Person-specific Face Spoofing Detection for Replay Attack based on Gaze Estimation

Lijun Cai, Lei Huang, and Changping Liu

Institute of Automation Chinese Academy of Sciences, Beijing, China

Abstract. Based on gaze estimation, we propose an effective person-specific spoofing detection method to counter replay attack using a non-invasive challenge and response technique. The points on the computer screen create the challenge, and the gaze positions of the user as they look at the computer screen form the response. Firstly, face identification is conducted to recognize identity. Secondly, gaze estimation model is trained for each subject by adaptive linear regression with incremental learning and used to predict gaze positions when user is looking at the computer screen. Finally, difference between predicted gaze positions and system point locations is used as fake score to evaluate the liveness of user. Our basic assumption is that a genuine access can be attacked by salient objects and follow them. Therefore, the lower the fake score is, the more probable the user is genuine. Experimental results show that proposed method obtains competitive performance in distinguishing replay attacks from genuine accesses.

Keywords: Face spoofing detection, replay attack, incremental learning, gaze estimation

1 Introduction

Due to the requirement of information security, face spoofing detection is attracting more and more attention and research nowadays. Generally speaking, there are three common manners to spoof face recognition system: print photograph, replayed video and 3D model of a valid user. Compared with real faces, print photograph faces are planar, as well as having quality degradation and blurring problems. Replayed video faces are reflective and 3D face models are rigid. Based on these clues, face anti-spoofing techniques can be roughly classified into three categories: motion-based [1–5], texture-based [6–9] and fusion methods combing motion and texture [10, 11]. Almost all these methods are effective for simple spoofing attacks for example, print photograph. However, very little attention has been paid to replay attacks. Existing methods dealing with replay attacks either by combing other biometric mode such as voice, gesture with face information [12–14] or by multiple spectrum device [15, 16] or operating in controlled environment such as a darkened room [17].

Considering that gaze is a kind of behavioral biometrics which is difficult to be detected by the surveillance due to the ambiguity of visual attention process,

it can be used as a clue for anti-spoofing with the following characteristics [18]. Firstly, it does not require physical contact between user and device. Secondly, gaze is difficult to be obtained by surveillance camera and other equipment. Ali et al. [19–21] present the first time to use gaze clue for anti-spoofing, in which user is required to follow a moving point showed on the computer screen. Features based on the collinearity of gaze are used to discriminate between genuine access and print photographs attack. However, they are invalid for still photographs and uncooperative users. We previously provided the first investigation in research literature on the use of gaze estimation model for face spoofing detection in [22], in which nonlinear regression model is trained including multiple subjects and used for gaze estimation, then information entropy on predicted gaze positions under the stimulus of random points suggests the uncertainty level of user's gaze movement. The higher the information entropy is, the more probable the user is genuine. Experimental results show the effectiveness of this method on photographs and replay attacks with still gaze (user in the video almost only watch one direction). However, it does not work to replay attacks with moving gaze. That is to say, this method misjudges relay attack in which video user changes his gaze directions frequently as genuine access.

In this paper, based on gaze estimation, we propose an improved version of [22] to counter replay attack using a noninvasive challenge and response technique. The points on the computer screen create the challenge, and the gaze positions of the user as they look at the computer screen form the response. Face spoofing detection is performed by evaluation the difference of the gaze positions and system point locations. Compared with [22], the different points and improvements in this proposed method are as follows. 1) Proposed method is person-specific. In this paper, gaze estimation model is trained for each subject, which dismisses the interferences among different subjects. 2) Compared with nonlinear regression, adaptive linear regression is adopted to estimate gaze positions for reducing computation complexity. To meanwhile obtain lower gaze error, incremental learning is integrated into adaptive linear regression for dynamically increasing the calibration-free training PoG (point of gaze). 3) Considering that moving point locations too random makes it impossible for the eyes to follow it, system points in this paper are generated following some distribution with random parameters, for example, Gaussian distribution with random mean and variance.

