# Directional Sensitivity of Gaze-Collinearity Features in Liveness Detection

Asad Ali, Farzin Deravi, and Sanaul Hoque
University of Kent, Canterbury, Kent, CT2 7NT, United Kingdom
E-mail: {aa623, f.deravi, s.hoque}@kent.ac.uk

**Abstract** - To increase the trust in using face recognition systems, these need to be capable of differentiating between face images captured from a real person and those captured from photos or similar artifacts presented at the sensor. Methods have been published for face liveness detection by measuring the gaze of a user while the user tracks an object on the screen, which appears at pre-defined, places randomly. In this paper we explore the sensitivity of such a system to different stimulus alignments. The aim is to establish whether there is such sensitivity and if so to explore how this may be exploited for improving the design of the stimulus. The results suggest that collecting feature points along the horizontal direction is more effective than the vertical direction for liveness detection.

**Keywords—**liveness; spoofing; challege/response; collinearity; biometrics.

## I. Introduction

Biometric recognition systems are vulnerable to increasingly sophisticated spoofing attacks with the use of fake artifacts. Face recognition systems are socially acceptable and convenient and used for a variety of security applications, however, they appear to be more vulnerable to abuse compared to other biometric modalities, because a simple photograph or video of a genuine user can be used to deceive such systems [1]. Therefore, by introducing a liveness detection mechanism, the trust in using face recognition can be increased.

Photographs, masks, and videos are some of the spoofing artifacts that may be used for attacking face recognition systems at sensor level. Photo spoofing attacks can be prevented by detecting motion, smiles, eye blinks, etc. However, presenting a video of the genuine user to the face recognition system can deceive such techniques. The subtle differences between a photograph (or video) of an individual and the live person need to be utilized to establish liveness of the presentation at the sensor.

Direct user interactions with the system in real time can provide an important source of liveness information. In this paper we present a challenge/response mechanism for a facial liveness detection system, using a standard webcam, based on tracking the gaze of the user. The stimulus is designed to facilitate the acquisition of distinguishing features based on collinear sets of points along the gaze trajectory. Here we explore the sensitivity of such a system to different stimulus alignments. The aim is to establish whether spoof detection performance is affected by different directional arrangements of stimulus groupings for feature extraction and if so explore how this may be exploited for improving the design of the stimulus

The paper is organized as follows. In Section II a brief overview of the state of the art is presented. Section III describes the proposed techniques while Section IV reports on their experimental evaluation. Finally Section V provides conclusions and offers suggestions for further work

## II. Related work

Various approaches have been presented in the literature to establish liveness and to detect presentation attacks. Jee *et al's* method [2] uses a single ordinary video camera and analyses the sequence of the images captured to calculate variations of each eye region and determine whether the input face is real or not. Wang et al [3] presented a liveness detection method in which physiological motion is detected by estimating the eye blink from a captured video sequence using an eye contour extraction algorithm. They use active shape models with a random forest classifier trained to recognize the local appearance around each landmark. Wang et al [4] proposed a method to counter spoofing attacks by recovering sparse 3D facial structure. They captured facial images from several viewpoints and located landmarks. Then, they recovered sparse 3D facial structure from the selected key frames. They used graph similarity to incrementally extract the key frames.

Komulainen et al [5] explored fusion of motion and micro-texture. They explored the fusion potential of different visual cues and showed that the performance of the individual methods can be vastly improved by performing fusion at score level. Kollreider et al [6-8] combined facial components (nose, ears, etc.) detection and optical flow estimation to determine a liveness score. They assumed that a 3D face produces a special 2D motion. This motion is higher at central face parts (e.g. nose) compared to the outer face regions (e.g. ears). Parts nearer to the camera move differently to parts which are further away in a live face. A translated photograph, by contrast, generates constant motion at various face regions. Li et al [9] explored a technique based on the analysis of 2-D Fourier spectra of the face image. They proposed the principle that as the size of a photograph is smaller than the real image and the photograph is flat, it therefore has fewer high frequency components than real face images. Kim et al [10] proposed a method for detecting a single fake image based on frequency and texture analyses. They exploited frequency and texture information using power spectrum. They also used Local Binary Pattern (LBP) features for analyzing the textures. They fused information of the decision values from the frequency-based classifier and the texture based classifier for detecting the fake faces. Pinto et al [11] used the noise signatures generated by the recaptured video to discriminate between live and fake attempts. They suggested that noise is an artifact generated when video is captured from playback attacks (and not from real scenes). They used the Fourier spectrum, computation of 'visual rhythm' and grey level co-occurrence matrices as feature descriptors.

