# Rotation Invariant Image and Video Description with Local Binary Pattern Features

Guoying Zhao, Timo Ahonen, Jiří Matas, *Member, IEEE* and Matti Pietikäinen, *Senior Member, IEEE*

***Abstract*—In this paper, we propose a novel approach to compute rotation invariant features from histograms of local, non-invariant patterns. Here we apply this approach to both static and dynamic Local Binary Pattern descriptors. For static texture description, we present Local Binary Pattern Histogram Fourier features (LBP-HF), and for dynamic texture recognition, two rotation invariant descriptors computed from the LBP-TOP (Local Binary Patterns from Three Orthogonal Planes) features in the spatiotemporal domain. LBP-HF is a novel rotation invariant image descriptor computed from discrete Fourier transforms of local binary pattern (LBP) histograms. The approach can also be generalized to embed any uniform features into this framework and combining the supplementary information, e.g. sign and magnitude components of LBP together can improve the description ability. Moreover, two variants of rotation invariant descriptors are proposed to the LBP-TOP, which is an effective descriptor for dynamic texture recognition as shown by its recent success in different application problems, but it is not rotation invariant. In the experiments, it is shown that LBP-HF and its extensions outperform non-invariant and earlier versions of rotation invariant LBP in rotation invariant texture classification. In experiments on two dynamic texture databases with rotations or view variations, the proposed video features can effectively deal with rotation variations of dynamic textures. They also are robust with respect to changes in viewpoint, outperforming recent methods proposed for view-invariant recognition of dynamic textures.**

***Index Terms*—Rotation invariance, feature, classification, texture, dynamic texture, LBP, Fourier transform.**

## I. Introduction

TEXTURE analysis is a basic vision problem [26], [30] with application in many areas, e.g. object recognition, remote sensing, and content-based image retrieval. In many practical applications, textures are captured in arbitrary orientations.

For static textures, rotation invariant features are independent of the angle of the input texture image [22], [27], [30]. Robustness to image conditions such as illumination is often required/desirable. Describing the appearance locally, e.g., using co-occurrences of gray values or with filter bank responses and then forming a global description by computing statistics over the image region is a well established technique [26]. This approach has been extended by several authors to produce rotation invariant features by transforming each local descriptor to a canonical representation invariant to rotations of the input image [2], [22], [27]. The statistics describing the whole region are then computed from these transformed local descriptors. The published work on rotation invariant texture analysis is extensive.

We have chosen to build our rotation invariant texture descriptor on LBP. LBP is an operator for image description that is based on the signs of differences of neighboring pixels. It is fast to compute and invariant to monotonic gray-scale changes of the image. Despite being simple, it is very descriptive, which is attested by the wide variety of different tasks it has been successfully applied to. The LBP histogram has proven to be a widely applicable image feature for, e.g., texture classification, face analysis, video background subtraction, and interest region description. LBPs have been used for rotation invariant texture recognition before. The original one is in [22], where the neighboring $n$ binary bits around a pixel are clockwise rotated $n$ times that a maximal number of the most significant bits is used to express this pixel. The more recent dominant local binary pattern (DLBP) method [18], which makes use of the most frequently occurred patterns to capture descriptive textural information, has also the rotation invariant characteristics. Guo et al. developed an adaptive LBP (ALBP) [13] by incorporating the directional statistical information for rotation invariant texture classification. In [14], LBP variance (LBPV) was proposed to characterize the local contrast information into the one-dimensional LBP histogram. The performance evaluation using rotation invariant LBP, Coordinated Clusters Representation and Improved LBP was conducted on granite texture classification [11]. The sign and magnitude of LBP, and the binary code of intensity of center pixels were combined together in CLBP [15] to improve the texture classification. But the intensity information is very sensitive to illumination changes, so this method needs image normalization to remove global intensity effects before feature extraction.

Dynamic textures (DT) are image sequences with visual pattern repetition in time and space, like sea-waves, smoke, foliage, fire, shower and whirlwind. For DT analysis, feature description is the key element. Local features have been obtaining increasing attention due to their ability of using micro textons to describe the motions, while there is an argument against global spatiotemporal transforms on the difficulty to provide rotation invariance [7]. Dynamic textures in video sequences can be arbitrarily oriented. The rotation can be



G. Zhao and M. Pietikäinen are with the Center for Machine Vision Research, Department of Computer Science and Engineering, University of Oulu, P. O. Box 4500, FI-90014, Finland. E-mail: {gyzhao, mkp}@ee.oulu.fi.

T. Ahonen was with the Center for Machine Vision Research, University of Oulu, P. O. Box 4500, FI-90014, Finland and is currently with the Nokia Research Center, Palo Alto, California, USA. Email: timo.ahonen@nokia.com

J. Matas is with the Center for Machine Perception, Dept. of Cybernetics, Faculty of Elec. Eng., Czech Technical University in Prague. Email: matas@cmp.felk.cvut.cz.

Fig. 1. (a)(b):Rotations of static textures and (c) a different texture.

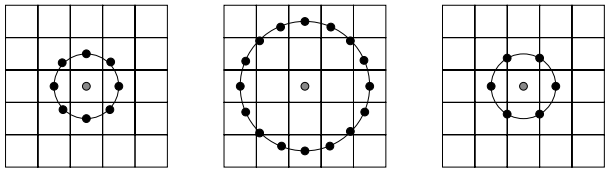

Fig. 2. Three circular neighborhoods: (8,1), (16,2), (6,1). The pixel values are bilinearly interpolated whenever the sampling point is not in the center of a pixel.

caused by the rotation of cameras and the self-rotation of the captured objects. Rotation invariant DT analysis is an important but still open research problem. It aims at providing features that are invariant to rotation angle of the input texture image sequences along the time axis. Moreover, these features should also capture the appearance and motions, as well as be robust to other challenges such as illumination changes, and allow multi-resolution analysis. Fazekas and Chetverikov [10] studied the normal flow and complete flow features for DT classification. Their features are rotation-invariant, and the results on ordinary DT without rotations are promising. Lu et al. proposed a method using spatiotemporal multi-resolution histograms based on velocity and acceleration fields [19]. Velocity and acceleration fields of different spatio-temporal resolution image sequences are accurately estimated by the structure tensor method. This method is also rotation-invariant and provides local directionality information. But both of these methods cannot deal with illumination changes and did not consider the multi-scale properties of DT. Even though there are some methods which are rotation invariant in theory, like [10], [19], but to our best knowledge, there are very few results reported about their performance evaluation using rotated sequences.

The main contribution of this paper is the observation that invariants constructed globally for the whole region by histogramming non-invariant are superior to most other histogram based invariant texture descriptors which normalize rotation locally. In [14], the authors also considered how to get a rotation invariant strategy from non-rotation invariant histograms. Our approach is different from that. The method in [14] keeps the original rotation variant features, but finds a match strategy to deal with rotation. Our method generates new features from rotation variant features and does not need any special match strategy.

Most importantly, as each local descriptor (e.g., filter bank response) is transformed to canonical representation independently, the relative distribution of different orientations is lost. In Fig. 1, (a) and (b) represent different rotations of the same texture, whereas (c) is clearly a different texture. Considering the case that each texture element (black bar) is rotated into canonical orientation independently, (a) and (b) will correctly get the same representation, but also the difference between textures (a) and (c) will be lost. Furthermore, as the transformation needs to be performed for each texton, it must be computationally simple if the overall computational cost needs to be low.

We apply this idea to static texture recognition (Sections II-III) and dynamic texture recognition (Sections IV-VI). Preliminary results for static texture recognition were presented in [1].

On the basis of LBP, we propose novel Local Binary Pattern Histogram Fourier features (LBP-HF) for static texture recognition. LBP-HF is a rotation invariant image descriptor based on uniform Local Binary Patterns (LBP) [22]. Unlike the earlier local rotation invariant features which are histograms of rotation-invariant version of LBPs, the LBP-HF descriptor is formed by first computing a non-invariant LBP histogram over the whole region and then constructing rotationally invariant features from the histogram. This means that rotation invariance is attained globally, and the features are thus invariant to rotations of the whole input signal but they still retain information about relative distribution of different orientations of uniform local binary patterns. Again, considering Fig. 1, if rotation is compensated for globally, textures (a) and (b) get the same description, but the difference between (a) and (c) is retained. In addition, this approach is generalized to embed any uniform features, e.g. sign and magnitude components of LBP, into this framework to improve the description ability.

Later, this idea is extended to the spatiotemporal domain: two variants of rotation invariant LBP-TOP operators are developed and the experiments on two databases show their effectiveness for rotation variations and view changes in dynamic texture recognition.

## II. Rotation Invariant Image Descriptors

In this section, we will focus on the rotation invariant image features for static texture description. Because it is based on uniform local binary pattern, first, the LBP methodology is briefly reviewed.

### A. Local Binary Pattern Operator

The local binary pattern operator [22] is a powerful means of texture description. The original version of the operator labels the image pixels by thresholding the $3\times3$-neighborhood of each pixel with the center value and summing the thresholded values weighted by powers of two.