2 Proposed Face Spoofing Detection Method

The general framework of proposed method is illustrated in Fig. 1, which consists of three main steps: face identification, gaze estimation and liveness judgement.

For a test sample, his identity should be obtained firstly. Note that face identification is not the focus of this paper, any effective face identification methods, for example [23, 24], can be used here to identify the test sample. After obtaining the identity of the input face images, person-specific gaze estimation model is trained and used to predict the gaze positions of input images. Finally, Eu-

clidean distance between predicted gaze positions and system points locations is computed as the fake score of the test sample. According to the visual attention mechanism, people are always attracted by some certain regions or objects. Therefore, the lower the fake score is, the more possible the user is judged as a genuine access. Next we will detail the gaze estimation and liveness judgement.

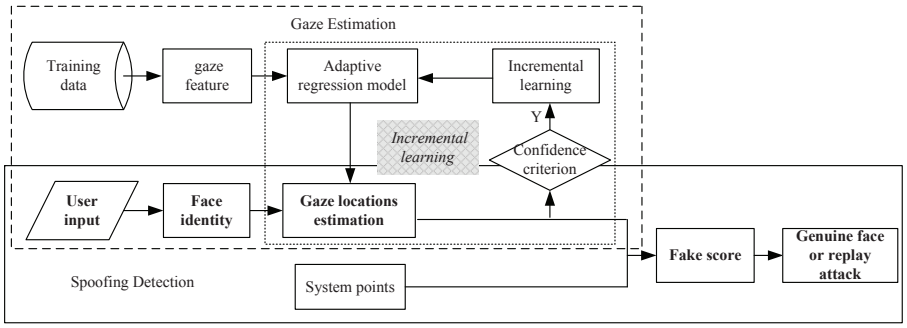


Fig. 1. System architecture.

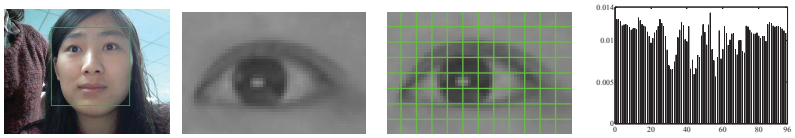
2.1 Gaze Estimation

Existing gaze estimation methods can be roughly classified into two categories: feature-based methods and appearance-based methods. Feature-based methods [25, 26] map the gaze feature (for example iris outline, pupil, cornea) to gaze position. However, this kind of methods generally require high quality camera, even multiple light sources. Appearance-based [27, 28] methods directly map the whole eye region to gaze position, which takes full advantage of gaze information. Considering proposed method is conducted under the condition of nature light and a generic camera, we choose an effective appearance-based method, adaptive linear regression [28], to establish gaze estimation model. Generally speaking, the gaze error will become lower with the increasing of training PoG (Point of Gaze) number which, however, brings more users' calibration burden. To get lower gaze error meanwhile not bring additional burden on user, incremental learning is added to adaptive linear learning in this work for online dynamically increasing the number of calibration-free training PoG.

Gaze Feature Extraction. Gaze feature extraction consists of two steps: eye region crop and feature generation. In the first step, face region and inner and outer eye corners are detected by adaptive boosting algorithm [29] (Fig. 2(a), left eye is used in this paper). To deal with small head motion, an additional alignment procedure is performed. Firstly we define an eye image template with 60×40 pixels, and the location of inner eye corner is set at (55, 25) and outer corner (5, 25). The aligned eye region is obtained by rotating and scaling the face region based on the locations of eye corners in template (Fig. 2(b)).

In the feature generation step, similar to [28], the cropped eye region is further divided into $r \times c$ subregions (here 8×12, Fig. 2(c)). Let S_j denote

the sum of pixel intensities in j-th subregion, then gaze feature is generated by $e = \frac{[S_1, S_2, \cdots, S_{r\times c}]^{\mathrm{T}}}{\sum_j S_j}$ (Fig. 2(d)).



