Facial Biometric Presentation Attack Detection Using Temporal Texture Co-occurrence

S.Pan
School of Engineering and Digital Arts
University of Kent
Canterbury, UK
sp641@kent.ac.uk

F.Deravi
School of Engineering and Digital Arts
University of Kent
Canterbury, UK
f.deravi@kent.ac.uk

Abstract

Biometric person recognition systems based on facial images are increasingly used in a wide range of applications. However, the potential for face spoofing attacks remains a significant challenge to the security of such systems and finding better means of detecting such presentation attacks has become a necessity. In this paper, we propose a new spoofing detection method, which is based on temporal changes in texture information. A novel temporal texture descriptor is proposed to characterise the pattern of change in a short video sequence named Temporal Co-occurrence Adjacent Local Binary Pattern (TCoALBP). Experimental results using the CASIA-FA, Replay Attack and MSU-MFSD datasets; the proposed method shows the effectiveness of the proposed technique on these challenging datasets.

1. Introduction

As the popularity of biometric person recognition systems have grown, the threat of spoofing or presentation attacks have also increased, potentially undermining their usefulness. This is particularly serious for facial recognition systems as it is relatively easy to create an attack artefact which may be hard to detect. The popularity of social networks where much identity-bearing facial information is freely available and cheap photographic devices aggravates this situation. Hence, face spoofing/liveness detection research has been expanding in recent years.

Facial spoofing attack detection is highly sensitive to specific spoofing attack approach, and the type of attack artefact used to subvert the system: e.g. photographic paper, video projection or (3D) mask. The quality of the artefacts and the sample acquisition sub-systems of the biometric system are crucial in the effectiveness of any spoofing attempt. Video replay attacks may need the high-resolution screen of a smartphone or tablet.

2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA)
978-1-5386-2248-3/18/$31.00 © 2018 IEEE

Paper-based attack artefacts may require the use of high quality papers and printers to reproduce facial images. Compared to other attack methods, mask attacks may be more complicated and costly. The main focus of this paper is therefore on spoofing attacks using photographic paper and video replays.

Zhang Z et, al. [15] suggested that paper-based attacks may be categorised by different attack schemes, namely (1) cut-paper attack, and (2) wrapped paper attack. Also, Chingovska I et al. [16] suggested that video attacks may be categorised by screen resolution, screen size, and whether the screen is held by hand. The different result in several scenarios that can be considered to simulate real usage.

Detecting facial spoofing attacks from static texture patterns is a fast and low-cost strategy. Some recent works using static features suggest the efficiency of this approach [8][17][18]. However, intuition suggests that it should be easier to distinguish a spoofing attack using video input rather than just a static image. These anti-spoofing schemes, using both temporal and spatial information, are categorised as feature-level dynamic facial spoofing approaches [11]. Some feature level dynamic facial spoofing approaches rely on the challenge-response strategy, which remains popular to protect against paper-based attacks and mask-based attacks. However, this strategy will lose accuracy against video-based attacks. Moreover, a long time duration for challenge-response strategy violates the non-invasive requirement and user-friendly requirement for a biometric system. Other dynamic-based anti-spoofing approaches are restricted by the high computational complexity, because the data volume of a video is larger than a static frame [11]. These difficulties emphasise the necessity of novel anti-spoofing research by adding temporal information.

2. Related work

Face spoofing detection approaches may be classified into three groups [11]: (1) Sensor-Level Techniques, (2) Feature–Level Techniques, (3) Score-Level Techniques. The Feature–Level Techniques drew more attention due to their low-cost and high-efficiency. Also, this category can be further divided into static and dynamic groups,

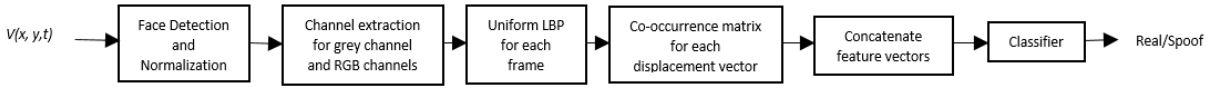

Figure 1: The block-diagram for the proposed method.

depending on whether they use temporal information [11].