The operator can also be extended to use neighborhoods of different sizes [22] (See Fig.2). To do this, a circular neighborhood denoted by $(P, R)$ is defined. Here $P$ represents the number of sampling points and $R$ is the radius of the neighborhood. These sampling points around pixel $(x, y)$ lie at coordinates $(x_p, y_p) = (x + R\cos(2\pi p/P), y - R\sin(2\pi p/P))$. When a sampling point does not fall at integer coordinates, the pixel value is bilinearly interpolated. Now the LBP label for the

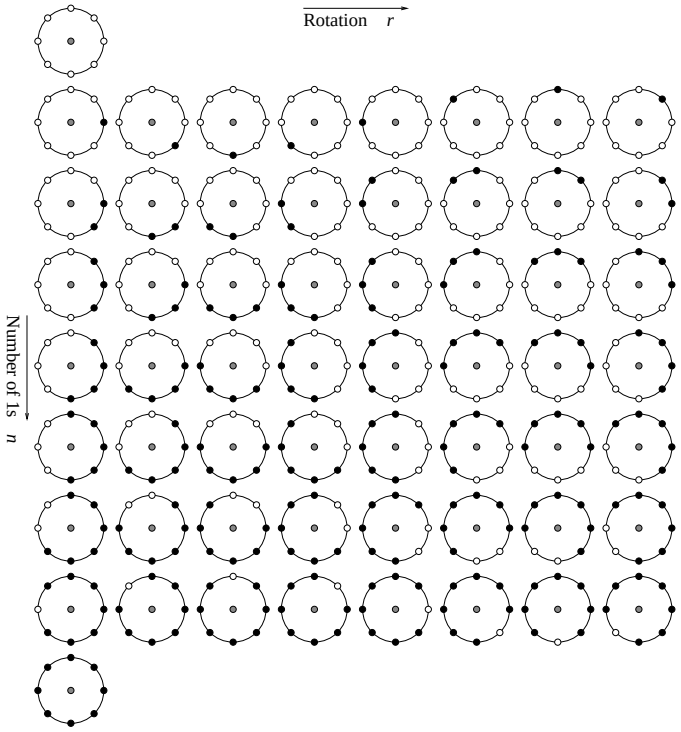


Fig. 3. The 58 different uniform patterns in (8,R) neighborhood

.

center pixel $(x, y)$ of image $f(x, y)$ is obtained through

$$LBP_{P,R}(x, y) = \sum_{p=0}^{P-1} s(f(x, y) - f(x_p, y_p))2^p, \quad (1)$$

where $s(z)$ is the thresholding function

$$s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases} \quad (2)$$

Further extensions to the original operator are so called *uniform* patterns [22]. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is considered circular. In the computation of the LBP histogram, uniform patterns are used so that the histogram has a separate bin for every uniform pattern and all non-uniform patterns are assigned to a single bin. The 58 possible uniform patterns in neighborhood of 8 sampling points are shown in Fig. 3.

The original rotation invariant LBP operator based on uniform patterns, denoted here as $\text{LBP}^{riu2}$, is achieved by circularly rotating each bit pattern to the minimum value. For instance, the bit sequences 10000011, 11100000 and 00111000 arise from different rotations of the same local pattern and they all correspond to the normalized sequence 00000111. In Fig. 3 this means that all the patterns from one row are replaced with a single label.

### *B. Rotation Invariant Descriptors from LBP Histograms for Static Texture Analysis*

Let us denote a specific uniform LBP pattern by $U_P(n, r)$. The pair $(n, r)$ specifies an uniform pattern so that $n$ is the number of 1-bits in the pattern (corresponds to row number in Fig. 3) and $r$ is the rotation of the pattern (column number in Fig. 3). Table I lists the notations and corresponding meanings used in this section.

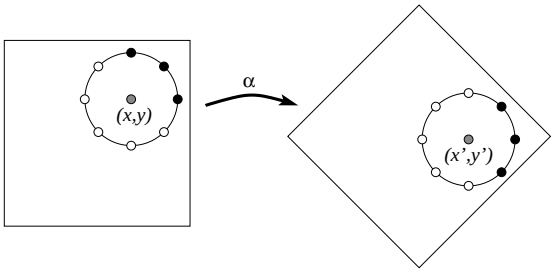


Fig. 4. Effect of image rotation on points in circular neighborhoods

Now if the neighborhood has $P$ sampling points, $n$ gets values from 0 to $P + 1$, where $n = P + 1$ is the special label marking all the non-uniform patterns. Furthermore, when $1 \leq n \leq P - 1$, the rotation of the pattern is in the range $0 \leq r \leq P - 1$.

Let $I^{\alpha^\circ}(x, y)$ denote the rotation of image $I(x, y)$ by $\alpha$ degrees. Under this rotation, point $(x, y)$ is rotated to location $(x', y')$. If we place a circular sampling neighborhood on points $I(x, y)$ and $I^{\alpha^\circ}(x', y')$, we observe that it also rotates by $\alpha^\circ$. See Fig. 4.

If the rotations are limited to integer multiples of the angle between two sampling points, *i.e.* $\alpha = a\frac{360^\circ}{P}, a = 0, 1, \ldots, P-1$, this rotates the sampling neighborhood exactly by $a$ discrete steps. Therefore the uniform pattern $U_P(n, r)$ at point $(x, y)$ is replaced by uniform pattern $U_P(n, r + a \mod P)$ at point $(x', y')$ of the rotated image.

Now consider the uniform LBP histograms $h_I(U_P(n, r))$. The histogram value $h_I$ at bin $U_P(n, r)$ is the number of occurrences of uniform pattern $U_P(n, r)$ in image $I$.

If the image $I$ is rotated by $\alpha = a\frac{360^\circ}{P}$, based on the reasoning above, this rotation of the input image causes a cyclic shift in the histogram along each of the rows,

$$h_{I^{\alpha^\circ}}(U_P(n, r + a \mod P)) = h_I(U_P(n, r)) \quad (3)$$

For example, in the case of eight neighbor LBP, when the input image is rotated by $45^\circ$, the value from histogram bin $U_8(1, 0) = 00000001b$ moves to bin $U_8(1, 1) = 00000010b$, value from bin $U_8(1, 1)$ to bin $U_8(1, 2)$, etc.

Based on the property, which states that rotations induce shift in the polar representation $(P, R)$ of the neighborhood, we propose a class of features that are invariant to rotation of the input image, namely such features, computed along the input histogram rows, that are invariant to cyclic shifts.

We use the Discrete Fourier Transform to construct these features. Let $H(n, \cdot)$ be the DFT of $n$th row of the histogram $h_I(U_P(n, r))$, *i.e.*

$$H(n, u) = \sum_{r=0}^{P-1} h_I(U_P(n, r))e^{-i2\pi ur/P}. \quad (4)$$

TABLE I
NOTATIONS AND THEIR CORRESPONDING MEANINGS IN THE DESCRIPTION OF LBP-HF.

| Notations | Meaning |
|---|---|
| $U_P(n,r)$ | uniform LBP pattern |
| $n$ | number of 1-bits in the pattern |
| $r$ | rotation of the pattern |
| $P$ | number of neighboring sampling points |
| $a$ | discrete steps of rotation |
| $h_I(U_P(n,r))$ | number of occurrences of uniform pattern $U_P(n,r)$ in image $I$ |
| $H(n,\cdot)$ | DFT of nth row of the histogram $h_I(U_P(n,r))$ |
| $\overline{H(n_2,u)}$ | complex conjugate of $H(n_2,u)$ |
| $\lvert H(n,u)\rvert$ | Fourier magnitude spectrum |

Now for DFT it holds that a cyclic shift of the input vector causes a phase shift in the DFT coefficients. If $h'(U_P(n,r)) = h(U_P(n,r-a))$, then

$$H'(n,u) = H(n,u)e^{-i2\pi ua/P}, \tag{5}$$

and therefore, with any $1 \le n_1, n_2 \le P-1$,

$$\begin{aligned} H'(n_1,u)\overline{H'(n_2,u)} &= H(n_1,u)e^{-i2\pi ua/P}\overline{H(n_2,u)}e^{i2\pi ua/P} \\ &= H(n_1,u)\overline{H(n_2,u)}, \end{aligned} \tag{6}$$

where $\overline{H(n_2,u)}$ denotes the complex conjugate of $H(n_2,u)$.

This shows that with any $1 \le n_1, n_2 \le P-1$ and $0 \le u \le P-1$, the features

$$\text{LBP}^{u2}\text{-HF}(n_1,n_2,u) = H(n_1,u)\overline{H(n_2,u)}, \tag{7}$$

are invariant to cyclic shifts of the rows of $h_I(U_P(n,r))$ and consequently, they are invariant also to rotations of the input image $I(x,y)$. The Fourier magnitude spectrum which we call LBP Histogram Fourier (LBP-HF) features,

$$|H(n,u)| = \sqrt{H(n,u)\overline{H(n,u)}} \tag{8}$$

can be considered a special case of these features. Furthermore it should be noted that the Fourier magnitude spectrum contains $\text{LBP}^{riu2}$ features as a subset, since

$$|H(n,0)| = \sum_{r=0}^{P-1} h_I(U_P(n,r)) = h_{\text{LBP}^{riu2}}(n). \tag{9}$$

An illustration of these features is in Fig. 5

Actually, the LBP-HF can be thought as a general framework, in which $U_P(n,r)$ in Eq. 4 does not have to be the occurrence of that pattern, and instead it can be any features corresponding to that uniform pattern. As long as the features are organized in the same way to uniform patterns so that they satisfy the Eq. 3, they can be embedded into Eq. 4 to replace $U_P(n,r)$ and generate new rotation invariant descriptors. One example is CLBP [15]. CLBP contains the sign-LBP (sign of the difference of neighboring pixel against central pixel, i.e. it equals LBP) and magnitude-LBP (the magnitude of the difference of neighboring pixel against central pixel) components. Sign LBP can be calculated using Eq. 1 and magnitude LBP can be obtained using the following equation [15]:

$$LBP\_M_{P,R}(x,y) = \sum_{p=0}^{P-1} s(|f(x,y)-f(x_p,y_p)|-c)2^p, \tag{10}$$

where $c$ is a threshold to be determined adaptively. It can be set as the mean value of $|f(x,y)-f(x_p,y_p)|$ from the whole image.

Both two parts can also be organized into uniform sign-LBP (equal to the uniform LBP) and uniform magnitude-LBP components. We can embed these two parts to Eq. 4 separately, concatenate the produced histogram Fourier features and obtain the $LBPHF\_S\_M$.

## III. EXPERIMENTS ON STATIC TEXTURE CLASSIFICATION

For static textures, we carried out experiments on Outex_TC_00012 database [23] for rotation invariant texture classification. Experiments with other databases are presented in [1].

The proposed rotation invariant LBP-HF features were compared against non-invariant $LBP^{u2}$, and the original rotation invariant version $LBP^{riu2}$. To show the generalization of LBP-HF, we also put $LBPHF\_M$ and $LBPHF\_S\_M$ into comparison.

Table II lists the abbreviation of the methods used in comparison and their corresponding meaning.

For a fair comparison, we used the Chi-square metric, since many previous works, e.g. [14], [15], also used it, assigning a sample to the class of model minimizing the $L_{Chi}$ distance

$$L_{Chi}(h^S,h^M) = \sum_{b=1}^{B}(h^S(b)-h^M(b))^2/(h^S(b)+h^M(b)), \tag{11}$$

where $h^S(b)$ and $h^M(b)$ denote the bin $b$ of sample and model features, respectively.