(a) (b) (c) (d)

Fig. 2. Gaze feature extraction. (a) Face and eye corners detection. (b) Cropped eye region (60×40 pixels). (c) Uniform partition of eye region. (d) Gaze feature (96D)

Adaptive Linear Regression with Incremental Learning Adaptive linear regression aims to find a subset of training data for reconstructing the test data. Compared with linear regression, adaptive linear regression can neglect irrelevant training data, thus is helpful to predict. Based on adaptive linear regression, incremental learning is combined for lower gaze error. The mathematical definition of adaptive linear regression with incremental learning is described as follows.

Let matrices $F = [f_1^d, \cdots, f_n^d] \in \mathrm{R}^{m\times n}$ and $P = [p_1^d, \cdots, p_n^d] \in \mathrm{R}^{2\times n}$ include all the gaze features and gaze positions of training samples belonging to the person with identity d (d is obtained by face identification), where m is the feature dimension and n is the samples number. For a test frame I_t with identity d and gaze feature $\hat{f}$, the corresponding gaze position can be estimated as $\hat{p} = P\hat{w}$ by adaptive linear regression

$$\hat{w} = \arg\min_w ||w||_1 \quad s.t. \quad ||Fw - \hat{f}||_2 < \epsilon, \sum_i w_i = 1 \tag{1}$$

Assuming $\{Q_0, \cdots, Q_N\}$ are the system points and Q_{j_0} is appearing on the computer screen when I_t is captured by system camera. If confidence criterion $||\hat{(p)} - Q_{j_0}||_2 < \varepsilon$ (ε is a small positive number) is met, F and P can be extended to $\tilde{F} = [F\, \hat{f}]$ and $\tilde{P} = [P\, \hat{p}]$. Therefore, for the next captured image I_{t+1} with gaze feature f^*, its gaze position can be estimated as $p^* = \tilde{P}\hat{w}$ by solving

$$\hat{w} = \arg\min_w ||w||_1 \quad s.t. \quad ||\tilde{F}w - f^*||_2 < \epsilon, \sum_i w_i = 1 \tag{2}$$

Based on above description, system points satisfying confidence criterion in the test phase can be added into the training PoG set one by one without calibration, which is the main idea of adding incremental learning to adaptive linear regression.

2.2 Liveness Judgement

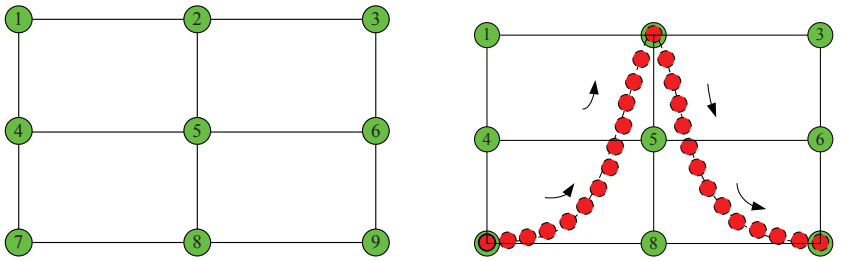
In this paper, system points $\{Q_1, \cdots, Q_N\}$ are generated by Gaussian distribution with random mean μ and standard deviation σ. Given another random positive number a, Q_i can be represented as $Q_i = (Q_x^{(i)}, Q_y^{(i)})^{\mathrm{T}}$, where

$Q_x^{(i)} = \mu - a + \frac{2a}{N} i$ and $Q_y^{(i)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(Q_x^{(i)} - \mu)^2}{2*\sigma^2}}$. That is to say, system points are different for each of the test runs and their locations are determined by three random parameters: μ, σ and a. From above analysis, $Q_x^{(i)} \in [\mu - a, \mu + a], Q_y^{(i)} \in [\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{a^2}{2\sigma^2}}, \frac{1}{\sqrt{2\pi}\sigma}]$. In order to be showed on the computer screen in a suitable way, the coordinate range of system random points have to be transformed according to the original training PoGs.