There are many texture descriptors that have been used in static Feature-Level spoofing attack detection. For instance, Agarwal et al. [18] developed a spoofing detection algorithm which introduced the Haralick texture descriptor to generate their feature vector and achieved HTER=0% in 3D MAD dataset [22]. The Local Binary Pattern (LBP) descriptor proposed by Ojala et al. [3] has been used extensively by researchers for spoofing attack detection, in part due to its computational efficiency [6]. Various extensions of the LBP are also introduced to improve its performance; for instance, the Fisher score is used to reduce the overlapping block LBP operator [6]. The usability of colour extensions of LBP (CLBP) by modelling the colour characteristics of spoofing artefacts is explored in [8].Also, Boulkenafet Z et al. [17] discussed the effect of a commonly used technique "multiscale space texture representation" and developed a multi-scale LBP based anti-spoofing algorithm.

On the other hand, the dynamic Feature-Level spoofing approaches use temporal information for face spoofing detection. Here, computational complexity becomes a significant consideration due to the large data volumes involved. The early research works using temporal information aim to deploy the Fourier transform or matrix decomposition for feature extraction. Li et al [13] assume the high frequency component differences between real and fake traits at Fourier spectra. More recently, the authors in [9] use dynamic mode decomposition (DMD) to represent temporal information of a video as a single frame. They reported, Equal Error Rate (EER) =21.8% for the CASIA-FA dataset and Half Total Error Rate (HTER) =8% for the Replay-Attack dataset. However, their method is computationally intensive as it concatenates the whole video into a very large matrix and computes a matrix decomposition. Another work [10] enhance micro and macro motions by using motion magnification technique, but they are also computationally expensive. Lakshminarayana et al. [19] combine temporal information and convolutional neural network (CNN) architecture to detect spoofing attacks. The results of their experiment are good, but a CNN architecture is more computationally expensive than many other methods. Also, they didn't provide an end-to-end architecture which is normally considered as the most advantage in using deep neural network (DNN).

To solve this computational complexity problem, some researchers aims to combine LBP with temporal information. The main challenge of using LBP directly to extract temporal information is that the original LBP can only process 2D texture information. However, a 3D extension of the 2D Local Binary Pattern can be envisaged where the third dimension is time as represented by the sequence of video frames.

There are two challenges in the design of a temporal extension of the LBP feature. Firstly, there is the problem of defining a meaningful time-series extension of the LBP descriptor. Secondly, there is the problem of coping with the large data volume inherent in video processing. Selecting a set of neighbouring points in 3D can be considered as an equidistant sampling problem on a sphere, which may still be difficult to implement. Zhao and Pietikainen's proposed the Volume LBP (VLBP), in an attempt to extend 2D LBP to 3D volume data [1], [2]. They suggest their approach can be used in both video and RGB-D images. An alternative approach proposed in [5] encodes 3D local texture in a video sequence by sampling the neighbouring points defined on the surface of a ball using the Uniform LBP. However, this method may encode different textures with the same binary code [5].

For the second problem, researchers consider data selection methods as solutions. The data volume of a video cube is related to the video length/duration. T de Freitas Pereira et al. [4] followed the data selection idea, which selects three orthogonal planes X-Y, X-T, and Y-T for a simple implementation named LBP-TOP, which compresses temporal-related data by generating X-T and Y-T orthogonal planes. Inspired by this work, a number of other three orthogonal planes have been investigated. For instance, a more recent work, LDP-TOP [13], combines the high-order Local Derivative Pattern and three orthogonal planes for better performance. However, the disadvantage of using only three orthogonal planes is that some crucial information may be missed. For example in [4], the orthogonal planes are selected at the middle of frames and may thus miss some crucial information such as eye blinks.

This paper also addresses the computational efficiency problem and tries to provide a novel method to incorporate temporal information. The standard LBP feature represent entire image as a histogram of numerous LBPs. Unlike the standard LBP texture features, calculating local histograms of intensity variations, the proposed approach computes co-occurrence matrices to represent the adjacency relationships between the LBP features at various

spatiotemporal displacements. Generating a co-occurrence matrix to represent spatial information is not a new idea. The co-occurrence adjacent LBP (CoALBP) [7] was originally designed for texture pattern recognition, and was used for facial spoofing attack detection as a colour texture descriptor in [8]. This work reported promising results for the CAISA-FA dataset (EER=14.8% in the grey channel and EER=11% in the RGB colour space) as well as the Replay-attack Dataset (HTER=4.7%). However, CoALBP is restricted to spatial information, which ignores that a co-occurrence matrix can also represent the repetitiveness of temporal information. This repetitiveness suggests that a high-dimensional video cube may be represented by a low-dimensional feature vector.