Frischholz et al [12] investigated a challenge/response approach to enhance the security of the face recognition system. The users were required to look in certain directions, which were chosen by the system randomly. The system estimated the head pose and compared the real time movement (response) to the instructions asked by the system (challenge) to verify user authenticity. Kollreider et al [13] proposed a method, which uses lip-motion (without audio information) to assess liveness. Ali et al [14] presented a method based on gaze tracking. They presented a video of a moving object on the screen. Users are required to follow the object with their head/gaze movement. The camera captures images of the user's face while the challenge moves. The path of the object is designed in such a way that a number of collinear points can be identified.

In this paper we investigate the work of Ali et al further to explore the sensitivity of the system to different gaze directions. The aim is to establish if there is such sensitivity and if so to explore how this may be used for improving the design of the stimulus.

## III. System structure

A spoofing attack (fake attempt or presentation attack) occurs when an impostor presents a photograph of the target individual to the camera to gain unauthorized access. A genuine (real) attempt occurs when an authorized individual attempts authentication using the biometric system. A general set up of the system is shown in Figure 1. The visual stimulus object (challenge) appeared on the display screen and the camera (sensor) captured the frames. The system extracts facial landmarks in the captured frames. Pupil centre coordinates are used to extract a feature vector and used for the classification of live and fake attempts. The visual stimulus, landmarks extraction and the feature vector are described below. The classification schemes used are discussed in detail in section IV-C.

The object appears at pre-defined locations randomly on the screen and the users were required to find it with eye or head movements and gaze at it. The stimuli locations were designed not to appear too close to one another on the screen, and should be located such that it is possible to find sets of collinear points along the horizontal and vertical directions. At each appearance of the stimulus, the camera captures an image of the user's face. The stimulus appears at 30 different places over the screen in 5 rows and 6 columns, giving 5 sets of collinear points horizontally and 6 sets of collinear points vertically.

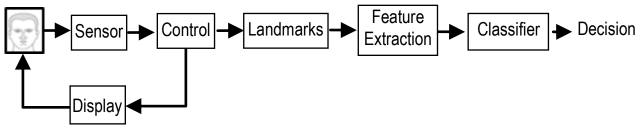

Fig. 1. System block diagram

The object appears in a random sequence to prevent predictive video attacks. The object visits positions where x- or y-coordinates of the current location is the same as that for one of the previous locations. In this way collinear sets of points of gaze can be identified.

The images captured were analyzed using STASM [15] to extract facial landmark points. STASM returns 68 different landmarks on the face region using the active shape model technique. The coordinates of the center of the pupils were used to extract features

Collinearity feature vectors were extracted from the facial images captured when the stimulus locations appeared along horizontal and vertical lines. When the stimulus locations on screen are along a vertical line, the x-coordinate values of these locations are the same. Similarly when the stimulus locations are arranged along a horizontal line, the y-coordinates are the same. Therefore, it may be assumed that the x- and y-coordinates of the corresponding centers of the pupils for these sets of stimulus locations should also be very similar in genuine attempts. This should result in a very small variance in the observed coordinate values for these sets of collinear points compared to that obtained for fake attempts. The collinearity feature vector is therefore a set of variances of face landmark coordinates extracted from multiple sets of collinear challenges/targets.

## IV. Experiments

### A. Hardware setup and Database

The experimental setup consists of a PC with a webcam, and a display monitor. The distance between the camera and the user was approximately 750 mm. This distance was not a tight constraint but had to be small enough so that the facial features could be easily acquired by the camera.

