FACE LIVENESS DETECTION BASED ON MULTIPLE FEATURE DESCRIPTORS

Jingjing Li[†], *Xinfeng Zhang*[‡], *Yongbing Zhang*[†], *Haoqian Wang*[†] *and Fang Yang*[†]

[†]Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China
[‡]School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, China

ABSTRACT

The face liveness detection module is one of the most important parts in the state-of-the-art face recognition system. In this paper, we present an efficient method to further improve its accuracy by leveraging multiple feature descriptors. Firstly, a data-driven feature descriptor is proposed based on the Karhunen-Loève Transform (KLT) learned from both client and imposter face images. Moreover, the Completed Local Binary Pattern (CLBP) algorithm is utilized to represent the local structure and the high-middle spectra components of 2D Fourier transform are also utilized to reflect the global structure. These features are fed into the support vector machine (SVM) to learn a classifier for face liveness detection. Experimental results on NUAA illustrate that our proposed method outperforms most of the widely utilized feature descriptors.

Index Terms— Face liveness detection, difference of Gaussian, 2D Fourier Spectra, CLBP, KLT

1. INTRODUCTION

With the application of face recognition in many areas such as face scan payment, identity verification, it is not a neglectable problem that someone may try to bypass a face biometric system using the photo of a valid user, which can be easily obtained through internet downloading or simply recapturing. As depicted in [1, 2], some face recognition systems can be easily fooled by recaptured 2D face images. The key point of this mater is that some face recognition systems have paid more attentions to the "image matching" module without analyzing whether the matched face is from a real face or a printed face image.

In recent years, face liveness detection has been one of the important tasks in face recognition systems, and relevant works have been proposed in literatures [3–8]. These methods can be divided into three categories according to the type of liveness indicator, *i.e.,* motion analysis, life sign detection and texture analysis, the pros and cons of which are discussed in [9]. The motion analysis methods mainly utilize the motion difference between 2D planar faces and 3D real faces. In [5], Kollreider *et al.* proposed a optical-flow based method to capture the subtle motion from multiple face images. In [7], Anjos *et al.* proposed a correlation analysis method based on optical flow, which achieves very good performance on popular datasets. The life sign detection methods mainly focus on the specific life signs, *e.g.* eye blinking, the opening of mouth and involuntary movements of parts of face and head. In [3, 4], the researchers focused on the eye-blinking motion and utilized it as a cue to detect the spoof attacks such as paper attack. However, these methods need the live video or multiple sequential images, which may not be applicable for some face recognition systems. Thus, many works have paid attention to the texture differences between the live faces and the spoof ones, which can be utilized for liveness detection from a single image. In [6], the Fourier analysis is used to capture the frequency distribution of live face images. Considering the inefficiency of single feature, in [8], Wild *et al.* fused the features from 2D Gabor filters, Gray level co-occurrence matrix (GLCM) and Fourier Transform (FT) to form hybrid features, which further improve the performance of face liveness detection.

Considering the various quality of input images, in [10], Galbally *et al.* proposed to utilize image quality assessment methods to enhance the security of biometric identification, and they put forward a face anti-spoofing assessment framework by leveraging 25 general image quality measures such as pixel difference, spectral distance, natural scene statistic, etc. In [11], the independent quantization of features and the joint quantization of rich local features are utilized to construct the new robust feature descriptors for face liveness detection. In [12], Wen *et al.* proposed a face spoof detection algorithm based on image distortion analysis (IDA), which extracted four different features (specular reflection, blurriness, chromatic moment, and color diversity) from normalized face images to form concatenated features, and then fed them into ensemble classifier. In [13], Boulkenafet *et al.* improve the robustness of the Local Binary Pattern (LBP) feature descriptors by extracting the joint color-texture information from face images. However, either the LBP or Fourier spectra cannot deal with various imposter images in face attacks, because the individual feature descriptor has limitation in representing image diverse structures.

