Gaze-based Presentation Attack Detection for Users Wearing Tinted Glasses

Asad Ali, Nawal Alsufyani, Sanaul Hoque* and Farzin Deravi
School of Engineering and Digital Arts
University of Kent
Canterbury, United Kingdom
s.hoque@kent.ac.uk

***Abstract*—Biometric authentication is vulnerable to presentation (spoofing) attacks. It is important to address the security vulnerability of spoofing attacks where an attacker uses an artefact presented at the sensor to subvert the system. Gaze-tracking has been proposed for such attack detection. In this paper, we explore the sensitivity of a gaze-based approach to spoofing detection in the presence of eye-glasses that may impact detection performance. In particular, we investigate the use of partially tinted glasses such as may be used in hazardous environments or outdoors in mobile application scenarios The attack scenarios considered in this work include the use of projected photos, 2D and 3D masks. A gaze-based spoofing detection system has been extensively evaluated using data captured from volunteers performing genuine attempts (with and without wearing such tinted glasses) as well as spoofing attempts using various artefacts. The results of the evaluations indicate that the presence of tinted glasses has a small impact on the accuracy of attack detection, thereby making the use of such gaze-based features possible for a wider range of applications.**

Keywords— biometrics, spoofing, liveness; mobile security, gaze tracking, challenge-response technique.

I. INTRODUCTION

Biometric systems have the potential to provide solutions for a variety of real time security applications. However, presentation attacks at the sensor is still a serious challenge to their use especially in un-supervised applications. This, however, can be addressed through a "Liveness" detection mechanism, which can be added to the existing biometric systems. Tracking users' gaze while responding to a visual stimulus has already been reported as a potential solution to this problem and a number of novel features that rely on such approach has been proposed [9-11,15-17]. In all these cases, data were acquired under ideal conditions where the users' eyes are clearly visible to facilitate gaze tracking easily. However, there are special cases where a user may be wearing tinted glasses especially in outdoor scenarios or may have to wear protective tinted glasses in hazardous work environments making gaze tracking for attack detection a more challenging task. In this paper, we explore the sensitivity of a gaze-based approach to spoofing detection in the presence of partially tinted eye-glasses and assess any impact on its attack detection accuracy. The system has been extensively evaluated using data captured from volunteers performing genuine attempts (with and without wearing such tinted glasses) as well as spoofing attempts using various artefacts.

The paper is structured as follows: Section II presents the state of the art of presentation attack detection (PAD) systems for facial biometrics. Section III presents the experimental setup and corresponding results are shown in Section IV. Conclusions are then provided in Section V.

II. RELATED WORKS

In the literature various approaches have been presented to establish liveness to detect presentation attacks. These approaches can be grouped into active and passive categories. Active approaches need user engagement with the biometric system to establish the liveness of the source through the sample captured at the sensor. Passive approaches do not require user co-operation or even user awareness but exploit involuntary physical movements. These can be spontaneous eye blinks and 3D properties of the image source.

A. Passive Techniques

Blinking is a natural phenomenon of the closing and opening of the eyelid. The blink helps spread fluid from the tear ducts across the eye and removes irritants from the surface of the cornea and conjunctiva [1]. Blinking has been used as a means of human interaction with computer [4, 5]. Eye blink has also been used for face liveness detection in the literature for biometric systems. Eye blink can be detected by classifying each image in a video sequence independently as one state (closed eye or open eye). A blink can then be defined by a procedure of eyes going from open to closed, and back to open. The blink speed can be affected by fatigue, injury/disease, medication, etc. Lin Sun et al. [2] presented an eye blink detection approach for detecting face liveness using Conditional Random Fields (CRFs) which has been further enhanced in [3]. In this method, they extracted the temporal information from the process of the eye-blink, namely the consecutive stages of open, half closed and closed, followed by half open and fully open all of which are sequential eye blink movements and constitute a complete eye blink pattern which was used to determine liveness.