Assuming original training PoGs are $\{P_1, \cdots, P_M\}$ (in this paper, $M = 9$) and $P_j = (P_x^{(j)}, P_y^{(j)})^{\mathrm{T}} \in \mathrm{R}^{2\times 1}$. The locations of training PoGs are shown in Fig. 3a, which are uniformly distributed on the computer screen. In this paper, linear transformation is used as follows.

$$\begin{aligned} \hat{Q}_x^{(i)} &= \frac{PH_x - PL_x}{QH_x - QL_x}(Q_x^{(i)} - QL_x) + PL_x \\ \hat{Q}_y^{(i)} &= \frac{PH_y - PL_y}{QH_y - QL_y}(Q_y^{(i)} - QL_x) + PL_y \end{aligned} \tag{3}$$

where $\hat{Q} = (\hat{Q}_x^{(i)}, \hat{Q}_y^{(i)})^{\mathrm{T}}$ is the transformed system random point. $AH_c = \max A_c^{(i)}$, $AL_c = \max A_c^{(i)}$, $A = \{P, Q\}$, $c = \{x, y\}$. An example of system random point is shown in Fig. 3b. Green big circles are training PoG and red small circles are system random points. The arrows represent the movement direction of system random points.



(b) The layout of training POGs (c) An example of test POGs layout

Fig. 3. Data collection system for gaze estimation.

Based on gaze estimation model, gaze positions of user can be predicted under the guide of system random points. Euclidean distance is used here as fake score to evaluate the matching degree between system point locations and predicted gaze positions. The lower the score is, the more probable the user is genuine.

3 Experiments

3.1 Database

Publicly available databases such as CASIA [10] and Replay-attack [22] don't contain gaze information, thus they are unsuitable for evaluating our proposed

method. In this paper, we collect a database composed of 18 subjects. For each subject, there are four kinds of data: training data for gaze estimation model, test data for gaze estimation model, data of genuine face and data of replay attack. In the following section, we use Data-Train-Gaze, Data-Test-Gaze, Data-Test-Genuine and Data-Test-Replay to represent these four kinds of data. Data-Train-Gaze and Data-Test-Gaze are used to train and test gaze estimation model. Data-Test-Genuine and Data-Test-Replay are Genuine access and replay attack data and used for evaluating proposed spoofing detection method.

In order to collect data, we develop a system on a desktop composed of a 19-inch computer screen with 1440×900 pixels resolution and a generic webcam with 640×480 pixels resolution. To collect Data-Train-Gaze, $M = 9$ fixed markers are showed on the computer screen (Fig. 3a). The system captures user's frontal appearance while his gaze is focusing on every marker shown on the screen. In this paper there are 20 images are captured at each marker for each user, totally $20 \times 9 \times 18 = 3240$ frontal images. By artificially removing eye-closed images, there are 2917 frontal images left. Considering the negative effect of optical reflection, users are required to remove glasses during the data collection.

To collect Data-Test-Genuine and Data-Test-Replay, $N = 51$ system points following Gaussian distribution with random mean and variance appear one by one (Fig. 3b) on the computer screen. The system camera captures user's frontal appearance while these points are shown on the screen. In this paper there are 10 images are captured at each system point and each kind of data, totally $51 \times 10 \times 2 = 1020$ frontal images for each subject. Considering that user may not respond to system points timely, for each system point, the first and last two frontal images are removed, Totally $51 \times 6 \times 2 = 612$ images left. It should be noted that during this process, user is not asked to watch point when system points are appearing.

The data collection process of Data-Test-Gaze is almost the same with that of Data-Test-Genuine. The difference is, during this process user is asked to watch these system points and follow them.

3.2 Experimental Results

In this section, we will verify the effectiveness of proposed method from the following three aspects: 1) Effectiveness of adaptive linear regression with incremental learning; 2) Effectiveness of proposed method for distinguishing replay attacks from genuine accesses; 3) Effectiveness of Euclidean distance based liveness score.