Implementation of LBP-HF Features for Matlab can be found in http://www.cse.oulu.fi/MVG/Downloads/LBPMatlab. The feature vectors are of the following form:

$$\begin{aligned} fv_{\text{LBP-HF}} = \; & [|H(1,0)|,\ldots,|H(1,P/2)|, \\ & \ldots, \\ & |H(P-1,0)|,\ldots,|H(P-1,P/2)|, \\ & h(U_P(0,0)),h(U_P(P,0)),h(U_P(P+1,0))]. \end{aligned}$$

We derived from the setup of [22] by using nearest neighbor (NN) classifier instead of 3NN because no significant performance difference between the two was observed.

We evaluated our methods on Outex database. Rotation variation is common in captured images. Outex database is widely used to evaluate texture methods for dealing with rotation variations [13], [14], [15], [22].

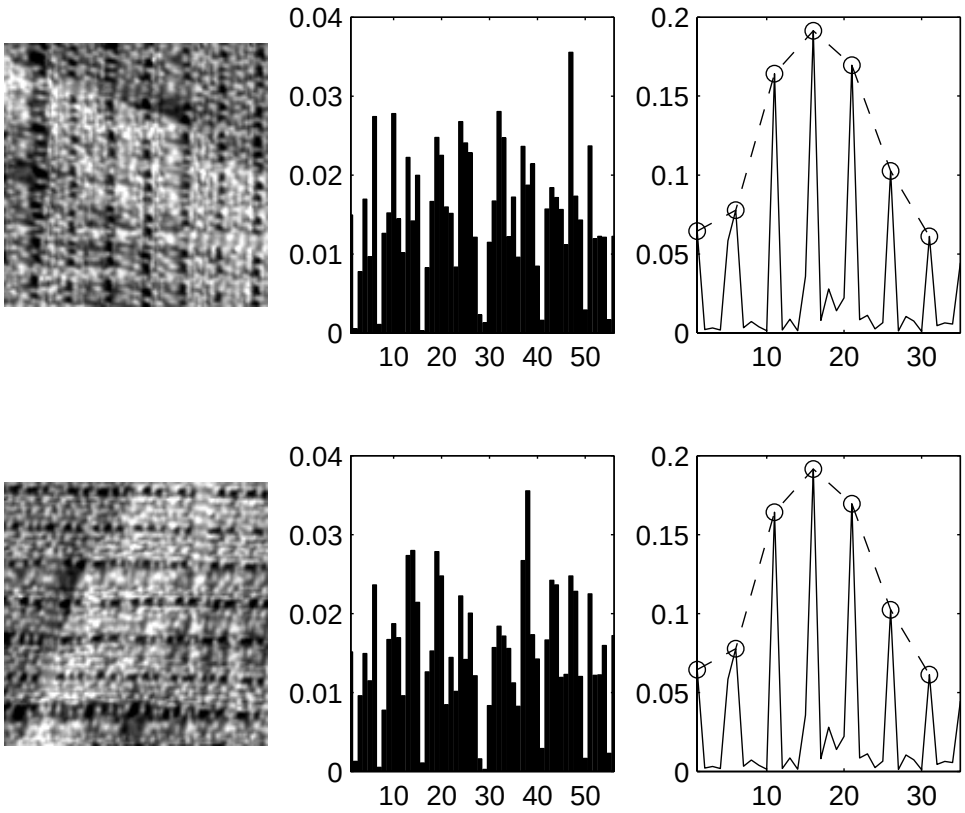


Fig. 5. 1st column: Texture image at orientations 0° and 90°. 2nd column: bins 1–56 of the corresponding LBP$^{u2}$ histograms. 3rd column: Rotation invariant features $|H(n,u)|, 1 \leq n \leq 7, 0 \leq u \leq 5$, (solid line) and LBP$^{riu2}$ (circles, dashed line). Note that the LBP$^{u2}$ histograms for the two images are markedly different, but the $|H(n,u)|$ features are nearly equal.

TABLE II
THE ABBREVIATION OF THE METHODS IN EXPERIMENTS AND THEIR CORRESPONDING MEANING.

| Abbreviation | Method |
|---|---|
| $LBP^{u2}$ | Uniform sign LBP |
| $LBP^{riu2}$ | Rotation invariant uniform sign LBP |
| $LBP\_M^{u2}$[15] | Uniform magnitude LBP |
| $LBP\_M^{riu2}$[15] | Rotation invariant uniform magnitude LBP |
| $LBP\_S\_M^{riu2}$ | Concatenation of Rotation invariant uniform sign LBP and magnitude LBP |
| $LBP-HF$ | Uniform LBP histogram Fourier |
| $LBPHF\_M$ | Uniform magnitude LBP histogram Fourier |
| $LBPHF\_S\_M$ | Concatenation of sign LBP histogram Fourier and magnitude LBP histogram Fourier |

We used the Outex_TC_00012 [23] test set intended for testing rotation invariant texture classification methods. This test set consists of 9120 images representing 24 different textures imaged under different rotations and lightings. The test set contains 20 training images for each texture class. The training images are under single orientation whereas different orientations are present in the total of 8640 testing images. We report here the total classification rates over all test images.

The results of the experiment are shown in Table III. As we can observe, rotation invariant features provide better classification rates than noninvariant features (here they are $LBP^{u2}$ and $LBP\_M^{u2}$). The performance of LBP-HF features and LBPHF_M is clearly higher than that of $LBP^{u2}$, $LBP^{riu2}$ and $LBP\_M^{u2}$, $LBP\_M^{riu2}$. When combining LBP-HF with magnitude-LBP together (LBPHF_S_M), much better results are obtained, (i.e. 0.949 for LBPHF_S_M with 24 neighboring points and radius three) than for all the other methods. By comparing the results of 1st column with 4th column, 2nd column with 5th column and 3rd column with 6th column, we can see that sign information usually plays more important roles than magnitude information, which is also consistent to the analysis in [15].

By varying $(P,R)$, multi-resolution analysis can be utilized to get improved classification accuracy, e.g. $LBPHF\_S\_M_{16,2+24,3}$ can achieve 96.2% accuracy. Similar improvement can be seen in [1], [13], [14], [15]. Moreover, the

TABLE III
TEXTURE RECOGNITION RATES ON OUTEX_TC_00012 DATASET.

| (P,R) | $LBP^{u2}$ | $LBP^{riu2}$ | $LBP-HF$ | $LBP\_M^{u2}$ | $LBP\_M^{riu2}$ | $LBPHF\_M$ | $LBP\_S\_M^{riu2}$ | $LBPHF\_S\_M$ |
|---|---|---|---|---|---|---|---|---|
| $(8,1)$ | 0.569 | 0.646 | **0.741** | 0.496 | 0.610 | **0.622** | 0.714 | **0.786** |
| $(16,2)$ | 0.589 | 0.789 | **0.903** | 0.567 | 0.731 | **0.856** | 0.860 | **0.940** |
| $(24,3)$ | 0.569 | 0.830 | **0.924** | 0.594 | 0.799 | **0.874** | 0.904 | **0.949** |

proposed LBPHF method can be applied to other features, like LBPV [14] and central pixel in CLBP [15], not just limited to sign and magnitude as shown in the above experiments. As long as they are organized in the same way to uniform patterns so that they satisfy the Eq. 3, they can be embedded into Eq. 4 to replace $U_P(n,r)$ and generate new rotation invariant descriptors.

## IV. DYNAMIC TEXTURE RECOGNITION AND LBP-TOP

In the previous sections, LBP Histogram Fourier features (LBP-HF) were constructed for static image analysis and obtained very good results on static texture classification. In the following sections, we will extend it to an Appearance-Motion (AM) description for dealing with rotation variation in video sequences. Dynamic textures (DT) recognition is utilized as a case study. Recognition and segmentation of dynamic textures have attracted growing interest in recent years [5], [8], [24]. Dynamic textures provide a new tool for motion analysis. Now the general assumptions used in motion analysis that the scene is Lambertian, rigid and static, can be relaxed [28].

Recently, two spatiotemporal operators based on local binary patterns [31] were proposed for dynamic texture description: Volume Local Binary Patterns (VLBP) and Local Binary Pattern histograms from Three Orthogonal Planes (LBP-TOP): $XY$, $XT$ and $YT$ planes. These operators combine motion and appearance together, and are robust to translation and illumination variations. They also can be extended to multi-resolution analysis. A rotation invariant version of VLBP has also been proposed, providing promising performance for DT sequences with rotations [32]. However, VLBP considers co-occurrences of neighboring points in subsequent frames of a volume at the same time, which makes its feature vector too long when the number of neighboring points used is increased. The LBP-TOP does not have this limitation. It has performed very well in different types of computer vision problems, such as dynamic texture recognition [31], segmentation [6] and synthesis [12], facial expression recognition [31], visual speech recognition [33], activity recognition [17], [20], [21], and analysis of facial paralysis [16]. But LBP-TOP is not rotation invariant, which limits its wide applicability.

Fig. 6 left (a) shows one image in $XY$ plane, (b) in $XT$ plane which gives the visual impression of one row changing in time, and (c) describes the motion of one column in temporal space. For each pixel in images from these three planes or slices, a binary code is produced by thresholding its neighborhood in a circle or ellipse from $XY$, $XT$, $YT$ slices independently with the value of the center pixel. A histogram is created to collect up the occurrences of different binary patterns from three slices which are denoted as XY-LBP, XT-LBP and YT-LBP, then concatenated into

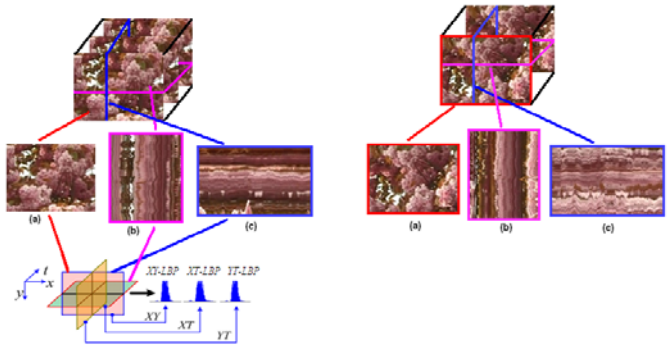


Fig. 6. Computation of LBP-TOP for "watergrass" with 0 (left) and 60 (right) degrees rotation.

a single histogram as demonstrated in last row of Fig. 6 left. In such a representation, DT is encoded by the LBP while the appearance and motion in two directions of DT are considered, incorporating spatial domain information and two spatiotemporal co-occurrence statistics together. For LBP-TOP, the radii in axes X, Y and T, and the number of neighboring points in the $XY$, $XT$ and $YT$ planes or slices can also be different, which can be marked as $R_X$,$R_Y$,$R_T$ and $P_{XY}$,$P_{XT}$,$P_{YT}$. The corresponding LBP-TOP feature is denoted as $LBP-TOP_{P_{XY},P_{XT},P_{YT},R_X,R_Y,R_T}$. Sometimes, the radii in three axes are same and so do the number of neighboring points in XY, XT and YT planes. In that case, we use $LBP-TOP_{P,R}$ for abbreviation where $P=P_{XY}=P_{XT}=P_{YT}$ and $R=R_X=R_Y=R_T$.