The proposed Temporal Co-occurrence Adjacent LBP (TCoALBP), will capture and summarise dynamic textural characteristics of a video sequence by encoding the co-occurrence of local texture features both in space and across time, as represented by the sequence of video frames. We tested our approach on grey-scale video frame sequences as well as on RGB colour frame sequences using the CASIA-FA and Replay-Attack datasets. Following a description of the proposed feature in Section 3 we present the results of these evaluations in Sections 4 and 5.

3. Temporal Co-occurrence Adjacent LBP

The proposed Temporal Co-occurrence Adjacent LBP encodes the spatiotemporal relationship between the uniform LBP of the original point and the uniform LBP of its selected neighbourhood. The uniform LBP histogram can provide a quick and robust encoding of local textures on one image. However, a 59 bins histogram may not adequately represent texture changes of a video cube. Concatenating the histogram of each frame is one way to solve this problem. However, different numbers of frames will cause varying feature lengths. The co-occurrence matrix approach solves this problem by encapsulating both information from LBP features of a video cube in an adjustable way. The length and descriptive capability of the proposed feature can be adjusted by the number, magnitudes and directions of displacement vectors used to compute the matrices. Therefore, the TCoALBP can be adapted to different applications by optimizing its parameters.

For any frame I in the frame sequence V, $I(x,y)$ represent the pixel value located at $l_{2D}=(x,y)$ and $V(x,y,t)$ represent the pixel values located at $l_{3D}=(x,y,t)$. The uniform LBP [3] are designed using (1) (2) (3).

$$LBP_{P,R}(I) = \Sigma_{p=0}^{P-1} Sig(g_p - g_c) * 2^p \quad (1)$$

$$Sig(Z) = \begin{cases} 1 & if\ Z > 0 \\ 0 & otherwise \end{cases} \quad (2)$$

$$U(\mathrm{LBP_{P,R}(I)}) = \Sigma_{p=1}^{P-1} |Sig(g_{p-1} - g_c) - Sig(g_p - g_c)| + |Sig(g_{P-1} - g_c) - Sig(g_0 - g_c)| \quad (3)$$

where P is the number of sampling points in a circular neighbourhood set of radius R centred at l_{2D}; and g_p indicates the pixel value of p -th point on this neighbourhood. The pixel value of the central point $g_c = I(l_{2D})$ is used as a threshold for g_p. The Uniform LBP $LBP_{P,R}^{U2}(l_{2D})$ only considers the binary patterns which $U(LBP_{P,R}) \leq 2$. For instance, the "00001111" and "00111100" are uniform patterns. Before we compute the co-occurrence matrix, we calculate uniform LBP at each frame in V. The result of this new video cube is represented by $V_{P,R}^{LBP,U2}(l_{3D})$.

The co-occurrence neighbour of any pixel point located at l_{3D} can be defined by an adjacent transition vector set $A = \left\{ \forall a \in A \middle| \begin{matrix} a = (a_1, a_2, a_3) where\ a_1, a_2, a_3 \in \{0, \pm 1\} \\ and\ a_1, a_2, a_3\ cannot\ be\ 0\ together \end{matrix} \right\}$ and a parameter set $\nabla = \{\forall \tau \in \nabla | \tau = (\nabla_x, \nabla_y, \nabla_t)\ where\ \nabla_x, \nabla_y, \nabla_t \in Z^+\}$. Thus, the neighbourhood location n_{3D} of l_{3D} can be defined by (4) where "·" represent element-wise multiply:

$$n_{3D} = l_{3D} \cdot a \cdot \tau \quad (4)$$

The co-occurrence matrix H can be computed using (5):

$$H(e,f) = \sum_{x=1}^{x=M-2\times\nabla_x} \sum_{y=1}^{y=N-2\times\nabla_y} \sum_{t=1}^{t=T-2\times\nabla_t} \delta\left(d\left(V_{P,R}^{LBP,U2}(l_{3D})\right), d\left(V_{P,R}^{LBP,U2}(n_{3D})\right), e, f\right) \quad (5)$$

$$\delta(u,v,e,f) = \begin{cases} 1 & if\ u = e\ and\ v = f \\ 0 & otherwise \end{cases} \quad (6)$$