When a photograph is printed on a paper, it introduces texture which is not present in the original images when captured from genuine user's faces. Schwartz et al. [4] proposed a texture based counter spoofing method for photo attack detection. They explored face texture, colour and shape to obtain a holistic representation. They generated feature vectors formed by combining low-level feature descriptors for each frame of a video that contains the facial information. Pinto et al. [5] presented a technique for video replay attack detection by analysing the noise signatures which were generated during video acquisition process. These noise properties, extracted from the captured video, were analysed using Fourier transform for spoofing attack detection. A compact representation, called visual rhythm, was proposed to detect the temporal information in the Fourier spectrum. Maatta et al. [6] proposed an approach based on reflection. According to the authors, genuine and fake facial images have differences in reflection. They used LBP-based micro texture analysis. Normalised facial images were divided into several local regions and three descriptors were extracted from each

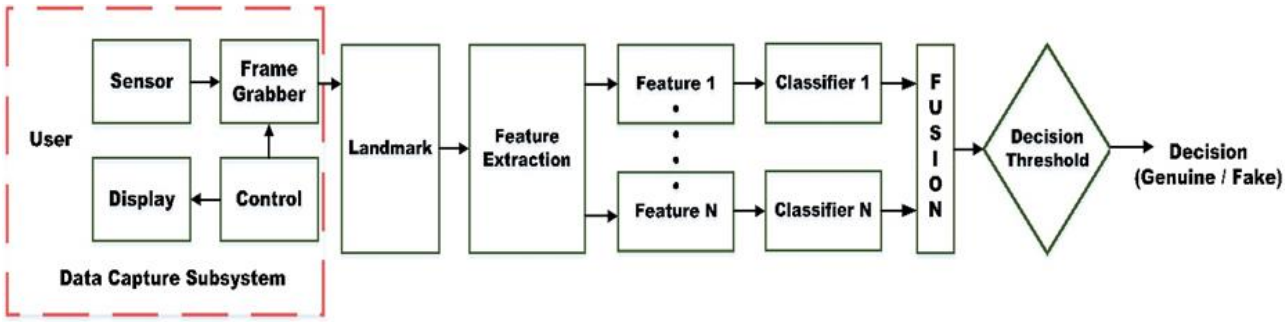

Fig. 1. The PAD system block diagram

block. The LBP operator was applied on the normalized face image.

Except 3D mask, all the commonly used artefacts are 2D in nature. The real face is 3D which carries depth information. Lagorio et al. [7] proposed a novel liveness detection method based on 3D facial structure. The proposed approach could be deployed in 2D or 3D face recognition system to detect spoofing attacks. The proposed algorithm computed the 3D features of the captured face data.

B. Active Techniques

Systems based on the challenge-response approach belong to the active category. In this category, the user is asked to perform specific activities to ascertain liveness such as uttering digits or changing his or her head pose. This approach may be useful against photo and video spoofing attacks. However, it could be challenging to stop mask attack using this method. Frischholz et al. [8] explored a challenge-response approach where users were required to look in certain directions randomly. The system estimated the head pose and compared it to the instructions given by the system. Ali et al. [9-11, 15] reported explicit gaze-based liveness detection approach for the first time. A moving object was shown randomly on the screen for user to follow with eye and head movement. The visual stimulus directed the gaze of the user to specific points on the screen. Features extracted from images where users were looking at collinear, colocated locations on the screen and corresponding gaze homography were used to estimate the liveness of the source. Experiments showed that the methods were effective in counter spoofing for all three types of attack (photo, 2D mask and video replay). Singh et al. [12] proposed a liveness detection scheme based on eye and mouth movements. The challenges were generated randomly such as eye or mouth movement (openness/ closeness). Responses were estimated using the corresponding eye and mouth movements. Smith et al. [13] proposed an approach to counter replay attacks on smart devices using a challenge-response technique. The bright white colour is considered as a challenge, and the reflection from the person's face due to this white colour is the response. They also used different colours as challenge and the corresponding reflections from the face due to these various colours were analysed to determine the presentation attack. Cai et al. [14] proposed system based on gaze estimation. The challenge creates points on the computer screen and the user is required to look at these points. Gaze estimation model was trained for each subject to predict the gaze positions when the user looking at computer screen. The difference between the predicted and the screen points are then used to differentiate between attacks and genuine attempts.