In this way, a description of DT is effectively obtained based on LBP from three different planes. The labels from the $XY$ plane contain information about the appearance, and in the labels from the $XT$ and $YT$ planes co-occurrence statistics of motion in horizontal and vertical directions are included. These three histograms are concatenated to build a global description of DT with the spatial and temporal features. However, the appearance-motion planes $XT$ and $YT$ in LBP-TOP are not rotation invariant, which makes LBP-TOP hard to handle the rotation variations. This needs to be addressed for DT description. As shown in Fig. 6 (right), the input video in top row is with 60 degrees rotation from that in Fig. 6 (left), so the $XY$, $XT$ and $YT$ planes in middle row are different from that of Fig. 6 (left), which obviously makes the computed LBP codes different from each other. Even if we sample the texture information in eight or 16 planar orientations, the orders of these planes would not change with the rotation of images.

On the basis of LBP-TOP, we propose two rotation invariant descriptors for LBP-TOP, based on using discrete Fourier transform for rotation invariant DT recognition. One is computing the 1D histogram Fourier transform for the uniform

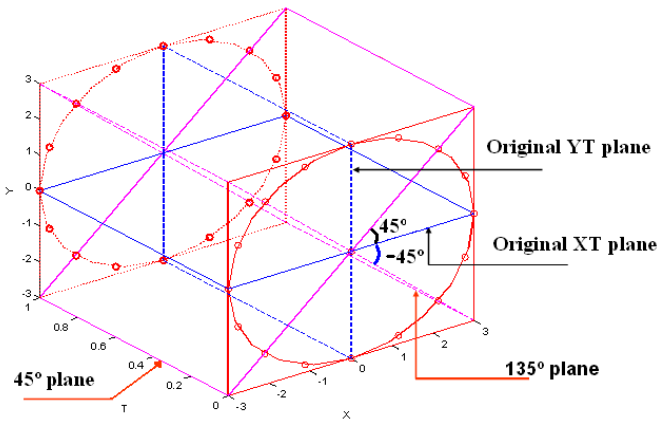


Fig. 7. Rotated planes from which LBP is computed.

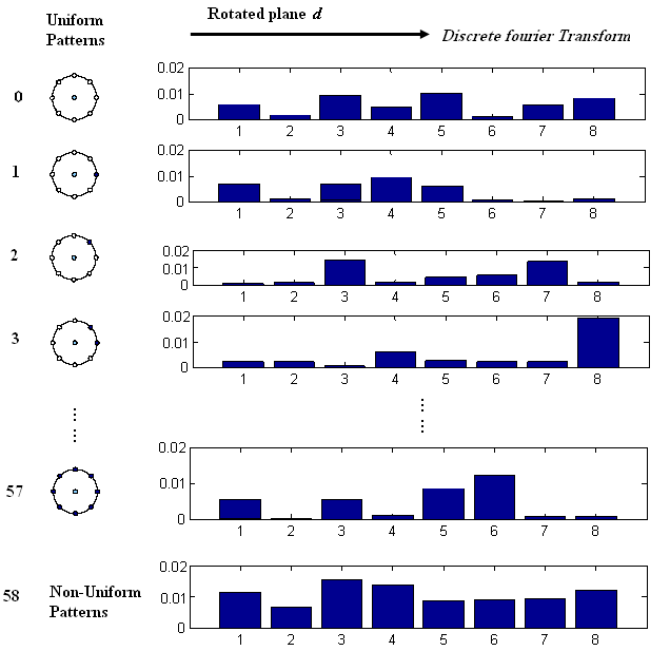


Fig. 8. LBP histograms for uniform patterns in different rotated motion planes with $P_{XY} = 8$ and $P_T = 8$.

patterns along all the rotated motion planes. The other one is computing the 2D Fourier transform for the patterns with the same number of "1"s along its rotation in bins and as well along all the rotated motion planes, which avoids to use the redundant information and makes the computation complexity much lower than the first one. We compare them with earlier methods in the classification of rotated DT sequences. The robustness of the descriptors on viewpoint variations is also studied, using a recently introduced test set for view-invariant dynamic texture recognition [24].

## V. Rotation Invariant LBP-TOP

Both $LBP-XT$ and $LBP-YT$ describe the appearance and motion. When a video sequence rotates, these two planes do not rotate accordingly, which makes the LBP-TOP operator not rotation invariant. Moreover, rotation only happens around the axis parallel to $T$ axis, so considering the rotation invariant descriptor inside planes does not make any sense. Instead, we should consider the rotations of the planes, not only just the orthogonal planes ($XT$ and $YT$ rotation with 90 degrees), but also the planes with different rotation angles, like the purple planes with rotations 45 and 135 degrees in Fig. 7. So the AM planes consist of $P_{XY}$ rotation planes around $T$ axis. The radius in $X$, $Y$ should be same, but can be different from that in $T$. Only two types for the number of neighboring points are included, one is $P_{XY}$ which determines how many rotated planes will be considered, the other one is $P_T$ which is the number of neighboring points in AM planes. The original $XT$ and $YT$ are not two separate planes any more, instead they are AM planes obtained by rotating the same plane zero and 90 degrees, respectively.

The corresponding feature is denoted as $LBP-TOP^{ri}_{P_{XY},P_T,R_{XY},R_T}$. Suppose the coordinates of the center pixel $g_{t_c,c}$ are $(x_c, y_c, t_c)$, we compute the LBP from $P_{XY}$ spatiotemporal planes. The coordinates of the neighboring points $g_{d,p}$ sampled from the ellipse in $XYT$ space with $g_{t_c,c}$ as center, $R_{XY}$ and $R_T$ as the length of axes, are given by $(x_c + R_{XY}\cos(2\pi d/P_{XY})\cos(2\pi p/P_T)$, $y_c - R_{XY}\sin(2\pi d/P_{XY})\cos(2\pi p/P_T), t_c + R_T\sin(2\pi p/P_T))$, $d(d = 0\cdots(P_{XY}-1))$ is the index of the AM plane and $p(p = (0\cdots(P_T-1))$ represents the label of neighboring point in plane $d$.

### A. One Dimensional Histogram Fourier LBP-TOP (1DHFLBP-TOP)

After extracting the uniform LBP for all the rotated planes, we compute the Fourier transform for every uniform pattern along all the rotated planes. Fig. 8 demonstrates the computation. For $P_T = 8$, 59 uniform patterns can be obtained as shown in the left column. For all the rotated $P_{XY} = 8$ planes, discrete Fourier transform is applied for every pattern along all planes to produce the frequency features. Eq. 12 illustrates the computation.

$$H_1(n,u) = \sum_{d=0}^{P_{XY}-1} h_1(d,n)e^{-i2\pi ud/P_{XY}}, \qquad (12)$$

where, $n$ is index of uniform patterns $((0\cdots N), N = 58\ for\ P_T = 8)$ and $u(u = 0\cdots P_{XY}-1)$ is frequency. $d(d = 0\cdots(P_{XY}-1))$ is the index of rotation degrees around the line passing through the current central pixel $g_{t_c,c}$ and parallel to $T$ axis. $h_1(d,n)$ is the value of pattern $n$ in uniform LBP histogram at plane $d$. To get the low frequencies, $u$ can use the value from 0 to $(P_{XY}/s+1)$ (e.g. $s = 4$ in experiments).

When $u = 0$, $H_1(n,0)$ means the sum of the pattern $n$ through all the rotated motion planes, which can be thought as another kind of rotation invariant descriptor of simply summing the histograms from all the rotated planes. Since it uses one dimensional histogram Fourier transform for LBP-TOP, we call it **one dimensional histogram Fourier LBP-TOP (1DHFLBP-TOP)**.

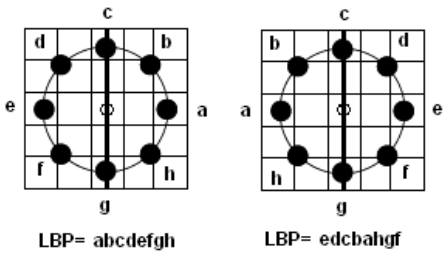


Fig. 9. LBP from plane with $g$ degrees (left) and $g + 180$ degrees (right). They are mirrored.

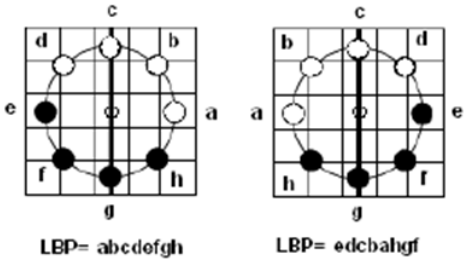


Fig. 10. Uniform pattern before mirror (left) and after mirror (right).

The feature vector $V1$ of 1DHFLBP-TOP is of the following form:

$$V1 = [|H_1(0,0)|, \cdots, |H_1(0, P_{XY}/s+1)|, \cdots, |H_1(N-1,0)|, \cdots, |H_1(N-1, P_{XY}/s+1)|].$$

$N$ is the number of uniform patterns with neighboring points $P_{XY}$. Here $s$ is the segments of frequencies. Not all the $P_{XY}$ frequencies are used. Instead only the low frequency, saying $[0\ P_{XY}/s+1]$ are utilized. Total length of $V1$ is $LRI1 = N \times (P_{XY}/s+2)$.