Where $e, f \in [1, Np^{LBP_{P,R}^{U2}}]$ represent the vertical and horizontal coordinates of the matrix H. The $Np^{LBP_{P,R}^{U2}}$ represent the number of all possible uniform patterns with parameter P and R. In (5), $d(.)$ represent a decimal conversion function which transform binary string to a decimal format. M and N represent the frame size and T means the number of frames of input video cube $V_{P,R}^{LBP,U2}(l_{3D})$.

4. Experimental setup

4.1. Implementation detials

Figure 1 is a block-diagram of the proposed system that was evaluated. The face areas are detected and normalised using facial landmarks to decrease the effect of the background. Some examples of fine-tuned face areas for different categories can be found in Figures 2 and 3. Then, the frames are transformed to gray channel or devided for

R,G and B channel. For each channel, the Uniform LBPs are calculated for each face area. Then, the feature historgram is calculatled by concatenating historgrams. Moreover, the number of co-occurrence matrices in our implementation are determined by the number of distinct elements in A. A different selection of A will cause a different feature size, which is fixed to 7 in our implementation to balance the speed and the performance of proposed feature. For the TCoALBP (RGB) in our experiment, the final feature vector H is generated by (7), where H_R, H_G, and H_B are TCoALBP feature vectors for red, green and blue channels of the input videos respectively. The classifier is a Support Vector Machine (SVM) with an RBF Kernel which has also been used by other researchers with whom we compare our results.

$$\mathrm{H} = \{H_R, H_G, H_B\} \quad (7).$$

4.2. Datasets and performance metrics

There are three widely used anti-spoofing benchmarking datasets which were used to evaluate the effectiveness of the proposed anti-spoofing algorithm: the CASIA Face Anti-Spoofing (CASIA_FA) dataset, the Replay-Attack database, and MSU MFSD dataset. All of these datasets include some recordings of genuine client access attempts and various presentation attacks. The pre-defined evaluation protocol for each dataset was followed for a fair evaluation and comparison with the state-of-the-art.

The CASIA-FA database [15] consists of 600 video clips which include both real and spoofing access attempts, totally, there are 50 individuals listed in the dataset, where the spoofing artefacts are produced from high-quality records of genuine faces. Three different attack artefacts are included: warped photo attacks, cut photo attacks, and video attacks. All of them are designed to simulate real attack attempts. For instance, the cut photo attack is a special photo attack, in which a high-quality face is printed on paper, but where the area surrounding the eyes is cut to subvert eye-motion-based spoofing attack detection methods. Three different image resolutions are used in this dataset to simulate different usage conditions, namely low resolution, normal resolution, and high resolution. In their evaluation scenarios, 50 subjects are split into two categories: the training set (20 subjects) and the test set (30 subjects). They also designed seven detailed scenarios which are: (1) l*ow-quality*, (2) *normal-quality* (3) *high-quality,* (4) *warped photo attacks*, (5) *cut photo attacks*, and (6) *video attacks*. The (1), (2), and (3) scenarios are used to test the robustness at different image quality conditions. The (4), (5), and (6) scenarios are used to simulate different attack behaviours. The overall test scenario (7) provides combined performance test results for all attack types and qualities.

The Replay-Attack database [16] is another widely used face spoofing dataset which contains various attack behaviours and contains 1300 video clips. There are 50 clients recorded for both real access attempts and 3 different attack behaviours. Two illumination conditions are considered: *controlled* and *adverse*. For each condition, three attack categories are included: (1) *print attacks*, (2) *mobile attacks*, and (3) *highdef attacks*. The mobile attacks and highdef attacks can both be categorised as video attacks but use different sizes of the screen with different resolutions. They also consider various conditions about whether the attack device is fixed in front of the camera:(1) hand based attack (the attack devices were held by hand) and (2) fixed-support attacks (the attack devices were fixed on a stand). The Replay-Attack database divides the whole datasets into three subsets, which are: the training set, the development set, and the testing set.