In this paper, we propose a more efficient liveness face detection framework based on multiple features to deal with the diversity of imposter images. Herein, we first design a data-

978-1-7281-4666-9/19/$31.00 ©2019 IEEE

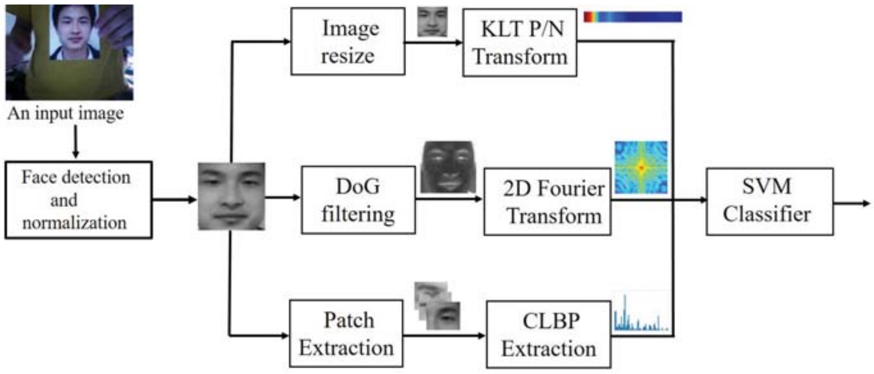

Fig. 1. Architecture of the proposed face anti-spoofing approach.

driven feature extractor based on the Karhunen-Loève Transform (KLT), where two transform matrices are learned from real and imposter face images. In addition, the 2D Fourier transform is utilized as a global feature extractor, which is applied on images after difference of Gaussian (DoG) filtering. This strategy makes the high-middle frequency spectra as global features to enhance their discrimination. Moreover, the Completed Local Binary Pattern (CLBP) is utilized to distinguish the local difference between real and imposter face images. These three kinds of features are jointly utilized and fed into support vector machine (SVM) to train a classifier for detecting the imposter images from live human face. Experimental results on the popular NUAA datasets further verifies the efficiency of the proposed method.

The remainder of this paper is organized as follows. The proposed multiple features based face liveness detection approach is introduced in Section 2. Section 3 shows the performance of the proposed method and analyzes the efficiency of the individual feature and their combination. . Finally, concluding remarks are introduced in Section 4.

2. THE PROPOSED MULTIPLE FEATURE BASED FACE LIVENESS DETECTION

Inspired by the importance of features in face liveness detection, we proposed a novel multiple feature method to improve the accuracy of 2D face liveness detection by dealing with the inefficient of individual feature. Fig.1 shows the framework of the proposed method, where there are three processing streams corresponding to each feature extraction. Before, feature extraction, the input image should be first normalized by converting into luminance space, detecting facial part, cropping and resizing the facial part into 64×64. The three kinds of features are further extracted from the normalized images and the features are cascaded into a vector and fed into the learned SVM classification model. The detailed feature extraction is introduced in the following subsections.

2.1. KLT Based Feature Extraction

The texture structure information is very important in distiguishing the real and fake face images. The previous works mainly follow the common characteristics for general images without considering the speciality of facial images. Different from general images, Facial images have its specific structures with eyes, nose and mouth in fixed positions, the structure of which can be well presented by the some learned atoms [14]. The KLT is a data-driven transform, the kernels of which are the unit eigenvectors of the covariance matrix of sampled data. It is the optimal transform in terms of energy compaction. In this paper, we propose to train two kinds of KLT matrices for real and imposter face images respectively, denoted as KLT P/N (Positive and Negative), as the KLT feature extractor. To make the feature compact, before KLT feature extraction, we first resize the normalized face image into 8×8 block and then apply the learned two KLT P/N matrices to transform it into feature domain. The joint positive and negative KLTs can better adapt to different facial image structures in both the real and imposter images.

2.2. 2D Fourier Spectra Feature

Based on the previous research in [6], the Fourier spectra feature has been verified its efficient as descriptors in distinguishing the texture difference between real and imposter face images. In this paper, we also take the 2D Fourier transform to extract the spectra information as a global feature. Considering the discrimination of different frequency bands, a difference of Gaussian filter is applied to the normalized face image before the 2D Fourier transform as follows,

$$I'_{\sigma_1,\sigma_2} = G_{\sigma_1} * I - G_{\sigma_2} * I \tag{1}$$

where the I is filtered by two Gaussian filters G_{σ_1} and G_{σ_2} with the standard deviations are σ_1 and σ_2 respectively. $*$ represents the convolution operation.