In the very early phase of face-PAD research [19, 20], spoof attack detection approaches were evaluated on proprietary datasets [21]. The use of private databases can be seen as somewhat reasonable when (random) challenge-response based methods demands specific user interaction [21]. Several publicly available datasets [21–28] are now available to test, evaluate, and compare face-PAD methods. However, the attack scenarios addressed in these databases are mostly photo and video attacks. This paper explores the impact of tinted glasses on gaze based spoof detection for which a database has been collected locally to simulate such attack as none of the public databases included tinted glasses.

III. The Gaze-based Presentation Attack Detection System

Fig. 1 shows the block diagram of the proposed system where a visual stimulus (as part of the challenge) appears on the display which the participant is asked to follow and the camera (sensor) captures facial images at each position of the stimulus on the screen. A control mechanism is used to ensure the placement of the target and the image acquisition are synchronized. The system extracts facial landmarks in the captured frames, computes various features from these landmarks, which are then used to classify whether the presentation attempt is by a genuine user.

A. Visual Stimulus and User Response Acquisition

The restricted geometry of a mobile device display (Tablet device, 15.87×21.18 cm) is simulated in the experiments using

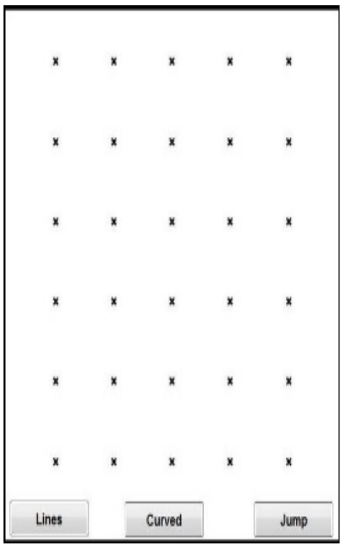

Fig.2. Sample of Points Challenge grid.

a limited area of a desktop computer screen. A small shape ("x") is presented, at distinct locations on the screen as shown in Fig. 2. The cross sign appears in a grid of 30 distinct pre-defined locations (Points Challenge) (Fig. 2). The order of points is randomised for each challenge attempt. During each presentation, images were acquired at every location of the challenge. The presentation of the challenge sequence lasted approximately 1 minute, however, only a small portion of each session was used for spoofing detection.

Data was collected from 80 participants. This number of participants is sufficient to illustrate the potential of the proposed approach and is in line with current state-of-the-art. Participants were of both male and female gender aged over 18 years old. The volunteers were from Africa, Asia, Middle-East, and Europe. Three spoofing attack types (projected photo, 2D mask and 3D mask), one genuine attempt without wearing tinted glasses and one genuine attempt wearing tinted glasses were recorded for each participant. Fig. 3(a) shows genuine attempt, Fig. 3(b) shows projected photo attempt, Figs. 3(c) and 3(d) show 2D mask and 3D mask attacks respectively. A participant wearing tinted glasses is shown in Fig. 4.

B. Facial Landmark Detection and Feature Extraction

The images thus captured during the challenge-response operation were processed using Chehra Version 3.0 [18] in order to extract facial landmark points. Chehra returns 59 different landmarks on the face region. The coordinates of some of these landmarks were used for feature extraction in the proposed scheme. Features proposed here are based on the eye movement during the challenge.