The MSU Mobile Face Spoof Database [21], which consists of 280 video recordings of real and fake faces, addresses the challenge of using a low qualitymobile camera. They use a built-in camera of MacBook Air 13-inch laptop (640 × 480) and a front camera of a Google Nexus 5 Android phone (720 ×480) to capture videos with at least nine seconds duration for all 35 subjects. This dataset includes both video and paper attacks. The high-quality biometric samples are taken by the Canon 550D camera and the back camera of an iPhone 5S. There are two screens used to generate video attacks which are an iPad Air screen and an iPhone 5S screen. For the printed attacks, HD pictures (5184×3456) were printed on A3 paper using an HP colour Laserjet CP6015xh printer. They require two subject-disjoint subsets for training and testing (15 and 20 subjects) in their evaluation protocol.

Various environmental conditions can affect performance; i.e., different image qualities, different distances between the face and the camera, different face angles, and background changes. Thus, we follow the pre-defined scenarios at CASIA-FA dataset for a detailed performance test at various conditions and attack types. Also, the evaluation protocols defined for these two datasets are followed for fair comparison with the state-of-the-art methods. The CASIA-FA and MSU MFSD datasets do not contain a development set for fine-tuning of parameters. Thus, we use four-fold subject-disjoint cross-validation on the training set to train the classifier and fine-tune parameters. After that, the experiments evaluate the performance by calculating the Equal Error Rate (EER) on the test set [15]. The Replay-Attack database contains a development subset for parameter fine-tuning. The experiments follow their protocol to produce the EER on the development set and the Half Total Error Rate (HTER) on the test set [16]. To make the error trade-off clearly, we report some results with True Positive Rate (TPR) where False Accept Rates (FAR) are fixed to 0.01 and 0.1.

5. Result and analysis

The effectiveness and robustness of the proposed system using the TCoALBP features have been tested for different attack types and environment conditions using the CASIA-FA, Replay-Attack, and MSU-MFSD databases and the results are reported in this section and compared with some state-of-the-art techniques. Additionally the selection of different parameters are tested at different dataset.

5.1. Parameter optimization

Three different parameter selection experiments were conducted to improve the final result and show some properties of the proposed method. All of them provide results using the pre-defined overall test in the protocol for each dataset by implementing TCoALBP with different parameters. All tables include three different columns of results which are EER for CASIA-FA overall test, HTER for Replay-Attack overall Dataset, and EER for MSU MFSD overall test.

Table 1. Performance for TCoALBP for different *A* sets (grey-scale video, 30 frames and parameters are fixed to $(P,R,\nabla_x, \nabla_y, \nabla_t)=(4,1,1,1,1)$).

A	CASIA overall test (EER)	R-A overall test (HTER)	MSU MFSD overall (EER)
{(1,0,0),(0,1,0),(-1,0,0),(1,1,0),(-1,1,0),(1,-1,0),(0,-1,0)}	12.1%	11.8%	18.7%
{(1,0,0),(0,1,0),(1,1,0),(-1,1,0),(1,-1,0),(-1,0,0),(0,0,1)}	10.16%	7.01%	20.01%
{(1,0,1),(0,1,1),(-1,0,1),(1,1,1),(-1,1,1),(1,-1,1),(0,-1,1)}	8.69%	6.07%	16.60%

Firstly, the neighbourhood location is an important parameter for TCoALBP. Table 1 includes the result of TCoALBP($(P,R,\nabla_x, \nabla_y, \nabla_t)=(4,1,1,1,1)$) for grey-scale video input with 30 frames but with different *A* sets. The first row of the *A* set only contains spatial correlation with $\nabla_t = 0$, which can be considered as CoALBP. The second row of set *A* includes two parts: (1) neighbour sub-set only including spatial displacements (2) neighbour sub-set only including temporal displacement. The third row of set *A* includes neighbours with both spatial and temporal displacements. Clearly including both special and temporal displacements improves performance.

Secondly, the video duration is considered as a parameter in this paper. Table 2 shows performance at different video durations, where TCoALBP($(P,R,\nabla_x, \nabla_y, \nabla_t)=(4,1,1,1,1)$) using grey-scale video as input. Generally, longer video duration can improve system performance especially from the row of 30 frames to the row of 60 frames.