The utilization of DoG filter is based on the assumption that the most discriminative information for live and imposter

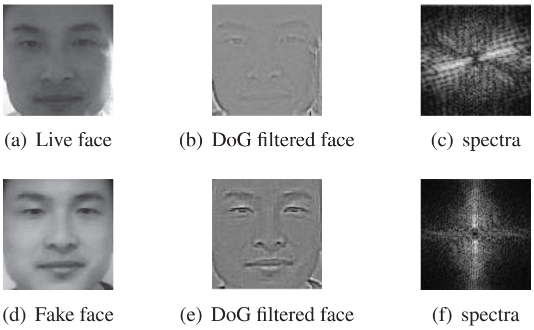

Fig. 2. Comparison of live face and fake face.

face images existing in a special frequency bands due to the different imaging conditions. The very high frequency bands maybe too noisy which may interfere with detection accuracy, and the very low frequency bands associated with DC component are not so discriminative. Therefore, the DoG filter can be regarded as a band-passing filter to output the most discriminative spectra by adjusting the two parameters σ_1 and σ_2. In this paper, we set the two parameters as 0.5 and 1 respectively, and Fig.2 shows the spectra distribution of the live and fake faces images after DoG filtering, where they show obviously different characteristics.

2.3. Completed Local Binary Pattern (CLBP)

Besides the two global features, KLT features and Fourier spectra features, we further proposed to utilize the CLBP features to distinguish the local difference between real and imposter face images in our work. The classical LBP features have been widely utilized in textural classification and recognition tasks [15, 16], which represent the local structures by comparing the values of center pixel with its neighbors as,

$$\mathrm{LBP}_{P,R} = \sum_{n=0}^{N-1} s(I_n - I_c)2^p, s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}, \quad (2)$$

where I_c and I_n are the gray value of central pixel and its neighbors respectively. Since the LBP feature only utilizes sign information of the comparison between central pixel and its neighbors, the magnitude information among them is totally removed, which degenerate efficiency of the local structural description.

In [17], Guo *et al.* proposed the CLBP feature descriptor, which utilizes the magnitude information and the value of central pixel besides sign information. Herein, the magnitude and central pixel information are denoted as CLBPM and CLBPC,

$$\mathrm{CLBPM} = \sum_{n=0}^{N-1} t(m_I, c)2^n, t(x,c) = \begin{cases} 1, & x \geq c \\ 0, & x < c \end{cases}, \quad (3)$$

$$\mathrm{CLBPC} = t(I_c, m_I), \quad (4)$$

where m_I is the mean value of the whole image. In our method, we first divide the normalized face images into patches and then extract CLBP features patch by patch. The final CLBP feature is the cascaded patch CLBP vectors, denoted as $P = [P^{(1)}...P^{(M)}]$, where $P^{(i)}, \{i = 1 : M\}$ is the CLBP feature extracted from the i^{th} patch.

2.4. SVM based Face Liveness Classification

To classify imposter face images from the live faces, the widely utilized support vector machine is utilized to train a classier with the cascaded three kinds of features. Herein, the LIBSVM [18] with linear kernel and default parameters is utilized in our work.

3. EXPERIMENTS

3.1. Dataset

To verify the efficiency of the proposed hybrid features, we conduct the experiments on the widely utilized 2D face liveness detection dataset, NUAA [19]. Herein, the NUAA is designed for printed photo attacks, and it was released publicly in 2010. There are 15 genuine subjects (80% of men and 20% of women) with the ages from twenty to thirty invited to attend this work.

Since the main goal of face liveness detection is to distinguish genuine faces from photographs instead of distinguishing different people as face recognition, there are fewer requirements for a large number of subjects than for the richness of the changes contained in the dataset. The fake faces were generated from image capturing of genuine subjects, each of which was asked to face the camera with a neutral expression, without obvious movements, such as blinking or head movement. That is to say, the genuine faces and fake faces look very similar, which makes the face liveness detection be challenging. Some statistics of NUAA are shown in Table 1. All original images in the dataset are color pictures captured by the webcam with the definition of 640×480 pixels.