Effectiveness of Incremental Learning Gaze error [28] is commonly used to evaluate the gaze estimation model.

$$error = arctan\left(\frac{\|y - \hat{y}\|_2}{d_{user}}\right) \tag{4}$$

where $\|y - \hat{y}\|_2$ represents the Euclidean distance between ground truth and predicted value, and d_{user} refers to the distance between user's eye with computer screen.

In order to verify the effectiveness of incremental learning, we compare proposed adaptive linear regression with incremental learning with that without incremental learning. In addition, to show the effectiveness of person-specific gaze estimation model, we also compare proposed method with subject-dependent adaptive linear regression. Subject-dependent experiment is conducted by training samples of all subjects instead of one subject. Compared results are illustrated in Fig. 4, which shows that 1) compared with original adaptive linear learning, proposed method with incremental learning achieves lower average gaze error; 2) compared with subject-dependent method, subject-specific methods obtain lower gaze error. Therefore, proposed method for gaze estimation is effective. Considering the gaze errors for 18 subjects are tolerated, adaptive linear regression with incremental learning can be embedded into proposed spoofing detection system.

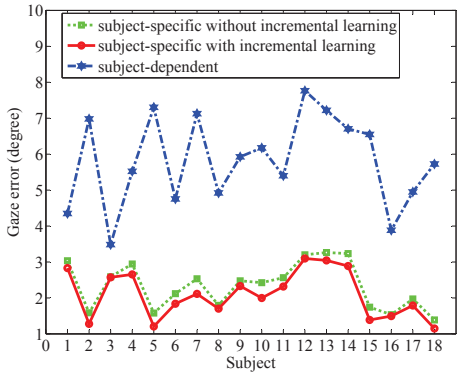


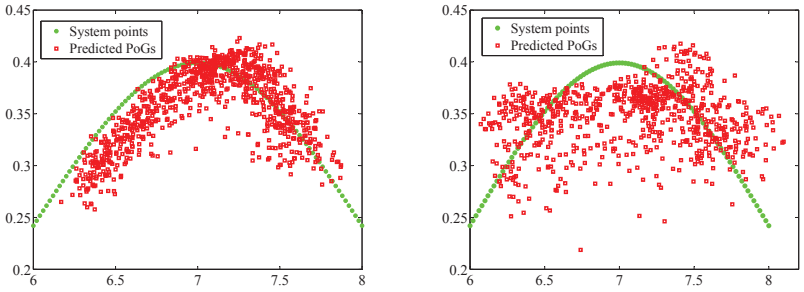
Fig. 4. Compared results on gaze error.

Effectiveness of Proposed Face Spoofing Detection Method EER (Equal Error Rate) and recognition accuracy are adapted as evaluation metrics. EER is the value when FRR (False Rejection Rate) equals to FAR (False Acceptance Rate). Recognition accuracy is reported based on cross-validation sets. 18 subjects are divided into 6 cross-validation sets and for each set there are 30 training samples (15 sequences of genuine faces and 15 sequences of replay attacks) for 15 subjects and 6 test samples (3 sequences of genuine faces and 3 sequences of replay attacks) for another 3 subjects. At each round, thresh is selected on training samples and used on test samples. The final recognition accuracy is achieved by averaging all the results on 6 sets of test samples. what's more, FRR values are also reported when FAR = 0.1, 0.01 and 0.001. Experimental results are listed in Table 1, which shows that proposed method perform excellent in distinguishing replay attacks from genuine accesses.

EER	FRR(FAR=0.1)	FRR(FAR=0.01)	FRR(FAR=0.001)	accuracy
0%	0%	0%	0%	100%

Table 1. EER FRR and recognition accuracy.

To further verify the effectiveness of proposed method, predicted gaze trajectories of genuine access and replay attack for one subject under the system point challenge are given in Fig. 5. Fig. 5a shows that genuine face is completely attracted by system points and follows them well. However, the replay video collected can not respond the challenge and predicted gaze trajectory is disorder (Fig. 5b). Therefore, proposed method is reasonable and effective.