Table 2. Performance for TCoALBP on different frame numbers and grey-scale video (parameters are fixed to $(P,R,\nabla_x, \nabla_y, \nabla_t)=(4,1,1,1,1)$).

Different frame numbers	CASIA overall test (EER)	R-A overall test (HTER)	MSU MFSD overall (EER)
5	23.33%	18.08%	34.35%
10	16.67%	15.97%	27.54%
15	15.42%	11.44%	27.74%
20	13.53%	9.81%	24.01%
25	11.15%	7.33%	15.77%
30	8.69%	6.07%	16.60%
60	7.96%	**5.32%**	14.29%
100	8.02%	5.94%	**12.33%**
All frames	**7.62%**	5.88%	14.97%

Table 3. Performance for different TCoALBP parameters on 60 frames and grey-scale video (bold type indicates best result).

Different parameters $(P,R,\nabla_x, \nabla_y, \nabla_t)$	CASIA overall test (EER)	R-A overall test (HTER)	MSU MFSD overall test (EER)
(4,1,1,1,1)	9.96%	6.07%	16.60%
(4,1,1,2,2)	**7.33%**	6.54%	15.72%
(4,1,1,1,3)	13.33%	5.70%	16.94%
(4,1,2,2,1)	15.88%	11.04%	24.80%
(4,1,2,2,2)	16.27%	10.34%	18.89%
(4,1,2,2,3)	11.21%	8.09%	17.41%
(8,1,1,1,1)	11.51%	8.37%	19.81%
(8,1,1,1,2)	9.97%	7.88%	15.87%
(8,1,1,1,3)	10.28%	6.73%	17.24%
(8,1,2,2,1)	11.70%	8.99%	16.29%
(8,1,2,2,2)	10.33%	7.81%	18.86%
(8,1,2,2,3)	9.16%	6.90%	16.13%
(4,2,1,1,1)	19.66%	8.50%	23.17%
(4,2,1,1,4)	15.55%	10.41%	25.42%
(4,2,1,1,6)	17.68%	9.27%	**14.09%**
(4,2,4,4,1)	14.83%	7.57%	16.37%
(4,2,4,4,4)	16.91%	6.79%	18.10%
(4,2,4,4,6)	15.00%	**5.21%**	15.88%
(4,2,6,6,1)	17.70%	11.60%	23.73%
(4,2,6,6,4)	14.30%	13.60%	19.03%
(4,2,6,6,6)	11.24%	9.11%	14.36%

Thirdly, there are five parameters *(P, R,*$\nabla_x, \nabla_y, \nabla_t$) that can directly change the way TCoALBP encodes temporal information. The experiments of Table 3 are designed for testing the influence of these parameters. Here, the relationship between R and $\nabla_x, \nabla_y, \nabla_t$ is explored in Table 3.

Table 3 illustrate results for different parameter sets. This work explored uniform LBP with different parameters

(P, R)={(4,1),(8,1),(4,2)} to test the impact of different radius *R* and the different number of sampling points *P* with different ∇ sets. The magnitude of displacement $\nabla_x, \nabla_y, \nabla_t$ can be roughly split into three subtypes: (1) ∇>2R,(2) ∇=2R,(3) ∇<2R. From the result of these three subtypes, the selection of $\nabla_x, \nabla_y, \nabla_t$ does appear to affect system performance, $\nabla_t \geq 2R$ slightly improves the system performance. Also, Table 3 implies that optimizing the ($P, R, \nabla_x, \nabla_y, \nabla_t$) parameter set may not be a convex optimization problem.

5.2. Intra-dataset results and comparison

Tables 4 and 5 provide the results of the grey-scale TCoALBP descriptor (TCoALBP (grey)), the RGB channel concatenated TCoALBP descriptor (TCoALBP (RGB)), the grey-scale CoALBP (CoALBP (grey)), and the grey-scale LBP. All of these descriptors use fine-tuned parameters for improving performance. The grey-scale LBP and CoALBP (grey) are provided as baseline results for comparison. For some static local texture descriptors, colour channels are believed to provide more information than the grey-scale image [8]. Thus, we design the TCoALBP (RGB), which divide the video cube into separate RBG colour channels in order to compute TCoALBP on different channels independently and concatenate the resulting feature vectors from different colour channels before classification.

