (2016) 1230 – 1241
33. Ren, J., Jiang, X., Yuan, J., Wang, G.: Optimizing LBP structure for visual recognition using binary quadratic programming. *IEEE Signal Processing Letters* **21**(11) (2014) 1346–1350
34. Nguyen, T.P., Vu, N., Manzanera, A.: Statistical binary patterns for rotational invariant texture classification. *Neurocomputing* **173** (2016) 1565–1577