ANALYSIS OF TEXTURAL FEATURES FOR FACE BIOMETRIC ANTI-SPOOFING

Muhammad-Adeel Waris, Honglei Zhang, Iftikhar Ahmad, Serkan Kiranyaz, Moncef Gabbouj

Department of Signal Processing,
Tampere University of Technology, Tampere, Finland

ABSTRACT

True authentication predicted on biometrics has received upsurge attention during the last few years, as it provides facile way to access the system through basic physical and behavioral characteristics. Face recognition being a non-intrusive recognition requires less participation from the user compared to iris, speech and fingerprint based biometric. Resistance to false authentication from photographs and video playbacks is a vigorous issue for successful biometric system. This paper analyzes different textural features and proposes a novel approach for anti-spoofing solution. Experiments were conducted on a publicly available face spoofing database REPLAY-ATTACK to validate textural analysis over a database containing printed photographs, photos and videos displayed on electronic screens. Results show that the approach is superior to the other existing state of art approaches tested on same database.

Index Terms— anti-spoofing, face biometric, textural analysis, counter measure.

1. INTRODUCTION

The use of biometric as a primary source of authentication in commercial application has been communal in last few years [1]-[3]. Contrary to other biometric techniques face recognition is more evident as it requires a lot less interaction from the user. Despite of the advances in biometric authentication systems they can still be deceived in one way or the other, e.g., Duc et al. in [4] showed how to successfully spoof a laptop verification system using only a printed photograph. Face biometric spoofing can be categorized in mainly in three categories [6];

- Photo attacks, showing printed attacks or a video sequence of pictures of the authorized user.
- Video attacks, displaying a dynamic scene video of the valid user.
- Showing a 3D face model of the valid user to the biometric system.

Producing an accurate 3D face model of a valid user might be demanding and need some expertize but other two attack scenarios can be implemented easily due to the fact of growing social media forums such as Facebook, Twitter, Linkedin, etc.

Schwartz et al. in [7] categorized current anti-spoofing methods into four groups: data driven characterization, user behavior modeling, user interaction need, and presence of additional devices. Possible solutions to this problem may be engaging additional devices such as deploying an additional 2-D camera, depth camera, thermal sensor, or implementing a human computer interaction interface asking the user to make a particular gesture for authentication. Since, such solutions are intrusive and may not be feasible in the existing systems. So, there is an imminent need to introduce an approach for detecting the spoofing attempts without any additional hardware.

In this context, this paper focuses on textural analysis for the cases of photo and video attacks. An extensive set of experiments and due observations showed that spoof video and photos intimate textural changes due to noise enforcement during the recapturing processes [5]. Our hypothesis is that noise and textural information is sufficient to classify real and spoof access. We captured distinct textural characteristics of real and spoof video sequences based on textural features to classify real and spoof videos.

The rest of the paper is organized as follows: Section 2 discusses current state-of-the-art for the facial anti-spoofing measures. In Section 3 different experiments along with region of interest selection, textural feature extraction and classification are explained, followed by experimental results in Section 4. Conclusions and directions for future research are given in Section 5.

2. LITERATURE REVIEW

Different counter measure techniques have been developed to avoid spoofing attempts such as liveness detection, analysis of Fourier spectrum, global motion detection and textural analysis of valid and fake accesses [6]- [16]. Pan et al. [6] proposed a real-time liveness detection approach against photograph spoofing, by conditional modeling of spontaneous eye blinks. The later work by the same authors [8] proposed counter-measure, which include a background context matching that helps avoiding video-spoofing in fixed face biometric systems. Tan et al. [9], proposed a solution based on extracting Difference-of-Gaussian (DoG) and variational retinex features to estimate the Lambertian reflectance properties and distinguish between valid and fake users on NUAA Database [9]. Kollreider et al. [10] used a heuristic classifier based on optical flow analysis that evaluates the trajectories of selected parts of a face region. Anjos et al. in [14] presented a motion-based solution that detects correlations between the person's head movements and the scene context. Pinto et al. in [15] proposed a face classification method based on Gray Level Co-occurrence Matrix (GLCM) feature after extracting noise signatures and calculating the Fourier spectrum on logarithmic scale to create visual rhythms in spoofed videos. Schwartz et al. in [7] presented an anti-spoofing solution based on a set of low level descriptor Histogram of Oriented Gradients (HOG), GLCM and Histograms of Shearlet Coefficients (HSC) using partial least square regression. Kose et al. [16] in proposed an anti-spoofing solution based on textural and contrast measure using Local Binary Patterns Variance (LBPV) with global matching. Chingovska et al. in [12] tested the variants of LBP features on face regions concluded that histogram of Uniform Local Binary Patterns $(LBP_{8,1}^{u2})$ produced the best result. Similar work was proposed by Pereira et al. in [13] against face spoofing attacks using the LBP−TOP (LBP from Three Orthogonal Planes) descriptor combining both space and time information into a single descriptor.