Table 4 also shows the performance results of different scenarios in CASIA-FA dataset. Results presented in Table 4 and 5 suggest that the TCoALBP features can significantly improve the system performance by combining temporal information and static texture information. Comparing the results of the original LBP and TCoALBP (RGB), the proposed method shows a 65.2% performance improvements for the CASIA-FA and 92.7% for the Replay-Attack datasets. Also, TCoALBP (RGB) shows an improved the performance on the CASIA-FA databases by 41.6% respectively compared with the grey-scale CoALBP.

Table 4. CASIA-FA test results in terms of EER (%) at different Scenarios:(1) low quality, (2) normal quality and (3) high-quality (4) warped photo attacks, (5) cut photo attacks, (6) video attack, and (7) overall test

Scenarios / Features	1	2	3	4	5	6	7
LBP-baseline[15]	16.5	17.2	23.4	25.1	17.6	26.7	25.0
CoALBP(grey)[8]	16	15.2	14.6	13.7	14.6	17.3	14.9
TCoALBP(grey)	9.7	**8.1**	8.9	10.3	9.1	8.4	8.69
TCoALBP(RGB)	**5.7**	7.3	**6.6**	**8.1**	**6.9**	**7.1**	**6.71**

Table 5. Replay-Attack DB overall results.

	EER (%)	HTER (%)
LBP^{U2}[8]	17.9	13.7
CoALBP(grey)[8]	12.9	16.7
CoALBP(RGB)[8]	6.2	8.0
TCoALBP(grey)	2.4	5.7
TCoALBP(RGB)	**0.1**	**0.6**

Table 6. Comparisons with the state-of-the-art.

	CASIA-FA (EER %)	Replay-Attack (HTER %)	MSU MFSD (EER%)
CoALBP(RGB) [8]	11.1	8.0	17.7
DMD[9]	21.8	3.8	N/A
Motion-meg[10]	14.4	**0.0**	N/A
LBP-TOP[12]	10.6	N/A	N/A
LDP-TOP[13]	8.9	1.7	N/A
CNN[19]	1.1	0.8	N/A
Multi-scale LBP(RGB) [17]	10.7	5.1	11.7
Proposed method	6.71	0.6	10.07

Table 6 shows the results of the comparison between the proposed method and some state-of-the-art methods, which contains the best results for the dynamic features attempting to use temporal information: DMD [9], Motion-seg [10], and LBP-TOP [4] [12]. The proposed method shows very competitive results on the challenging CASIA-FA database and the Replay-Attack database. Some approaches included in Table 6 report better results than the proposed method for some of the datasets. However, the proposed approach outperforms these methods in other datasets. For instance, a CNN-based method [19] is reported to have a very competitive result for the CASIA-FA dataset. However, the proposed method produces better results than which is reported in [19] for the Replay-Attack dataset.

6. Conclusion and future works

In this paper, we proposed a novel feature for biometric presentation attack detection using temporal texture co-occurrence in a video sequence of facial images. The effectiveness of different temporal texture representations was studied by extracting TCoALBP features from grey-channel image sequences as well as RGB colour channel image sequences. Extensive experiments showed good results on three challenging spoofing detection databases: CASIA-FA, MSU-MFSD and Replay Attack Dataset. On CASIA-FA database, the result of TCoALBP (RGB) feature reaches the state-of-the-art level. Furthermore, in the intra-database evaluation, TCoALBP (RGB) feature shows very promising generalisation capabilities. The TCoALBP algorithm requires the optimisation of several parameters for different datasets to reach the best performance. Also, the inclusion of colour information did not result in a significant performance improvement in the

experiments. More experiments may be needed to optimise the parameters for different colour channels. For future work, the effect of different temporal and special displacements for texture co-occurrence will be studied using larger and more challenging datasets. Also, heuristic search algorithms may be explored for optimizing parameter sets.

References

[1] Zhao G and Pietikäinen M, "Experiments with Facial Expression Recognition using Spatiotemporal Local Binary Patterns," presented at the Multimedia and Expo, 2007 IEEE International Conference on, pp. 1091–1094.