We utilize the positive predictive value (PPV), negative predictive value (NPV) and accuracy to evaluate the performance of face liveness detection, which are calculated as follows,

$$PPV = \frac{TP}{TP + FP}, \quad (5)$$

$$NPV = \frac{TN}{TN + FN}, \quad (6)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (7)$$

where TP (True Positive) is the number of the live faces correctly identified as liveness, TN (True Negative) represents the number of the fake faces correctly identified as fake, FP (False Positive) is the number of the fake faces incorrectly identified as liveness, and FN (False Negative) is the number of the live faces incorrectly identified as fake.

Table 1. The NUAA images in the training and test set.

	Client	Imposter	Total	Resolution
Training Set	1743	1748	3491	640×480
Test Set	3362	5761	9123	640×480

3.2. Experimental Results

Table 2 shows the accuracy of the two individual features, CLBP and 2D Fourier spectra. We can see that the 2D Fourier spectra feature achieves much higher accuracy compared with that of CLBP feature, about 83.91%. When the features are combined as an intergraded feature, the classification accuracy is further improved as shown in Table 3, up to 94.28%. These results shows that the individual feature descriptor is inefficient to deal with images with various structures, and the combination of multiple features is an promising solution to improve the classification for face liveness detection. Moreover, in table 3, we further show the performance of the proposed method by combining the three features, and the classification accuracy is further improved by around 1%, which verifies the efficiency and the complementarity of the proposed three features in face liveness detection problem.

Table 2. Performance index (%) of upper stream and lower stream in the training and testing set.

	CLBP		Spectra	
	Training	Testing	Training	Testing
PPV	98.05	53.93	100	96.67
NPV	99.94	98.96	100	76.46
Accuracy	99.10	82.36	100	83.91

Table 3. Performance index (%) of fusion with and without PCA features in the training and testing set.

	w/o KLT features		fusion all features	
	Training	Testing	Training	Testing
PPV	100	98.67	100	98.96
NPV	100	91.60	100	93.02
Accuracy	100	94.28	100	95.21

To verify the efficiency of the proposed method, we further compare it with some state-of-the-art methods on face liveness detection features. In Fig.3, we show the ROC curves for the proposed method, individual LBP feature and the High Frequency Descriptor method in [20]. We can see that our method achieves obvious higher TP rate values at different FP rates compared with others. Moreover, in Fig.4, we further demonstrates the more performance comparisons for the proposed method and the low-rank based method in [19], where the low-rank method used various feature descriptors as inputs including LTVu [21], LTVp, fusion of LTVu and LTVp (LTVfused), DoG filtered image and one third of the highest

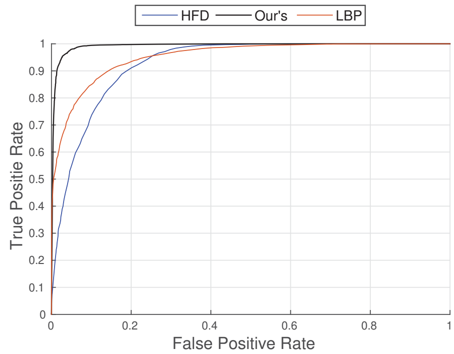

Fig. 3. The ROC curves of face liveness detection methods.

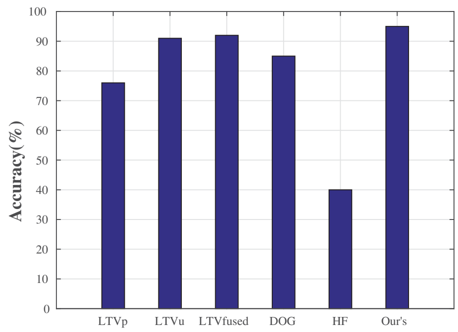

Fig. 4. Comparison of different features for face liveness detection on NUAA dataset.

frequency components (HF) [6]. We can see that the best detection accuracy of the low-rank based method is only around 90% which is obviously inferior to our method. Thanks to the combination of feature descriptors, the proposed method significantly improves the detection accuracy, and the optimal accuracy is up to 95.21%.