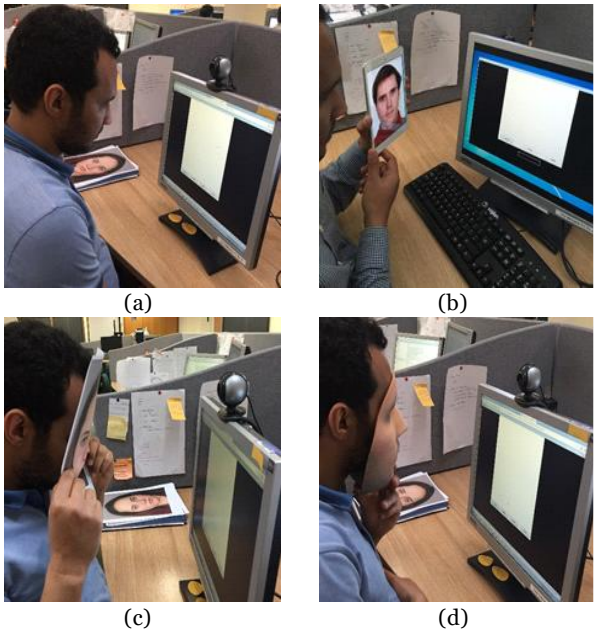

Fig. 3. Data collection process (a) genuine attempt without glass, (b) photo attack, (c) 2D mask attack, (d) 3D mask attack

Fig. 4. A participant wearing tinted glasses

C. Gaze based colocation feature

For the colocation feature, the Points stimulus is used letting the user to fixate on a number of randomly selected locations on the screen. At each stimulus location, the facial image of the user is captured. The gaze colocation features are extracted from images where the stimulus is at the same locations at different times. It can be assumed that the coordinates of the pupil centres in these corresponding frames should also be very similar. This should result in a very small variance in the observed coordinates of the pupil centres in genuine attempts. On the contrary, for presentation attacks, the gaze fixations are expected to be more variable resulting in much higher variances. A feature vector is thus formed from the variances of pupil centre coordinates for all the frames where the stimulus is colocated. The features are then passed to the classifier to discriminate between genuine and fake attempts.

Suppose u_i and v_i are the observed coordinates of a given landmark in response to the stimulus presented at the same location at different instances. To quantify the deviation from perfect colocation, the variances in the observed landmarks are calculated. Let σ_u^2 and σ_v^2 denote the variances of the observed landmarks.

$$\sigma_u^2 = \frac{1}{N}\sum_i (u_i - \bar{u})^2 \quad (1)$$

$$\sigma_v^2 = \frac{1}{N}\sum_i (v_i - \bar{v})^2 \quad (2)$$

where $\bar{u}$ and $\bar{v}$ is the mean of the observed landmark locations and N is cardinality of the corresponding subset of responses. These variances are concatenated together for feature vector as shown below:

$$F_{coloc} = [\sigma_u^2,\ \sigma_v^2, \dots] \quad (3)$$

IV. Experiments

Several sets of experiments were carried out to verify the performance of the proposed features in distinguishing genuine attempts from attacks. Initially, the effectiveness of the proposed colocation features was investigated with genuine presentation without wearing glasses. Subsequently experiments were carried out to investigate whether this feature would work effectively even if tinted glasses are worn by the genuine users.

Table I shows the performances for three attack scenarios (photo, 2D mask and 3D mask). Each set of colocated points takes around 3 seconds to present. It is clear from the table that this feature is effective in distinguishing photo attacks from

TABLE I. TPR AT FPR = 0.10 FOR VARIOUS SETS OF COLOCATION POINTS, TRAINED AND TEST WITH GENUINE PRESENTATION WITHOUT WEARING GLASSES

Attack Type	Sets of colocated points		
	3	***10***	***15***
Photo	92%	92%	93%
2D Mask	44%	42%	43%
3D Mask	40%	48%	57%