### *B. Two Dimensional Histogram Fourier LBP-TOP (2DHFLBP-TOP)*

We can notice that for the descriptor 1DHFLBP-TOP, the LBPs from a plane rotated $g$ degrees $(180 > g >= 0)$ and $g+180$ are mirrored along the $T$-axis through the central point, as the line $cg$ in Fig. 9, but they are not same. So to get the rotation invariant descriptor, all the rotated planes should be used, which increases the computational load.

But for planes rotated $g$ degrees and $g + 180$ degrees, the information is same, so there is no need to use both of them. We notice this in the left and right images of Fig. 9: even though the LBP codes $(a, b, c, d, e, f, g, h = 0\ or\ 1)$ are mirrored, the neighboring relationship still remains, e.g. $d$ is adjacent to $c$ and $e$ in both images.

According to the definition of uniform patterns: "a local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is considered circular", 1) if one LBP code $L$ is uniform and there is zero bitwise transition from 0 to 1 or 1 to 0, it means all the bits in this LBP are 0 or 1. So after mirror, for the produced LBP code $L'$, all the bits are still 0 or 1, which is still uniform. 2) if L is uniform and there are two bitwise transitions from 0 to 1 or 1 to 0, as shown in Fig. 10, the transitions happen between d, e, and h, a (Fig. 10 left). After mirror, the neighboring relationship is unchanged, so the transitions are also between e,d and a,h (Fig. 10 right) and the transition times are still two, which means the mirrored LBP is also uniform. 3) if $L$ is non-uniform, we first assume that after mirror, $L'$ is uniform. We can then mirror $L'$ again, and the obtained $L''$ should be equal to $L$. But according to 1) and 2), if $L'$ is uniform, the mirrored $L''$ is also uniform. But $L$ is non-uniform, which means $L'' <> L$. It is self-contradictory. Thus, if $L$ is non-uniform, after mirror, obtained $L'$ is also non-uniform. Because in the mirror transformation, only the location of the bits changes, the value of all bits keep same, so the number of 1's is unchanged no matter $L$ is uniform or non-uniform.

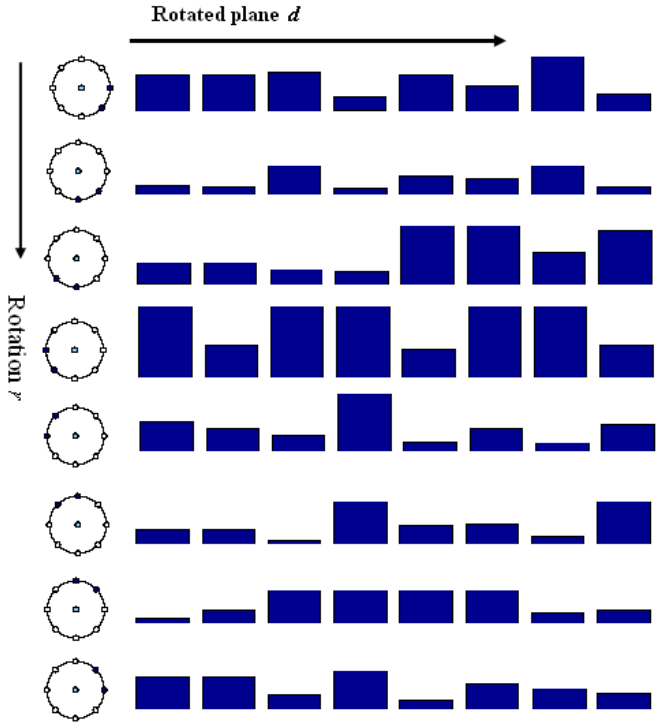


Fig. 11. The examples of LBP with number of "1"s two in different rotated motion planes with $P_{XY} = 8$ and $P_T = 8$.

So the uniform patterns in a plane rotated $g$ degrees are still uniform in a plane with $g + 180$ degrees, and with the same number of "1"s. The non-uniform patterns are still non-uniform in both planes.

We propose to make the uniform patterns with same number of "1"s into one group, but they can rotate $r(r = 0 \cdots (P_T - 1))$ bits as shown in the left column of Fig. 11. In total we have $P_{XY} - 1$ groups with the "1"s numbered as $1, 2, 3, \cdots, P_{XY} - 1$. For the uniform patterns with "1"s numbered as 0(all zeros), $P_{XY}$(all ones), and non-uniform patterns, no matter how video sequences rotate, they remain the same. For every group, the 2D Fourier transform is used to get the rotation invariant descriptor, as shown in Fig. 11.

Eq. 13 illustates the computation.

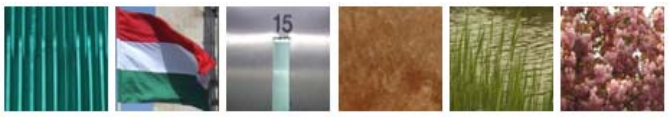

Fig. 12. DynTex database (http://projects.cwi.nl/dyntex/).

$$H_2(m,u,v) = \sum_{d=0}^{P_{XY}-1} \sum_{r=0}^{P_T-1} h_2(U(m,r),d) e^{-i2\pi ud/P_{XY}} e^{-i2\pi vr/} \quad (13)$$

where $m(m = 1 \cdots P_{XY} - 1)$ is the number of "1"s, and $u(u = 0 \cdots P_{XY} - 1)$ and $v(0 \cdots P_T - 1)$ are frequencies in two directions. $d$ is the index of rotation degree around $T$ axis, $r$ is the rotation inside the circle or ellipse with $P_T$ neighboring points. $U(m,r)$ is the uniform pattern with the $m$ of "l"s and $r$ is the rotation index of it. $h_2(U(m,r),d)$ is the number of occurrences of $U(m,r)$ at plane $d$. So, $H_2(0,0,0)$ is the sum of all zeros in all planes, $H_2(P_{XY},0,0)$ is the sum of all ones, and $H_2(P_{XY}+1,0,0)$ is the sum of all non-uniform patterns in all planes, which can be used with $H_2(m,u,v)$ together to describe the DT. The difference of this descriptor from the first one is that it computes the histogram Fourier transforms from two directions, considering the frequencies not only from planes but also from pattern rotations. We call it **two dimensional histogram Fourier LBP-TOP (2DHFLBP-TOP)**.

The final feature vector $V2$ is of the following form:

$$\begin{aligned} V2 = \; & [|H_2(1,0,0)|, \cdots, |H_2(1,P_{XY}/s+1,P_T/s+1)|, \cdots \\ & |H_2(P_{XY}-1,0,0)|, \cdots, \\ & |H_2(P_{XY}-1,P_{XY}/s+1,P_T/s+1)|, \\ & |H_2(0,0,0)|, |H_2(P_{XY},0,0)|, |H_2(P_{XY}+1,0,0)|]. \end{aligned}$$

Total length of $V2$ is $LRI2 = (P_{XY}-1) \times (P_{XY}/s+2) \times (P_T/s+2) + 3$.

## VI. Experiments on Dynamic Texture Classification

The performance of the proposed rotation invariant video descriptors was tested on dynamic texture classification. We carry out experiments on DynTex database for rotation variation evaluation and the dataset from [25] for view variation evaluation.

### A. Experiments on Rotation Variations

DynTex, a large and varied database of dynamic textures which originally included 35 DTs and now has been extended to have 656 DTs, was selected for the experiments. Fig. 12 shows example DTs from this dataset.

To evaluate the rotation invariance of the methods and compare with previous methods [32], we use the same setup as in [32]. The dataset used in experiments includes 35 classes from the original database and each sequence was rotated by 15 degrees intervals as shown in Fig. 13, obtaining 24 sequences. Every sequence was cut in length into two sequences. So totally we have 48 samples for each class. In our experiments, two sequences with 0 degree (no rotation) are used as training samples, and the remaining ones are test sequences. Hence, in this suite, there are $70(35 \times 2)$ training models and $1610(35 \times 46)$ testing samples.

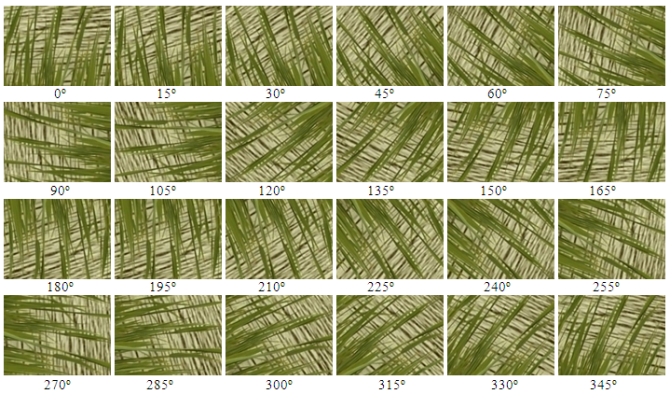


Fig. 13. Images after rotating by $15^\circ$ intervals.

The mean values of the rotation invariant LBP-TOP features of the two samples without rotation are computed as the feature for the class. The testing samples are classified or verified according to their difference with respect to the class using the $k$ nearest neighbor method $(k = 1)$.

Table IV demonstrates the results with different parameters of the proposed descriptors for all rotation tests. Here we show results for both L1 distance metric and Chi-square distance metric and it can be seen that L1 distance provides similar or better results than Chi-square. So in the following experiments of KNN classification, the reported results are for L1 distance measurement. The first two rows are the results using original LBP-TOP with four neighboring points. The results are very poor. Even when more neighboring points are used, like eight or 16, as shown in the sixth, seventh, 11th and 12th rows, the classification rates are still less than 60%, which demonstrates that LBP-TOP is not rotation invariant. We also did the experiments using oversampled LBP-TOP, which is denoted as $LBP-TOP_{P_{XY},P_T,R_{XY},R_T}$, i.e. we extend the fixed three planes in LBP-TOP to $P_{XY}$ planes in spatiotemporal domain and sample the neighboring points in each plane with radii $R_{XY}$ and $R_T$ in $XY$ direction and $T$ direction, respectively, for $P_T$ points in ellipse. The uniform histograms from each plane are then concatenated together as the oversampled LBP-TOP features. The results are shown in the third, eighth and 13th rows of Table IV. It can be seen that the oversampled LBP-TOP got better results than original LBP-TOP because oversampling can include more information. But the accuracy is lower than for the proposed rotation invariant descriptors, which shows that oversampling can not deal with rotation variations. Even though it contains more information, the order of planes keeps unchanged when there are rotations which makes oversampling not rotation invariant. When using four neighboring points, $2DHFLBP-TOP_{4,4,1,1}$ obtained 82.11% with only 30 features. When using 16 neighboring points in AM planes, $1DHFLBP-TOP_{16,16,2,2}$ got 98.57% and $2DHFLBP-TOP_{16,16,2,2}$ 97.33%, respectively. Both descriptors are effective for dealing with rotation variations.