[2] Zhao G, Pietikainen M. "Dynamic texture recognition using local binary patterns with an application to facial expressions" [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6), pp. 915 - 928.

[3] Ojala T, Pietikainen M, Maenpaa T. Multiresolution "Gray-scale and rotation invariant texture classification with local binary patterns" [J]. IEEE Transactions on pattern analysis and machine intelligence, 2002, 24(7):pp. 971-987.

[4] De Freitas Pereira, T., Komulainen, J., Anjos, A., De Martino, J. M., Hadid, A., Pietikäinen, M., & Marcel, S... "Face liveness detection using dynamic texture" [J]. EURASIP Journal on Image and Video Processing, 2014, 2014(1):pp. 1-15.

[5] Paulhac L, Makris P, Ramel J Y. "Comparison between 2D and 3D local binary pattern methods for characterisation of three-dimensional textures"[C]. International Conference Image Analysis and Recognition. Springer Berlin Heidelberg, 2008:pp. 670-679.

[6] Benlamoudi A, Samai D, Ouafi A, et al. "Face spoofing detection using Local binary patterns and Fisher Score"[C]. Control, Engineering & Information Technology (CEIT), 2015 3rd International Conference on. IEEE, 2015: 1-5.

[7] Nosaka R, Ohkawa Y, Fukui K. Feature extraction based on co-occurrence of adjacent local binary patterns [J]. Advances in image and video technology, 2012:pp. 82-91.

[8] Boulkenafet Z, Komulainen J, Hadid A. "Face spoofing detection using colour texture analysis" [J]. IEEE Transactions on Information Forensics and Security, 2016, 11(8):pp. 1818-1830.

[9] Tirunagari S, Poh N, Windridge D, et al. "Detection of face spoofing using visual dynamics" [J]. IEEE Transactions on Information Forensics and Security, 2015, 10(4):pp. 762-777.

[10] Bharadwaj S, Dhamecha T I, Vatsa M, et al. "Computationally efficient face spoofing detection with motion magnification"[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2013:pp. 105-110.

[11] Galbally J, Marcel S, Fierrez J. "Biometric antispoofing methods: A survey in face recognition" [J]. IEEE Access, 2014, 12:pp. 1530-1552.

[12] De Freitas Pereira T, Anjos A, De Martino J M, et al. "Can face anti-spoofing countermeasures work in a real world scenario?" [C]. Biometrics (ICB), 2013 International Conference on. IEEE, 2013: 1-8.

[13] Li J, Wang Y, Tan T, et al. "Live face detection based on the analysis of fourier spectra"[C]. Defense and Security. International Society for Optics and Photonics, 2004:pp. 296-303.

[14] Phan Q T, Dang-Nguyen D T, Boato G, et al. "FACE spoofing detection using LDP-TOP[C]". Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, 2016:pp. 404-408.

[15] Zhang Z, Yan J, Liu S, et al. "A face antispoofing database with diverse attacks[C]" Biometrics (ICB), 2012 5th IAPR international conference on. IEEE, 2012: 26-31.

[16] Chingovska I, Anjos A, Marcel S. "On the effectiveness of local binary patterns in face anti-spoofing[C]". Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the. IEEE, 2012: 1-7.

[17] Boulkenafet Z, Komulainen J, Feng X, et al. Scale-space texture analysis for face anti-spoofing[C]//Biometrics (ICB), 2016 International Conference on. IEEE, 2016: 1-6.

[18] Agarwal, Akshay, Richa Singh, and Mayank Vatsa. "Face anti-spoofing using Haralick features." Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on. IEEE, 2016.

[19] Lakshminarayana, Nagashri N., et al. "A discriminative spatio-temporal mapping of face for liveness detection." Identity, Security and Behavior Analysis (ISBA), 2017 IEEE International Conference on. IEEE, 2017.

[20] Yin, Wenze, Yue Ming, and Lei Tian. "A face anti-spoofing method based on optical flow field." Signal Processing (ICSP), 2016 IEEE 13th International Conference on. IEEE, 2016.

[21] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," IEEE Trans. Inf. Forensics Security, vol. 10, no. 4, pp. 746–761, Apr. 2015

[22] Erdogmus, Nesli, and Sébastien Marcel. "Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect." Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on. IEEE, 2013.