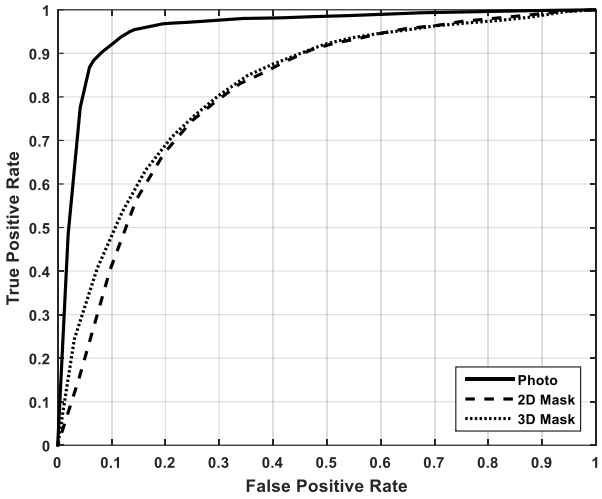

Fig. 5. ROC curves for photo, 2D mask and 3D mask for 10 sets of colocation points, trained and tested with genuine presentation without wearing glasses

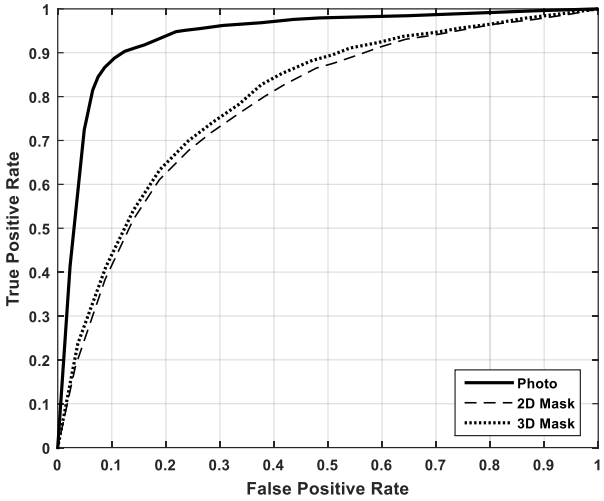

Fig. 6. ROC curves for photo, 2D and 3D mask for 10 sets of colocation points, trained with genuine attempts without wearing tinted glasses but tested with genuine presentations wearing tinted glasses

genuine presentations. It is apparent that the performance of the system generally improves with the number of colocated sets of points used as features. At 10% FPR, the TPR is 93% for photo attack using 15 sets of colocation points. This is changed only slightly when even as few as three sets of colocation points are used. The detection accuracy of the system was found to be much lower for 2D mask and 3D mask attacks. At 10% FPR, the performance is 43% and 57% TPRs for 2D mask and 3D mask using 15 sets of colocation points. When fewer sets of colocation points are used the performance under 3D mask attacks is dropped more than that for 2D mask attacks which remains more or less stable. The performance of the system for 3D mask attack detection is slightly better than that achieved for 2D mask attacks detection. At 10% FPR, 3D mask detection TPR is about 48% for 10 set of colocation points.

Fig. 5 presents the ROC curves for attack detection. The performance of the system for 3D mask attack detection is slightly better than that achieved for 2D mask attack detection. These results may be explained due to the ease with which an attacker may be able to look through the holes cut in the 2D and 3D masks. Given the flexibility of the 2D masks used compared to the rigid 3D masks, it may be easier for the attacker to subvert the system using the 2D mask, hence the slightly worse detection results for this attack type.

A similar set of tests was conducted with the system being trained with genuine presentations without wearing tinted glasses, but tested with users wearing tinted glasses. The purpose of this set of experiments was to investigate the performance of the system when genuine users attempted to access the system with tinted glasses.