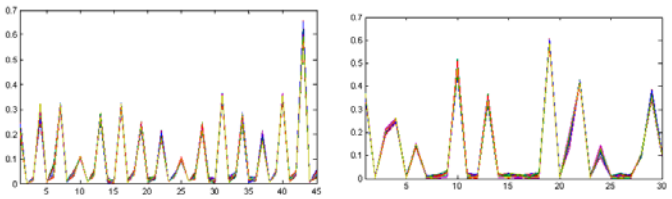

Fig. 14. 1DHFLBP-TOP histograms (left) and 2DHFLBP-TOP histograms (right) of DT "square sheet" with 48 rotation samples.

One may have noticed that the proposed 2DHFLBP is better than 1DHFLBP when the number of neighboring pixels is four, but worse than 1DHFLBP for a larger number of these pixels. That is because 2DHFLBP considers only the number of "1"s for the uniform patterns, i.e. in that way only half of the rotated planes need to be used, while 1DHFLBP considers the uniform patterns for all planes. When the number of neighboring pixels increases, there is more information missing for 2DHFLBP compared with 1DHFLBP. But 2DHFLBP-TOP gets a comparative accuracy to the 1DHFLBP-TOP with a much shorter feature vector and only using half of the rotated planes which will save computation time. It is a good compromise for computational efficiency and recognition accuracy, which could make it very useful for many applications.

Fig. 14 shows the 1DHFLBP-TOP histograms and 2DHFLBP-TOP histograms for the "square sheet" dynamic texture with 48 rotation samples. It can be seen that both descriptors have good characteristics for rotation variations.

The magnitude of LBP-HF is also extended to video description and utilized as supplemental information to the proposed rotation invariant descriptors. The mean difference of the neighboring point's grey scale against the central pixel in each spatiotemporal plane is calculated and utilized as threshold for getting the binary code. Results of 1DHFLBP-TOP and 2DHFLBP-TOP which are actually the sign LBP-TOP histogram Fourier, 1DHFLBP-TOP_M and 2DHFLBP-TOP_M which are the magnitude LBP-TOP histogram Fourier, and their combination on DynTex database are demonstrated in Table V. As we can see, the magnitude information does not work as well as sign information, which is consistent to the conclusion in [15]. The combination of the magnitude information together with sign information yields much better result than using either of them solely. Especially when the number of neighboring points and planes are fewer, as for $1DHFLBP-TOP_{4,4,1,1}$ and $2DHFLBP-TOP_{4,4,1,1}$, the improvement is significant, from 79.50% to 84.22% and from 82.11% to 86.15%, respectively.

Table VI lists the accuracies using different methods. The first three rows give the results using LBP histogram Fourier transform [1] which only considers the rotation invariance in appearance. The accuracy of 40%-60% shows its ineffectiveness for rotations happening on videos. The middle four rows show the results using different versions of rotation invariant VLBP [32], which can get quite good results using short feature vectors, e.g. 87.14% with only 26 features. But because VLBP considers the cooccurrence of all the neighboring points in three frames of a volume at the same time, it is hard to be extended to use more neighboring information. Four neighboring points is almost the maximum as concluded in [31]. Comparatively, the proposed two rotation invariant LBP-TOP descriptors inherit the advantage of the original LBP-TOP, they can be easily extended to use many more neighboring points, e.g. 16 or 24, and they obtained almost 100% accuracy DT recognition with rotations. For comparing with LBP-TOP on DTs without rotation, we carried out experiments on DynTex database. A 92.9% accuracy is obtained with 1DHFLBP-TOP, while LBP-TOP obtained 93.4% accuracy with four neighboring points. We can see that on videos without rotation the proposed rotation invariant descriptors give similar or slightly worse results than non-invariant LBP-TOP. More importantly, they clearly outperform LBP-TOP when there are rotations. We also computed the maximal of uniform patterns along all the rotated planes as a simple rotation invariant descriptor for comparison. We got 70.93% (vs. 82.11% for the proposed method) when using four neighboring points and radius one, and 91.06% (vs. 98.07%) when using eight neighboring points and radius two. So with this simple normalization, we can get some rotation invariance, but it works much poorer than proposed methods.

TABLE V
RECOGNITION RESULTS USING ROTATION INVARIANT SIGN HFLBP-TOP AND MAGNITUDE HFLBP-TOP.

| Features | Length | Results (%) |
|---|---|---|
| $1DHFLBP-TOP_{4,4,1,1}$ | 45 | 79.50 |
| $1DHFLBP-TOP_{4,4,1,1}\_M$ | 45 | 63.60 |
| $1DHFLBP-TOP_{4,4,1,1}\_S\_M$ | $45 \times 2$ | **84.22** |
| $2DHFLBP-TOP_{4,4,1,1}$ | 30 | 82.11 |
| $2DHFLBP-TOP_{4,4,1,1}\_M$ | 30 | 61.86 |
| $2DHFLBP-TOP_{4,4,1,1}\_S\_M$ | $30 \times 2$ | **86.15** |
| $1DHFLBP-TOP_{8,8,2,2}$ | 236 | 98.07 |
| $1DHFLBP-TOP_{8,8,2,2}\_M$ | 236 | 86.71 |
| $1DHFLBP-TOP_{8,8,2,2}\_S\_M$ | $236 \times 2$ | **98.45** |
| $2DHFLBP-TOP_{8,8,2,2}$ | 115 | 94.41 |
| $2DHFLBP-TOP_{8,8,2,2}\_M$ | 115 | 88.32 |
| $2DHFLBP-TOP_{8,8,2,2}\_S\_M$ | $115 \times 2$ | **96.27** |

TABLE VI
RESULTS USING DIFFERENT METHODS.

| Features | Length | Results (%) |
|---|---|---|
| $HFLBP_{4,1}$ | 12 | 43.60 |
| $HFLBP_{8,2}$ | 31 | 50.37 |
| $HFLBP_{16,2}$ | 93 | 61.93 |
| ri #1 $VLBP_{1,4,1}$ | 864 | 75.34 |
| ri #2 $VLBP_{2,4,1}$ | 4176 | 79.38 |
| riu2 #2 $VLBP_{1,4,1}$ | 16 | 78.07 |
| riunu2 #2 $VLBP_{1,4,1}$ | 26 | 87.20 |
| $1DHFLBP-TOP_{16,16,2,2}$ | 1458 | **98.57** |
| $2DHFLBP-TOP_{16,16,2,2}$ | 543 | **97.33** |

### *B. Experiments on view variations*

View variations are very common in dynamic textures. View-invariant recognition DTs is a very challenging task. To our best knowledge, most of the proposed methods for dynamic texture categorization validated their performance on the ordinary DT databases, without viewpoint changes, except [24], [29]. Woolfe and Fitzgibbon addressed shift invariance [29] and Ravichandran et al. [24] proposed to use bag of

TABLE IV
RECOGNITION RESULTS USING DIFFERENT PARAMETERS. ($u2$ DENOTES UNIFORM PATTERNS).

| Features | Length | L1 Results(%) | Chi-square Results(%) |
|---|---|---|---|
| $LBP-TOP_{4,1}$ | 48 | 43.11 | 39.81 |
| $LBP-TOP^{u2}_{4,1}$ | 45 | 42.98 | 39.32 |
| $LBP-TOP^{u2}_{4,4,1,1}$ | 60 | 51.30 | 48.39 |
| $1DHFLBP-TOP_{4,4,1,1}$ | 45 | **79.50** | 71.99 |
| $2DHFLBP-TOP_{4,4,1,1}$ | 30 | **82.11** | 71.93 |
| $LBP-TOP_{8,2}$ | $256\times 3$ | 58.76 | 51.86 |
| $LBP-TOP^{u2}_{8,2}$ | $59\times 3$ | 58.63 | 52.86 |
| $LBP-TOP^{u2}_{8,8,2,2}$ | $59\times 8$ | 69.25 | 65.71 |
| $1DHFLBP-TOP_{8,8,2,2}$ | 236 | **98.07** | 97.83 |
| $2DHFLBP-TOP_{8,8,2,2}$ | 115 | **94.41** | 94.10 |
| $LBP-TOP^{u2}_{16,1}$ | $243\times 3$ | 52.86 | 47.58 |
| $LBP-TOP^{u2}_{16,2}$ | $243\times 3$ | 56.71 | 51.12 |
| $LBP-TOP^{u2}_{16,16,2,2}$ | $243\times 16$ | 72.05 | 68.32 |
| $1DHFLBP-TOP_{16,16,2,2}$ | 1458 | **98.57** | 99.69 |
| $2DHFLBP-TOP_{16,16,2,2}$ | 543 | **97.33** | 97.76 |

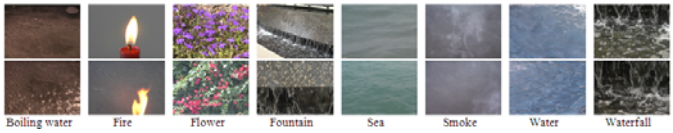


Fig. 15. Sample images from reorganized dataset from [25].

system (BoS) as the representation for dealing with viewpoint changes.