Most of the previously counter measures using on textural analysis [7] [12] [13] have taken textural analysis only over the face region and thus they are directly dependent on the face detection. Due to the fact that face detection is an erroneous process, such an approach may lead to performance degradations. Besides that, there are crucial clues around the face that can contribute to the accuracy of the spoof classification. Therefore, the proposed method in this paper performs textural analysis over the entire image, which significantly improves classification of real and spoofing images/videos. Also, concatenating different features assists gathering more distinguishable textural differences in both classes. Detailed information on the proposed approach is presented in the following section.

3. TEXTURAL ANALYSIS OF SPOOF ATTACKS

In this section, we describe the details of experiments performed to validate textural analysis on spoofing attacks. We tried to capture texture characteristics of non-live video sequence based on textural features, namely, Rotation Invariant Uniform Local Binary Patterns features $(LBP_{P,R}^{riu2})$ Gabor, GLCM and their different variations. In the notation $LBP_{P,R}^{riu2}$, the superscript $riu2$ stands for rotation invariant uniform LBP, while the subscripts P, R refer to the number of points P which form the LBP code and are taken on a circle of radius R around the central pixel.

Videos are re-encoded to Audio Video Interleave (AVI) file format from QuickTime Movie (MOV) file format at bit rate 576 kbps without changing the resolution of the video. All experiments are solely based on textural analysis such that no pre-filtering is performed, because pre-filtering will eventually reduce the effect of noise signatures and artifacts present in the videos.

REPLAY-ATTACK database [12] contains a training set of 360 videos including 60 real accesses, 60 printed, 120 photo and 120 video attacks. Photo and video attacks were further categorized into mobile and high definition photographs. The test set contains 480 videos including 80 real accesses, 80 printed, 160 photo and 160 video attacks. The resolution of each video is 320 (width) x 240 (height) pixels with a frame rate of 25 frames-per-second and contains 240 frames for each attack videos and 375 frames for each real access video [12]. So each attack video is about 10 seconds in duration and can either be:

- Hand-based attacks (i.e., the video is recaptured while the operator holds the attack media or device using their own hands).
- Fixed-support attacks (i.e., the attack device is fixed to some support such as a tripod, or in case of print attack, attached to the wall so that they do not move during the spoof attempt).

3.1. Region of interest selection

As aforementioned, previously implemented textural counter measures [7] [12] [13] extract textural features only on facial region. We observed that extracting histogram of $LBP_{16,2}^{riu2}$ on full image scene adds more clues for detecting the spoofing attempts than applying on the face region alone. Since, the Fixed Pattern Noise (FPN) and Photo-Responsiveness of Non-Uniformity (PRNU) [5] induced in the recapturing process are more discriminative on the surrounding region. Added benefit is that the face detection process is not error free; therefore, the proposed technique provides far better performance as shown in Table 1.

Table 1: Classification performance of Rotation Invariant Uniform Local Binary Patterns on face region versus entire image

Features	Classification Rates (SVM)
$LBP_{16,2}^{riu2}$ (Face Region)	85.21%
$LBP_{16,2}^{riu2}$ (Entire Image)	97.50%

3.2. Feature Extraction and Classification

A video, V, can be defined as a 2-D sequence of N frames, each frame as a function $f(x,y)$ of luminous intensities. Rotation-Invariant Uniform LBP feature vector $f_1(V)$ for a video V is computed by averaging LBP histograms of all N