TABLE II. TPR AT FPR = 0.10 FOR VARIOUS SETS OF COLOCATION POINTS, TRAINED WITH GENUINE PRESENTATION WITHOUT WEARING TINTED GLASSES AND TESTED WITH GENUINE PRESENTATION WEARING TINTED GLASSES

Attack Type	Sets of colocated points		
	3	***10***	***15***
Photo	88%	88%	90%
2D Mask	40%	42%	43%
3D Mask	39%	44%	51%

Table II summarizes the TPRs at FPR = 10% for various sets of colocation points for these experiments. The performance of the system remains largely unchanged as the number of collocated point sets is reduced. In case of photo attack detection, at 10% FPR, the TPR is 90% using 15 sets of colocation points. The performance of the system was found to be worse when the colocation feature is used for 2D mask attack detection. At 10% FPR, the performance is about 43% TPR for 15 sets of colocation points. The performance of the system for 3D mask attack detection is slightly better than that achieved for 2D mask attacks detection. Fig. 6 shows the corresponding ROC curves for the three attack scenarios for 10 sets of colocation points.

Tables I and II provide comparison of TPRs for a range of FPRs for three attack scenarios with and without wearing tinted glasses. It is clearly seen that while wearing tinted glasses has generally lead to a reduction in accuracy, the overall performance level for photo attack detection is still significant.

V. Conclusion

This paper explored the impact of wearing tinted glasses on the detection accuracy of a gaze-based presentation attack detection technique. It is seen that wearing such glasses does have an impact on performance of such spoofing attack detection techniques. However, the impact is not very large for attack types that are amenable to this form of detection and the performance may still be acceptable in some application. This result may be of particular significance in applications when spoofing detection may need to take place in outdoor environments using mobile devices or in indoor hazardous environments that may demand the use of eyewear for protection. Future work will explore the impact of tinted glasses when using other challenge-response spoofing detection techniques as well as the possibility for the use of reflective glasses in gaze-based liveness detection.

References

[1] C. Roberts, "Biometric attack vectors and defences," *Computers & Security*, vol. 26(1), pp. 14-25, February 2007.

[2] L. Sun, G. Pan, Z. Wu, and S. Lao, “Blinking-based live face detection using conditional random fields,” in *Advances in Biometrics (ICB 2007)*. Seoul, Korea, Lecture Notes in Computer Science, vol. 4642. Springer, pp. 252-260, August 2007.

[3] G. Pan, L. Sun, Z. Wu, and S. Lao, “Eyeblink-based anti-spoofing in face recognition from a generic webcamera,” in *Proc. of the IEEE 11th International Conf on Computer Vision (ICCV)*, Rio de Janeiro, Brazil, pp. 1-8, October 2007.

[4] W. R. Schwartz, A. Rocha, and H. Pedrini, “Face spoofing detection through partial least squares and low-level descriptors,” in *Proc. of the International Joint Conference on Biometrics (IJCB)*, Washington, DC, USA, pp. 1-8, October 2011.

[5] A. Pinto, W. R. Schwartz, H. Pedrini, and A. d. R. Rocha, “Using visual rhythms for detecting video-based facial spoof attacks,” *IEEE Transactions on Information Forensics and Security*, vol. 10(5), pp. 1025-1038, May 2015.

[6] J. Maatta, A. Hadid, and M. Pietikainen, “Face spoofing detection from single images using texture and local shape analysis,” *IET Biometrics*, vol. 1(1), pp. 3-10, March 2012.

[7] A. Lagorio, M. Tistarelli, M. Cadoni, C. Fookes, and S. Sridharan, “Liveness detection based on 3d face shape analysis,” in *Proc. of International Workshop on Biometrics and Forensics (IWBF)*, Lisbon, Portugal, pp. 1-4, April 2013.

[8] R. W. Frischholz and A. Werner, “Avoiding replay-attacks in a face recognition system using head-pose estimation,” in *Proc. of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG '03)*, pp. 234-235, October 2003.

[9] A. Ali, F. Deravi, and S. Hoque, “Liveness detection using gaze collinearity," in *Proc. of the 2012 Third International Conference on Emerging Security Technologies (EST)*, Lisbon, Protugal, pp. 62-65, September 2012.

[10] A. Ali, F. Deravi, and S. Hoque, “Spoofing attempt detection using gaze colocation,” in *Proc. of the International Conference of the Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, pp. 1-12, September 2013.