To evaluate the robustness of our proposed descriptors to view variations, we use the same dataset to [24], [25]. This dataset consists of 50 classes of 4 video sequences each. Many previous works [4], [25] are based on the 50 class structure and the reported results are not on the entire video sequences, but on a manually extracted patch of size $48\times 48$ [24]. In [24], the authors combine the sequences that are taken from different viewpoints, and reduce the dataset to a nine class dataset with the classes being boiling water (8), fire (8), flowers (12), fountains (20), plants (108), sea (12), smoke (4), water (12) and waterfall (16). The numbers in parentheses represent the number of sequences in the dataset. To compare the performance of handling view variations with the methods proposed in [25] and [24], we use the same experimental setup, i.e. 1) the plants class is left out since the number of sequences of plants far outnumbered the number of sequences for the other classes, so remaining eight classes are used in our experiments; 2) four different scenarios are explored in this paper. The first set is an easy two class problem namely the case of water vs. fountain. The second one is a more challenging two class problem namely the fountain vs. waterfall. The third set of experiments is on a four class (water, fountain, waterfall and sea) problem and the last set is on the reorganized database with eight classes. We abbreviate these scenarios as W-F (Water Vs. Fountain), F-WF (Fountain Vs. Waterfall), FC (Four Classes) and EC (Eight Classes). Sample frames from the video sequences in the database are shown in Fig. 15. For every scenario, we train using 50% of the data and test using the rest.

We utilize Support Vector Machine (SVM) and the Nearest Neighbor (1NN) of L1 distance as classifiers. For SVM, the second degree polynomial kernel function is used in the experiments. Fig. 16 compares our results using $1DHFLBP-TOP_{8,8,1,1}$ and $2DHFLBP-TOP_{8,8,1,1}$, and the results with term frequency (TF) and soft-weighting (SW) from [24], and DK (DS) from [25]. TF and SW are the two kinds of representation utilized in [24] to represent the videos using the codebook. DK and DS depict the methods proposed in [25], in which a single LDS is used to model the whole video sequence and the nearest neighbor and SVM classifiers are used for categorization, respectively. Fig. 16 shows the results for four scenarios using SVM and 1NN as classifier. In all four experimental setups, the second (F-WF) and last (EC) setups are considered more challenging, because in the second scenario, the viewpoints in testing are not used in the training, which makes them totally novel. In the last scenario, we have a total of 92 video sequences with varying number of sequences per class and they are from various viewpoints. It can be seen from Fig. 16 that our proposed descriptors obtain leading accuracy for all four scenarios. Especially for more challenging fountain vs. waterfall and all eight class problems, our results from $1DHFLBP-TOP_{8,8,1,1}$ with SVM are 87.5% and 86.96%, and with 1NN are 87.5% and 73.91%, respectively. These are much better than for TF (70% and 63%) and SW (76% and 80%) with SVM [24], DK (50% and 52%) and DS (56% and 47%) [24]. In addition, we also carried out the experiments using $LBP-TOP$ without rotation invariant characteristics. $LBP-TOP^{u2}_{8,1}$ obtained 75% and 71.74% with SVM for more difficult F-WF and EC scenarios, which is inferior to either of the proposed descriptors. Rotation invariant VLBP [32] was also implemented for comparison. For F-WF and EC scenarios, 78.15% and 80.43% are achieved with SVM classifier, which are better than those from LBP-TOP, but worse than those from the proposed descriptors. From these results and comparison on the database with real 3D viewpoint changes, it can be seen that even though our descriptors are not designed as view-invariant, they can deal with this problem very effectively.

Table VII lists the results using sign LBP-TOP histogram Fourier, the combination of magnitude LBP-TOP histogram

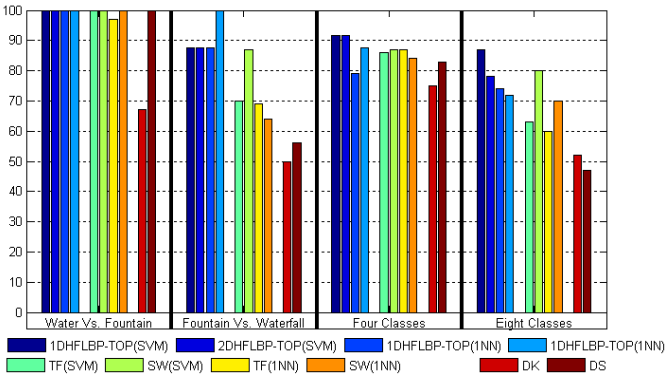


Fig. 16. Classification results for the four scenarios with different methods.

Fourier and sign LBP-TOP histogram Fourier, and oversampled LBP-TOP features for the most difficult EC problem. In this table, we can see when combining sign and magnitude components together, we achieve higher accuracy than using sign alone. In this database, the in-plane rotation is not the main problem, but view variation. Even though view changes can cause appearance differences with translation, rotation and scaling, the local transition in appearance and motion would not change much. With oversampling, it can catch much more information about local appearance and motion transition than the original LBP-TOP providing comparable results to the proposed rotation invariant LBPTOP histogram Fourier descriptors but with much longer feature vectors when the number of neighboring points increases to be over eight.

NN is one of the easiest machine learning algorithms. In the classification it is only determined by the sample closest to the test sample. So when there is big overlapping in classes, NN works pretty well. But it is prone to over-fitting because of the highly non-linear nature. SVM classifier is well founded in statistical learning theory and has been successfully applied to various object detection tasks in computer vision. SVM finds a separating hyperplane with the maximal margin to separate the training data in feature space. The classification of SVM is determined by the support vectors, so it is usually more robust than NN. As noticed in Table VII, when the number of planes is small, like $HFLBP-TOP_{4,4,1,1}$, the features are not very discriminative, thus the overlapping of classes is big and NN got similar or even higher results than SVM. But in most cases, SVM obtained better results than NN. In addition, the computational complexity of SVM depends on the number of support vectors, not the dimension of feature vectors. When the dimension is very high, SVM is computationally less expensive. However, SVM is designed for two-class problems. If there are more than two classes, some strategy, like reorganizing a multi-class problem into multiple 2-class problems needs to be employed. Therefore it is generally easier to deal with multiple-class problems with KNN than SVM.

## VII. Discussion and Conclusion

In this paper, we proposed rotation invariant image and video descriptors based on computing the discrete Fourier transform of LBP or LBP-TOP histograms, respectively.

The proposed LBP-HF features differ from the previous versions of rotation invariant LBP, since the LBP-HF features are computed from the histogram representing the whole region, i.e. the invariants are constructed from the histogram of non-invariant LBPs instead of computing invariant features independently at each pixel location.

This approach can be generalized to embed any features in the same organization of uniform patterns, like the LBPHF_S_M shown in our experiments. By embedding different features into the LBP-HF framework, we are able to generate different rotation invariant features and combine supplementary information together to improve the description power of the proposed method. The use of complementary magnitude of LBP will further improve the classification accuracy. In our experiments, LBPHF_S_M descriptors were shown to outperform the original rotation invariant LBP, LBP_M, LBP-HF and LBP_S_M features for the Outex database.

Moreover, two rotation invariant descriptors based on LBP-TOP were developed. One is computing the 1D histogram Fourier transform for the uniform patterns along all the rotated motion planes. The other one is computing the 2D Fourier transform for the patterns with the same number of "1"s along its rotation in bins and as well along all the rotated motion planes. Experiments on rotated DT sequences show that both descriptors achieve very promising results in recognizing video sequences with rotations. Another experiments on DTs captured from different views provided better results than the state-of-the-art, proving the effectiveness of our approach for dealing with view variations. So it can be seen the proposed method not only works on in-plane rotations but it can deal with out-of-plane rotations robustly, obtaining much better results than previous methods. As well, the second rotation invariant descriptors consider the mirror relations for the patterns produced from planes rotated $g$ degrees and $g + 180$ degrees. This reduces its computational complexity to the half of the first descriptor, because it only needs to compute the LBP histograms from the planes with rotation degrees less than 180 degrees. The proposed descriptors keep the advantages of LBP-TOP, such as robustness to illumination changes and ability to describe DTs at multiple resolutions, as well as have the capability of handling rotation and view variations. They also can be extended to embed other features, like the magnitude of LBP-TOP shown in our experiments in Section VI, which can be utilized solely or combined with sign of LBP-TOP as supplemental information to help improve the classification accuracy. They will widen the applicability of LBP-TOP to such tasks in which rotation invariance or view invariance of the features is important. For example, in activity recognition, there exist rotations due to rotation of cameras, or activity itself (e.g. some hand gestures), and we believe rotation invariant descriptor will be very useful in these problems.

## Acknowledgment

This work was supported by the Academy of Finland and Infotech Oulu. JM was supported by Czech Science Foundation project P103/10/1585.

TABLE VII
RECOGNITION RESULTS ON EC PROBLEM.

| Features | Length | SVM (%) | NN (%) |
|---|---|---|---|
| $1DHFLBP-TOP_{4,4,1,1}$ | 45 | 65.22 | 59.78 |
| $1DHFLBP-TOP_{4,4,1,1}\_S\_M$ | $90(45\times 2)$ | **73.91** | **77.17** |
| $2DHFLBP-TOP_{4,4,1,1}$ | 30 | 58.70 | 59.78 |
| $2DHFLBP-TOP_{4,4,1,1}\_S\_M$ | $60(30\times 2)$ | **71.74** | **73.91** |
| $LBP-TOP^{u2}_{4,4,1,1}$ | $60(15\times 4)$ | **70.65** | **61.96** |
| $1DHFLBP-TOP_{8,8,1,1}$ | 236 | 86.96 | 73.91 |
| $1DHFLBP-TOP_{8,8,1,1}\_S\_M$ | $472(236\times 2)$ | **86.96** | **78.26** |
| $2DHFLBP-TOP_{8,8,1,1}$ | 115 | 78.26 | 71.74 |
| $2DHFLBP-TOP_{8,8,1,1}\_S\_M$ | $230(115\times 2)$ | **79.35** | **82.61** |
| $LBP-TOP^{u2}_{8,8,1,1}$ | $472(59\times 8)$ | **85.87** | **73.91** |
| $1DHFLBP-TOP_{16,16,2,2}$ | 1458 | 71.74 | 65.22 |
| $1DHFLBP-TOP_{16,16,2,2}\_S\_M$ | $2916(1458\times 2)$ | **85.87** | **67.39** |
| $2DHFLBP-TOP_{16,16,2,2}$ | 543 | 68.48 | 54.35 |
| $2DHFLBP-TOP_{16,16,2,2}\_S\_M$ | $1086(543\times 2)$ | **79.35** | **71.74** |
| $LBP-TOP^{u2}_{16,16,2,2}$ | $3888(243\times 16)$ | **76.09** | **67.39** |

## REFERENCES


[1] T. Ahonen, J. Matas, C. He, and M. Pietikäinen, "Rotation invariant image description with local binary pattern histogram fourier features," *16th Scandinavian Conference on Image Analysis*, pp. 2037-2041, 2009.