4. CONCLUSION

In this paper, we proposed a novel approach to improve the performance of the 2D face liveness detection. Based on the diversity analysis of different feature extraction approaches, we designed a three-stream fusion model as the feature representation of face images. The support vector machine is utilized in our method as the predictor to classify the genuine and fake faces. Based on the experimental results on NUAA dataset, the proposed method achieved very high accuracy and outperforms many other feature descriptors on the accuracy. As further work, we will extend the proposed stream model to the video attack problems by utilizing the temporal correlation.

5. ACKNOWLEDGEMENT

This work was supported in part by the Projects of National Science Foundations of China under Grant 61571254, in part by the Shenzhen Fundamental Research Fund under Grant JCYJ20170817161409809 and Grant JCYJ20180508152042002, and in part by the Guangdong Natural Science Foundation under Grant 2017A030313353.

6. REFERENCES

[1] Anil K. Jain, Patrick Flynn, and Arun A. Ross, *Handbook of Biometrics*, Springer US, 2008.

[2] Yan Li, Ke Xu, Qiang Yan, Yingjiu Li, and Robert H. Deng, "Understanding osn-based facial disclosure against face authentication systems," in *ACM Symposium on Information, Computer and Communications Security*, 2014, pp. 413–424.

[3] Gang Pan, Lin Sun, Zhaohui Wu, and Shihong Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *IEEE International Conference on Computer Vision*, 2007, pp. 1–8.

[4] Lin Sun, Gang Pan, Zhaohui Wu, and Shihong Lao, "Blinking-based live face detection using conditional random fields," in *International Conference on Advances in Biometrics*, 2007, pp. 252–260.

[5] 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.

[6] Jiangwei Li, Yunhong Wang, Tieniu Tan, and Anil K. Jain, "Live face detection based on the analysis of fourier spectra," *Proc Spie*, vol. 5404, pp. 296–303, 2004.

[7] André Anjos, Murali Mohan Chakka, and Sébastien Marcel, "Motion-based counter-measures to photo attacks in face recognition," *Iet Biometrics*, vol. 3, no. 3, pp. 147–158, 2014.

[8] Peter Wild, Petru Radu, Lulu Chen, and James Ferryman, "Robust multimodal face and fingerprint fusion in the presence of spoofing attacks," *Pattern Recognition*, vol. 50, no. C, pp. 17–25, 2016.

[9] O. Kahm and N. Damer, "2d face liveness detection: An overview," in *Biometrics Special Interest Group*, 2012, pp. 1–12.

[10] J Galbally, S Marcel, and J Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE Trans Image Process*, vol. 23, no. 2, pp. 710–724, 2014.

[11] Diego Gragnaniello, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva, "An investigation of local descriptors for biometric spoofing detection," *IEEE Transactions on Information Forensics Security*, vol. 10, no. 4, pp. 849–863, 2015.

[12] Di Wen, Hu Han, and Anil K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics Security*, vol. 10, no. 4, pp. 746–761, 2015.

[13] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid, "face anti-spoofing based on color texture analysis," in *IEEE International Conference on Image Processing*, 2015, pp. 2636–2640.

[14] Amir Masoud Taheri and Homayoun Mahdavi-Nasab, "Facial image compression using adaptive multiple dictionaries," in *9th Iranian Conference on Machine Vision and Image Processing*, 2015, pp. 92–95.

[15] Timo Ojala, Matti Pietikainen, and Topi Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[16] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen, "Face recognition with local binary patterns," in *European conference on computer vision*. Springer, 2004, pp. 469–481.

[17] Guo Zhenhua, Zhang Lei, and Zhang David, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010.

[18] Chih Chung Chang and Chih Jen Lin, "Libsvm: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[19] Xiaoyang Tan, Yi Li, Jun Liu, and Lin Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *European Conference on Computer Vision*. Springer, 2010, pp. 504–517.

[20] Junyan Peng and Patrick P. K. Chan, "Face liveness detection for combating the spoofing attack in face recognition," in *International Conference on Wavelet Analysis & Pattern Recognition*, 2014.

[21] Terrence Chen, Wotao Yin, Xiang Sean Zhou, Dorin Comaniciu, and Thomas S Huang, "Total variation models for variable lighting face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 9, pp. 1519–1524, 2006.