[11] A. Ali, F. Deravi, and S. Hoque, “Directional sensitivity of gaze-collinearity features in liveness detection,” in *Proc. of the 4th International Conf on Emerging Security Technologies (EST)*, Cambridge, UK, pp. 8-11, September 2013.

[12] A. K. Singh, P. Joshi, and G. C. Nandi, “Face recognition with liveness detection using eye and mouth movement,” in *2014 International Conference on Signal Propagation and Computer Technology (ICSPCT)*, Ajmer, India, pp. 592-597, July 2014.

[13] D. F. Smith, A. Wiliem, and B. C. Lovell, “Face recognition on consumer devices: Reflections on replay attacks,” *IEEE Transactions on Information Forensics and Security*, vol. 10(4), pp. 736-745, April 2015.

[14] L. Cai, L. Huang, and C. Liu, “Person-specific face spoofing detection for replay attack based on gaze estimation,” in *Biometric Recognition (CCBR 2015)*, Tianjin, China, Lecture Notes in Computer Science, vol. 9428. Springer. pp. 201-211, November 2015.

[15] Ali, Asad, Sanaul Hoque, and Farzin Deravi. “Gaze stability for liveness detection,” *Pattern Analysis and Applications*, vol. 21(2), Springer, pp. 437-449, May 2018.

[16] N. Alsufyani, A. Ali, S. Hoque, and F. Deravi, "Biometric presentation attack detection using gaze alignment,” in *IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA)*, Singapore, Singapore, pp. 1-8, January 2018.

[17] A. Ali, N. Alsufyani, S. Hoque, and F. Deravi, "Biometric counter-spoofing for mobile devices using gaze information," in *Pattern Recognition and Machine Intelligence (PReMI)*, Kolkata, India, vol. LNCS-10597, Springer, pp. 11-18, December 2017.

[18] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Incremental face alignment in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, Ohio, USA, pp. 1859-1866, June 2014.

[19] J. Li, Y. Wang, T. Tan, and A. K. Jain, “Live face detection based on the analysis of fourier spectra,” in *Biometric Technology for Human Identification*, vol. 5404, Orlando, Florida, USA, pp. 296-304, April 2004.

[20] W. Bao, H. Li, N. Li, and W. Jiang, “A liveness detection method for face recognition based on optical flow field,” in *International Conference on Image Analysis and Signal Processing*, Taizhou, China, pp. 233-236, April 2009.

[21] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, “OULU-NPU: A mobile face presentation attack database with real-world variations,” in *Proc. of the 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, Washington, DC, USA, pp. 612-618, June 2017.

[22] X. Tan, Y. Li, J. Liu, and L. Jiang, “Face liveness detection from a single image with sparse low rank bilinear discriminative model,” *Computer Vision–ECCV 2010*, Heraklion, Crete, Greece, vol. LNCS 6316, Springer, pp. 504–517, September 2010.

[23] A. Anjos and S. Marcel, “Counter-measures to photo attacks in face recognition: a public database and a baseline,” in *Proc. of the International Joint Conference on Biometrics (IJCB)*, Washington, DC, USA, pp. 1–7, October 2011.

[24] K. Kollreider, H. Fronthaler, and J. Bigun, “Verifying liveness by multiple experts in face biometrics,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Anchorage, AK, USA, pp. 1–6, June 2008.

[25] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, “A face antispoofing database with diverse attacks,” in *Proc. of the 5th IAPR International Conference on Biometrics (ICB)*, New Delhi, India, pp. 26–31, March-April 2012.

[26] I. Chingovska, A. Anjos, and S. Marcel, “On the effectiveness of local binary patterns in face anti-spoofing,” in *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, pp. 1–7, September 2012.

[27] D. Wen, H. Han, and A. K. Jain, “Face spoof detection with image distortion analysis,” *IEEE Transactions Information Forensics and Security*, vol. 10(4), pp. 746–761, April 2015.

[28] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel, “The REPLAY-MOBILE face presentation-attack database,” in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, pp. 1-7, September 2016.