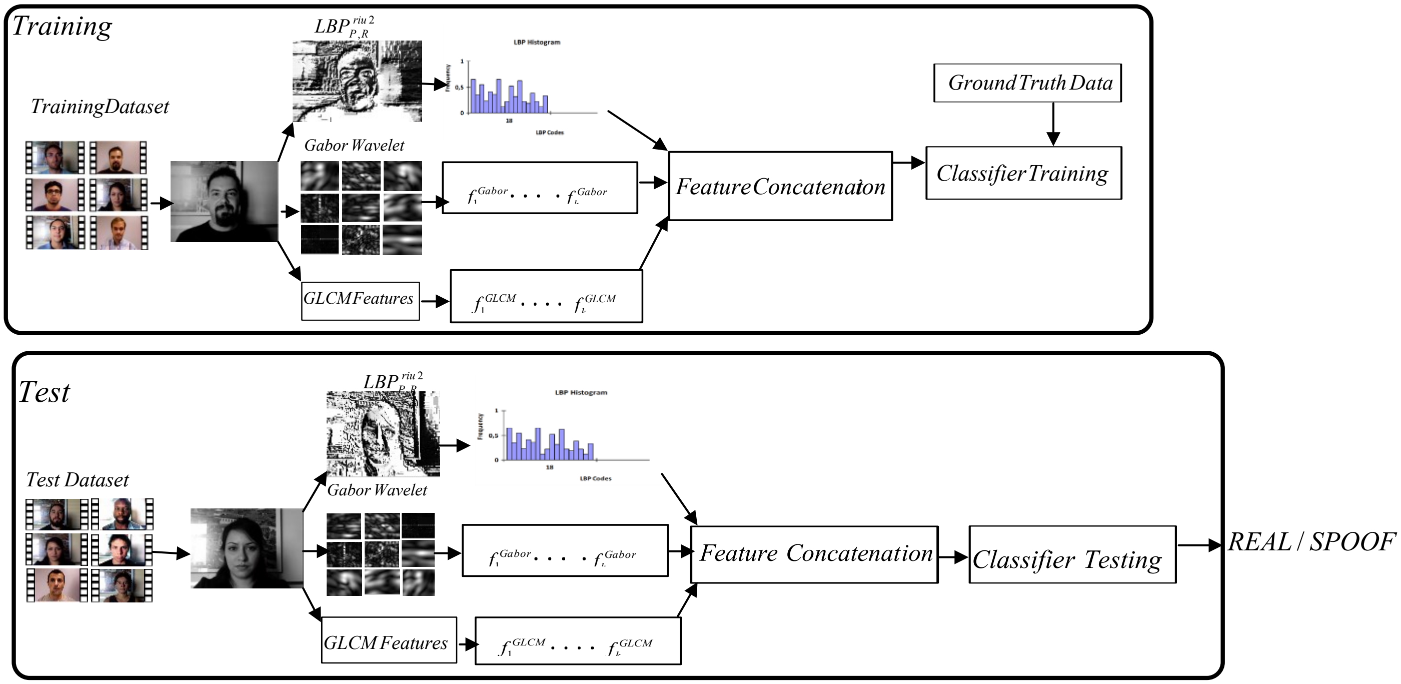

Figure 1: Block diagram of texture based Anti-Spoofing Experiments

frames of the video. According to equation (1), histogram per frame is individually computed as follows:

$$LBP^{riu2}_{P,R} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & if U(LBP_{P,R}) \le 2 \\ P+1 & otherwise \end{cases} \quad (1)$$

$$U(LBP_{P,R}) = \left| s(g_{P-1} - g_c) - s(g_0 - g_c) \right| + \sum_{p=1}^{P-1} \left| s(g_p - g_c) - s(g_{p-1} - g_c) \right| \quad (2)$$

where g_c corresponds to the grayscale value of the center pixel, g_p is the grayscale values of the *p* equally spaced pixels on the circle of radius *R*, and $s(g_p - g_c)$ is the threshold function of grayscale pixels g_c and g_p as described in [17].

$$f_1(V) = \sum_{i=1}^{N} \frac{hist(LBP^{riu2}_{P,R}(I_i(x,y)))}{N} \quad (3)$$

Similarly, for the video *V* its Gabor feature vector $f_2(V)$ is constructed by averaging *N* Gabor feature vectors computed on all *N* frames. Given the frame *I(x, y)*, its Gabor wavelet transform is defined as,

$$W_{mn}(x,y) = \int I(x_1, y_1) g^*_{mn}(x - x_1, y - y_1) dx_1 dy_1 \quad (4)$$

$$\mu_{mn} = \iint \left| W_{mn}(xy) \right| d_x d_y \quad (5)$$

$$\sigma_{mn} = \sqrt{\iint (\left| W_{mn}(x,y) \right| - \mu_{mn})^2 d_x d_y} \quad (6)$$

where, g^* in equation (4) specifies the complex conjugate. We extract Gabor wavelet features with 4 scales, *S*=4 and 6 orientations, K=6. The feature vector is then constructed using μ_{mn} and σ_{mn}, where μ_{mn} in equation (5) is the mean and σ_{mn} in equation (6) is standard deviation of the magnitude of transform coefficients [18].