(a) (b)

Fig. 5. An example of gaze movement for one subjects. (a) Gaze movement of Genuine face. (b) Gaze movement of replay attack.

Effectiveness of Euclidean Distance based Fake Score Fig. 6 illustrates the fake scores of samples for 18 subjects and shows that scores of real faces are averagely lower than that of replay attacks. Considering the fact that different from replay attacks, genuine accesses can be attracted by some objects or regions even in a long while. Experimental results show that the hypothesis of proposed method matches the real case, therefore, Euclidean distance based fake score is a good indicator.

4 Conclusion and Future Work

In this paper we propose an effective spoofing detection method for replay attack based on gaze estimation. Proposed spoofing detection method contains three key stages: face identification, gaze estimation and liveness judgement. In order to obtain lower gaze error, gaze estimation model is trained for each subject which dismisses the interferences among different subjects. In addition, adaptive learning regression is used for gaze estimation and improved with incremental learning. Then Euclidean distance based fake score is used to evaluate the difference between predicted gaze positions and system random point locations.

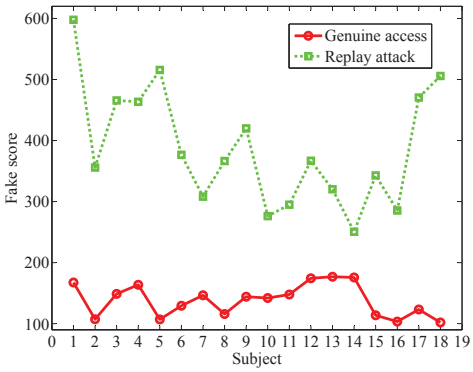


Fig. 6. Fake scores of real faces and replay attacks for 18 subjects.

Experimental results on collected database show that proposed method can effectively distinguish replay attacks from genuine accesses. We believe that with gaze estimation becoming more and more accurate, proposed spoofing detection method based on gaze estimation will have a good applicant prospect. However, how to deal with head pose in this work is still an challenge problem that we will research on.

References

1. Kollreider, K., Fronthaler, H., Bigun J.: Non-intrusive liveness detection by face images. Image Vision Computing 27, 223–244 (2009)
2. Bao, W., Li, H., Li, N., Jiang, W.: A liveness detection method for face recognition based on optical flow field. In: International Conference on Image Analysis and Signal Processing, pp. 233–236. IEEE Press, Taizhou (2009)
3. Anjos, A., Marcel, S.: Counter-measures to photo attacks in face recognition: a public database and a baseline. In: International Joint Conference on Biometrics Compendium, pp. 1–7. IEEE Press, Washington (2011)
4. Anjos, A., Mohan, M., Marcel, S.: Motion-based counter-measures to photo attacks in face recognition. IET Biometrics 3, 147–158 (2014)
5. Pan, G., Sun, L., Wu, Z., Wang, Y.: Monocular camera-based face liveness detection by combining eyeblink and scene context. J. Telecom. Syst. 47, 215–225 (2011)
6. Jukka, M.P., Hadid, A., Pietikinen, M.: Face spoofing detection from single images using micro-texture analysis. In: International Joint Conference on Biometrics Compendium, pp. 1–7. IEEE Press, Washington (2011)
7. Maatta, J., Hadid, A., Pietikainen, M.: Face spoofing detection from single images using texture and local shape analysis. IET Biometrics 1, 3–10 (2012)
8. Tan, X., Li, Y., Liu, J., Jiang, L.: Face liveness detection from a single image with sparse low rank bilinear discriminative model. In: 11 European Conference on Computer Vision, pp. 504–517. Springer, Greece (2010)
9. Komulainen, J., Hadid, A., Pietikainen, M.: Face spoofing detection using dynamic texture. In: Asian Conference on Computer Vision Workshops, pp. 146–157. Springer, Daejeon (2013)