[2] H. Arof, and F. Deravi,"Circular neighbourhood and 1-d dft features for texture classification and segmentation," *IEE Proceedings - Vision, Image and Signal Processing*, vol. 145, no. 3, pp. 167-172, 1998.

[3] B. Caputo, E. Hayman, and P. Mallikarjuna, "Class-specific material categorisation," *In: 10th IEEE International Conference on Computer Vision*, pp. 1597-1604, 2005

[4] A. Chan, and N. Vasconcelos, "Classifying video with kernel dynamic textures," *CVPR*, 2007.

[5] A. Chan, and N. Vasconcelos, "Variational layered dynamic textures," *CVPR*, pp. 1063-1069, 2009.

[6] J. Chen, G. Zhao, and M. Pietikäinen, "Unsupervised dynamic texture segmentation using local spatiotemporal descriptors," *ICPR*, 2008.

[7] D. Chetverikov, and R. Péteri, "A brief survey of dynamic texture description and recognition," *Int'l Conf. Computer Recognition Systems*, pp. 17-26, 2005.

[8] T. Crivelli, P. Bouthemy, and B. Cernuschi-Frias, and J.F. Yao, "Learning mixed-state markov models for statistical motion texture tracking," *ICCV workshop on Machine Learning for Vision-based Motion Analysis*, 2009.

[9] K.J. Dana, B. Ginneken, S.K. Nayar and J.J. Koenderink, "Reflectance and texture of real-world surfaces," *ACM Transactions on Graphics*, vol. 18, no. 1, pp. 1-34, 1999.

[10] S. Fazekas, and D. Chetverikov, "Analysis and performance evaluation of optical flow features for dynamic texture recognition," *Image Communication*, pp. 680-691, 2007.

[11] A. Fernandez, O. Ghita, E. Gonzalez, F. Bianconi, and P.F. Whelan, "Evaluation of robustness against rotation of LBP, CCR and ILBP features in granite texture classification," *Machine Vision and Applications*, 2010.

[12] Y. Guo, G. Zhao, J. Chen, and M. Pietikäinen, and Z. Xu, "Dynamic texture synthesis using a spatial temporal descriptor," *ICIP*, 2009.

[13] Z. Guo, L. Zhang, D. Zhang and S. Zhang, "Rotation Invariant Texture Classification Using Adaptive LBP with Directional Statistical Features," *ICIP*, 2010.

[14] Z. Guo, L. Zhang, and D. Zhang, "Rotation Invariant Texture Classification Using LBP Variance (LBPV) with global matching," *Pattern Recognition*, vol. 43, pp. 706-719, 2010.

[15] Z. Guo, L. Zhang, and D. Zhang, "A Completed Modeling of Local Binary Pattern Operator for Texture Classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657-1663, 2010.

[16] S. He, J.J. Soraghan, B. OReilly, and D. Xing, "Quantitative analysis of facial paralysis using local binary patterns in biomedical videos," *IEEE Transactions on Biomedical Engineering*, vol. 56, pp. 1864-1870, 2009.

[17] V. Kellokumpu, G. Zhao, and M. Pietikäinen, "Recognition of Human Actions Using Texture Descriptors," *Machine Vision and Applications*, in press.

[18] S. Liao, M. Law, and C.S. Chung, "Dominant local binary patterns for texture classication," *IEEE Transactions on Image Processing*, vol. 18, pp. 1107-1118, 2009.

[19] Z. Lu, W. Xie, J. Pei, and J. Huang, "Dynamic texture recognition by spatiotemporal multiresolution histogram," *IEEE Workshop Motion and Video Computing*, 2005.

[20] Y. Ma, and P. Cisar, "Event detection using local binary pattern based dynamic textures," *CVPR Workshop on Visual Scene Understanding*, 2009.

[21] R. Mattivi, and L. Shao, "Human action recognition using lbp-top as sparse spatio-temporal feature descriptor," *CAIP*, pp. 740-747, 2009.

[22] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution Gray Scale and Rotation Invariant Texture Analysis with Local Binary Patterns," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.

[23] T. Ojala, T. Mäenpää, M. Pietikäinen, J. Viertola, J. Kyllönen, and S. Huovinen, "Outex - new framework for empirical evaluation of texture analysis algorithms," *In: Proc. 16th International Conference on Pattern Recognition*, vol. 1, pp. 701-706, 2002.

[24] A. Ravichandran, R. Chaudhry, and R. Vidal, "View-invariant dynamic texture tecognition using a bag of dynamical systems," *CVPR*, pp. 1-6, 2009.

[25] P. Saisan, G. Doretto, Y.N. Wu, and S. Soatto, "Dynamic texture recognition," *CVPR*, pp. 58-63, 2001.

[26] M. Tuceryan, and A.K. Jain, *Texture analysis*, In C.H. Chen, L.F. Pau, and P.S.P. Wang, eds.: The Handbook of Pattern Recognition and Computer Vision (2nd Edition), World Scientific Publishing Co., pp. 207-248, 1998.

[27] M. Varma, and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision*, vol.62, no. 1-2, pp. 61-81, 2005. 61–81

[28] R. Vidal, and A. Ravichandran, "Optical flow estimation & segmentation of multiple moving dynamic textures," *CVPR*, pp. 516-521, 2005.

[29] F. Woolfe, and A. Fitzgibbon, "Shift-invariant dynamic texture recognition," *ECCV*, pp. 549-562, 2006.

[30] J. Zhang, and T. Tan, "Brief review of invariant texture analysis methods," *Pattern Recognition*, vol. 35, no.3, pp. 735-747, 2002.

[31] G. Zhao, and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE PAMI*, vol. 29, no. 6, pp. 915-928, 2007.

[32] G. Zhao, and M. Pietikäinen, "Improving rotation invariance of the volume local binary pattern operator," *IAPR Conf. on Machine Vision Applications*, pp. 327-330, 2007.

[33] G. Zhao, M. Barnard, and M. Pietikäinen, "Lipreading with local spatiotemporal descriptors," *IEEE Transactions on Multimedia*, vol. 11, pp. 1254-1263, 2009.

**Guoying Zhao received the PhD degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China in 2005. From July 2005 to August 2010, she was a Senior Researcher in the Center for Machine Vision Research at the University of Oulu. From September of 2010, she is an adjunct professor at the University of Oulu. Her research interests include gait analysis, dynamic texture recognition, facial expression recognition, human motion analysis, and person identification. She has authored over 70 papers in journals and conferences, and has served as a reviewer for many journals and conferences. She gave an invited talk in Institute of Computing Technology, Chinese Academy of Sciences, July 2007. With Prof. Pietikäinen, she gave a tutorial: "Local binary pattern approach to computer vision" in 18th ICPR, Aug. 2006, Hong Kong, and another tutorial "Local texture descriptors in computer vision" in ICCV, Sep. 2009, Kyoto, Japan. She has authored/edited three books and a special issue on IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics and is editing another special issue on Elsevier Journal on Image and Vision Computing. She was a co-chair of ECCV 2008 Workshop on Machine Learning for Vision-based Motion Analysis (MLVMA), and MLVMA workshop at ICCV 2009 and CVPR 2011.**

**Timo Ahonen received Ph.D. (with honors) in information engineering in 2009 from the University of Oulu, Finland, and is now a Senior Researcher at Nokia Research Center, Palo Alto, Ca. Currently he is working on computational photography on mobile devices. His research interests include computational photography, computer vision, object and face recognition, and image processing. He visited ETH Zurich in 2002 and University of Maryland in 2005.**

**Jiří Matas received the MSc degree in cybernetics (with honours) from the CTU Prague in 1987 and the PhD degree from the University of Surrey, UK, in 1995. He has published more than 150 papers in refereed journals and conferences. His publications have more than 4000 citations in the ISI Thomson-Reuters Science Citation Index and about 10000 in Google scholar. His h-index is 21 (ISI) and 34 (Google scholar) respectively. He received the best paper prize at the British Machine Vision Conferences in 2002 and 2005 and at the Asian Conference on Computer Vision in 2007. He served in various roles at major international conferences (e.g. ICCV, CVPR, NIPS, ECCV), co-chairing ECCV 2004 and CVPR 2007. He is on the editorial board of the International Journal of Computer Vision, Pattern Recognition and IEEE Transactions on Pattern Analysis and Machine Intelligence; from January 2009 serving as associate Editor-in- chief. He is a co-inventor of two patents.**

**Matti Pietikäinen received the Doctor of Science in Technology degree from the University of Oulu, Finland, in 1982. In 1981, he established the Center for Machine Vision Research at the University of Oulu. Currently, he is a professor of information engineering, scientific director of Infotech Oulu Research Center, and leader of the Center for Machine Vision Research at the University of Oulu. From 1980 to 1981 and from 1984 to 1985, he visited the Computer Vision Laboratory at the University of Maryland. His research interests include texture-based computer vision, face analysis, activity analysis, and their applications in human-computer/robot interaction, person identification and visual surveillance. He has authored about 250 refereed papers in international journals, books, and conference proceedings. His research on texture-based computer vision, local binary pattern (LBP) methodology and facial image analysis, for example, is frequently cited and its results are used in various applications around the world. He was an associate editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence and Pattern Recognition journals, and is currently an associate editor of Image and Vision Computing Journal. He was the president of the Pattern Recognition Society of Finland from 1989 to 1992. From 1989 to 2007 he served as a member of the Governing Board of the International Association for Pattern Recognition (IAPR), and became one of the founding fellows of the IAPR in 1994. He regularly serves on program committees of the top conferences of his field. Recently, he was an area chair of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07), a co-chair of Workshops of International Conference on Pattern Recognition (ICPR '08), ECCV 2008 Workshop on Machine Learning for Vision-based Motion Analysis (MLVMA), and MLVMA workshop at ICCV 2009 and CVPR 2011. He is a senior member of the IEEE, and was the vice-chair of IEEE Finland Section.**