$$f_{Gab}(I(x,y)) = [\mu_{00}\sigma_{00}\mu_{01}\sigma_{01} \ \dots \ \mu_{35}\sigma_{35}] \quad (7)$$

$$f_2(V) = \sum_{i=1}^{N} \frac{f_{Gab}(I_i(x,y))}{N} \quad (8)$$

Gray-Level Co-occurrence Matrix (GLCM) describes how often different combinations of gray levels co-occur in an image. Just like $f_1(V)$ and $f_2(V)$, GLCM feature $f_3(V)$ for every video is computed by averaging *N* feature vectors. The feature vector of the GLCM for single image *I(x, y)* is formed as follows:

$$f_{GLCM} I(x,y) = [r_1 r_2 r_3 \dots r_{23}] \quad (9)$$

where, $r_1, r_2, r_3 \ \dots, r_{23}$ refer to energy, entropy, contrast, homogeneity, correlation, autocorrelation, shade, dissimilarity, cluster shade, cluster prominence and other descriptors, e.g. see [19]-[21].

$$f_3(V) = \sum_{i=1}^{N} \frac{f_{GLCM}(I_i(x,y))}{N} \quad (10)$$

To take benefit of rich textural information retrieved from above three features obtained using equations (3), (8) and (10), three more features were formed by simple approach of concatenation.

Figure 1 shows the process of acquiring the features using equations (3), (8) and (10); formation of the feature vectors by concatenation; and finally the classification of real and spoof attacks based on concatenated feature vectors.

Support Vector Machines (SVM) [22] and PLS (Partial Least Square) regression method [23] are used to classify the extracted features. For SVM we used the linear kernel, to transform the original data onto higher dimensions. The SVM finds an optimal hyper plane to divide the input data into two distinct classes. PLS regression simplifies the problem and combines features from Principal Component Analysis (PCA) and multiple regression.

4. EXPERIMENTAL RESULTS

To examine the effectiveness of obtained feature set in Section 3.2, two types of experiments are done. The first experiment includes the training of two different classifiers using the entire training set, and the results are shown in Table 2.

Table 2: Classification performance of different features over entire training dataset

Features	Classification Rate (SVM)	Classification Rate (PLS)
$LBP_{16,2}^{riu2}$	98.54%	99.37%
Gabor	94.16%	96.25%
GLCM	91.04%	96.04%
$LBP_{16,2}^{riu2}$ + Gabor	98.12%	100.00%
$LBP_{16,2}^{riu2}$ +GLCM	97.91%	100.00%
GLCM+Gabor	94.79%	96.04%

For further validation to the fact that there are distinguishable textural patterns of noise present in the recaptured videos, a more challenging experimental setup is created which uses half of the training data of each sub-category of attack, real videos while keeping the testing dataset as is. The results are given in Table 3.

Table 3: Classification performance of different features over half training set

Features	Classification Rate (SVM)	Classification Rate (PLS)
$LBP_{16,2}^{riu2}$	97.50%	98.75%
Gabor	95.41%	93.95%
GLCM	88.75%	95.95%
$LBP_{16,2}^{riu2}$ + Gabor	97.70%	100.00%
$LBP_{16,2}^{riu2}$ +GLCM	97.91%	97.79%
GLCM+Gabor	95.41%	94.37%

A spoofing detection system is often subjected to two types of errors, an attack is accepted (false acceptance), or either the real access is rejected (false rejection). Performance is measured with Half Total Error Rate (HTER), which is half of the sum of the False Rejection Rate (FRR) and the False Acceptance Rate (FAR) [12]. The HTER percentage of aforementioned classification is shown in Table 4.

Table 4: HTER (%) for different features on half training dataset

Features	HTER % (SVM)	HTER % (PLS)
$LBP_{16,2}^{riu2}$	4.50%	2.25%
Gabor	9.25%	10.62%
GLCM	26.25%	10.12%
$LBP_{16,2}^{riu2}$ + Gabor	4.37%	0.00%
$LBP_{16,2}^{riu2}$ +GLCM	3.25%	0.12%
GLCM+Gabor	9.25%	10.37%

5. CONCLUSIONS

Due to the importance of providing non-vulnerable biometric systems based on facial traits, this paper investigates different textural characteristics between real and spoof videos. Many previously proposed solutions in this regard are only taking textural analysis of the face region alone. The experiments proved that the surrounding region of the face also contains dominant clues for detection of spoofing attempts. In order to employ the differences in textural characteristics between live and fake videos, texture information is exploited by using, Rotation Invariant Uniform Local Binary Patterns, Gabor and GLCM respectively. Experimental results showed Gabor feature along with Rotation Invariant Uniform Local Binary Patterns is more reliable for capturing the textural differences between both classes. Finally, a direction for future work includes the textural analysis on other publically available databases namely NUAA photo-impostor Database [9], BERC Webcam Database and BERC ATM Database.