10. Yan, J.J., Zhang, Z.W., Lei, Z., Yi, D., Li, S.Z.: Face liveness detection by exploring multiple scenic clues. In: International Conference on Control Automation Robotics and Vision, pp. 188–193. (2012)
11. Komulainen, J., Hadid, A., Pietikainen, M., Anjos, A., Marcel, S.: Complementary countermeasures for detecting scenic face spoofing attacks. In: International Conference on Biometrics Compendium, pp. 1–7. IEEE Press, Madrid(2013)
12. Frischholz, R.W., Dieckmann, U.: Bioid: A multimodal biometric identification system. Comput. 33 64–68 (2000)
13. Eveno, N., Besacier, L.: Co-inertia analysis for "liveness" test in audio-visual biometrics. In: 4th International Symposium on Image and Signal Processing and Analysis, pp. 257–261. IEEE Press (2005)
14. Chetty, G., Wagner, M.: Liveness verification in audio-video speaker authentication. In: 8th International Conference on Spoken Language Processing, pp. 363–385. JeJu (2004)
15. Zhang, Z.W., Yi, D., Lei, Z., Li, S.Z.: Face liveness detection by learning multispectral reflectance distributions. In: IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, pp. 436–441. IEEE Press, Santa BarBara (2011)
16. Kim, Y., Na, J., Yoon, S., Yi, J.: Masked fake face detection using radiance measurements. J. Opt. Soc. Am. A 24, 760–766 (2009)
17. Smith, D.F, Wiliem, A.,Lovell, B.C.: Face Recognition on Consumer Devices: Reflections on Replay Attacks. IEEE Trans. Inf. Foren. Sec. 10, 736–745 (2015)
18. Sireesha, M.V., Vijaya, P.A., Chellamma, K.: A survey on gaze estimation techniques. In: International Conference on VLSI, Communication, Advanced Devices, Signals and Systems and Networking, pp. 353–361. Springer (2013)
19. Ali, A., Deravi, F., Hoque, S.: Liveness detection using gaze collinearity In: Third International Conference on Emerging Security Technologies, pp. 62–65. IEEE Press, Lisbon (2012)
20. Ali, A., Deravi, F., Hoque, S.: Directional sensitivity of gaze-collinearity features in liveness detections. In: Fourth International Conference on Emerging Security Technologies, pp. 8–11. IEEE Press, Cambridge (2013)
21. Ali, A., Deravi, F., Hoque, S.: Spoofing attempt detection using gaze colocation. In: International Conference on Biometric Special Interst Group, pp. 1–12. IEEE Press, Damstadt (2013)
22. Cai, L.J., Xiong, C.S., Huang, L., Liu, C.P.: A novel face spoofing detection method based on gaze estimation. In: 12th Asian Conference on Computer Vision, pp. 547–561. Springer, Singapore (2014)
23. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: DeepFace: closing the gap to human-level performance in face verification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1707–1708. IEEE Press, Columbus (2014)
24. Lu, C.C., Tang, X.O.: Learning the Face Prior for Bayesian Face Recognition. In: European Conference on Computer Vision, pp. 119–134. Springer (2014)
25. Sigut, J.F., Sidha, S.A.: Iris center corneal reflection method for gaze tracking using visible light. IEEE Trans. Biomed. Eng. 58, 411-419 (2011)
26. Xiong, C.S., Huang, L., Liu, C.P.: Gaze Estimation based on 3D Face Structure and Pupil Centers. In: International Conference on Pattern Recognition, pp. 24–28. IEEE Press, Stockholm (2014)
27. Williams, O., Blake, A., Cipolla, R.: Sparse and semi-supervised visual mapping with the S3GP. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 230–237. IEEE Press (2006)

28. Feng, L., Sugano, Y., Takahiro, O., Sato, Y.: Inferring human gaze from appearance via adaptive linear regression. In: International Conference on Computer Vision, pp. 153–160. IEEE Press, Barcelona(2011)
29. Viola, P., Jones, M.: Robust Real-time Face Detection. Int. J. Comput. Vision 57, 137–154 (2004)