6. REFERENCES

[1] N. Zamani, M. Darus, S. Abdullah, and M. Nordin, “Multiple-frames Super-resolution for Closed Circuit Television Forensics”, International Conference on Pattern Analysis and Intelligent Robotics, vol. 1, pp.36–40, 2011.

[2] A. K. Jain and A. Ross, “Handbook of Biometrics”, Springer, ch.Introduction to Biometrics, pp. 1–22, 2008.

[3] A. Jain and B. Klare,“Matching Forensic Sketches and Mug Shots toApprehend Criminals”, Computer, vol. 44, no. 5, pp. 94–96, 2011.

[4] N. M. Duc and B. Q. Minh, “Your face is not your password”, Black Hat Conference, 2009.

[5] J. Lukas, J. Fridrich, and M. Goljan, “Digital Camera Identification from Sensor Pattern Noise”, IEEE Trans. on Information Forensics and Security, vol. 1, no. 2, pp. 205–214, 2006.

[6] G. Pan, Z. Wu, and L. Sun, “Recent Advances in Face Recognition”, ch. Liveness Detection for Face Recognition, InTech, pp. 235–252, 2008.

[7] W. R. Schwartz, A. Rocha, H. Pedrini, “Face Spoofing Detection through Partial Least Squares and Low-Level Descriptors”, International Joint Conference on Biometrics (IJCB), 2011.

[8] G. Pan, L. Sun, Z. Wu, S. Lao, “Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcamera”, 11th IEEE ICCV, 2007.

[9] X. Tan, Y. Li, J. Liu, and L. Jiang, “Face Liveness Detection from a Single Image with Sparse Low Rank Bilinear Discriminative Model”, European Conference on Computer Vision, pp. 504–517, 2010.

[10] K. Kollreider, H. Fronthaler, and J. Bigun, “Non-intrusive liveness detection by face images”, Image and Vision Computing, vol. 27, no. 3, pp. 233–244, 2009.

[11] G. Pan, L. Sun, Z. Wu, and Y. Wang, “Monocular camera-based face liveness detection by combining eye- blink and scene context”, Journal of Telecommunication Systems, 2009.

[12] I. Chingovska, A. Anjos and S. Marcel, “On the Effectiveness of Local Binary Patterns in Face Anti-spoofing”, International Conference of the Biometrics Special Interest Group, 2012.

[13] T. Pereira, A. Anjos, J. Martino and S. Marcel, “LBP − T OP based countermeasure against face spoofing attacks”, Asian Conference on Computer Vision, 2012.

[14] A. Anjos and S. Marcel, “Counter-Measures to Photo Attacks in Face Recognition: a public database and a baseline”, International Joint Conference on Biometrics, 2011.

[15] A. Pinto, H. Pedrini, W. R. Schwartz, A. Rocha, “Video-Based Face Spoofing Detection through Visual Rhythm Analysis”, SIBGRAPI, 2012.

[16] N. Kose, J. Dugelay “Classification of captured and recaptured images to detect photograph spoofing”, International Conference on Informatics, Electronics & Vision (ICIEV), 2012.

[17] T. Ojala, M. Pietikäinen, T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, 2002.

[18] B. S. Manjunath and W. Y. Ma, “Texture features for browsing and retrieval of image data”, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI - Special issue on Digital Libraries), vol. 18, no. 8, pp. 837-42, 1996.

[19] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features of Image Classification”, IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-3, no. 6, 1973.

[20] L. Soh and C. Tsatsoulis, “Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices”, IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 2, 1999.

[21] D. A. Clausi, “An analysis of co-occurrence texture statistics as a function of grey level quantization”, Can. J. Remote Sensing, vol. 28, no.1, pp. 45-62, 2002.

[22] C.-C. Chang and C.-J. Lin. LIBSVM: “A library for support vector machines”, ACM Transactions on Intelligent Systems and Technology, 2, 2011.

[23] De Jong, S. “SIMPLS: An Alternative Approach to Partial Least Squares Regression”, Chemometrics and Intelligent Laboratory Systems. Vol. 18, pp. 251–263, 1993.