There is no publically available database that could be used for the particular challenge response scheme proposed in this paper. Therefore a small amount of data was collected to investigate the performance of the proposed scheme. In total 8 subjects participate in the data collection phase. The data was captured in 3 sessions. A total of 26 fake and 26 live attempts were captured. The user presented a high quality colour photo of a target user in front of the camera of while attempting to follow the stimulus to spoof the system. Each attempt acquired 90 image frames of resolution 352×288 pixels. This resolution gave good enough picture quality to recognize the facial landmarks by STASM. In total, 11 x-y-coordinates, 6 x-coordinates and 5 y-coordinates from the centres of the pupils collinear gazes were extracted. There were a small number of frames where the pupil centres were not detected by STASM and such frames were excluded from the feature extraction process.

### B. Evaluvation framwork

There are four possible outcomes of the face liveness detection classification process: true positive, true negative, false negative and false positive. When a genuine (live/real) attempt is classified as genuine and a false (fake/spoof) attempt is classified as genuine, these are termed true positive (TP) and false positive (FP) classifications respectively. Similarly, when a genuine attempt is classified as a fake and fake attempt is classified as fake these are called true negative (FN) and false negative (TN) respectively.

Error rates are dependent on the classifier threshold in use. The Receiver Operating Characteristic (ROC) curve is used to illustrate the true positive rates (TPR) against the false positive rates (FPR) for different thresholds.

The database was divided into two parts for training and testing purposes. Of the 52 samples, 12 were chosen for testing and the remaining 40 for training the classifier. For training the classifier, 20 random samples from fake and 20 from genuine users were chosen. The experiments were repeated 300 times, and on each occasion the system chose random samples for testing and training. The mean error rates of the 300 iterations of the experiments are reported here.

### C. Proposed Various Combining Schemes

In this investigation, various coordinates of the features were used to analyse and compare the performance of the x and y coordinates of the features. In the first phase the x-coordinates from both eyes (left and right eye) were used for the classification of fake and live attempts. The x-coordinates from left and right eyes were passed to independent k-nearest neighbor (kNN) classifiers [16]. In this implementation, k has been optimized to minimize the leave-one-out error in the training data. The normalized scores (based on the posterior probabilities of class membership) from these classifiers were fused to produce a single score, using various rule-based fusion schemes.

In the second phase of the experiments the y-coordinates from both eyes were used for the classification of fake and live attempts. In the final phase x-coordinates and y-coordinates from both eyes were used, using fusion rules as shown in Figure 2.

The sum, product and majority-vote rules [17] were investigated for the fusion process. Figure 3 shows the ROC curves, which describe the performance of the various feature combinations. The performance of the system was found to be lower using the y-coordinate features. Using x-coordinate

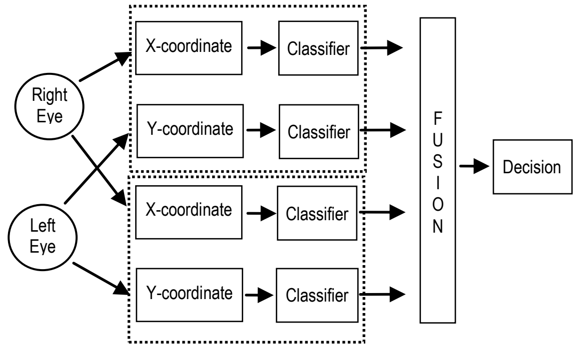


Fig. 2. Score fusion using left and right eye

features alone improved the performance but using both x- and y-coordinates performed best. Using both x- and y-coordinates of both eyes, the system performance reached 75% TPR (at 10% FPR). Using only the x-coordinate of both eyes the system achieved 72% TPR (at 10% FPR).

The scores were combined using the sum rule score fusion. At the lower FPR (<0.10), the ROC curve of the x-coordinate features is similar to the ROC curve of x and y coordinate features. The ROC curves for these features rise rapidly with increasing FAR and show much better performance than that for the y-coordinate features. In conclusion, these results suggest that the x-coordinate features are better compared to the y coordinate features. The improvement in the performance between using features based on the x-coordinate only and the x and y-coordinates together is very small. The system performance may be improved more efficiently by not using y-coordinate features and instead increasing the number or range of the x-coordinate features.

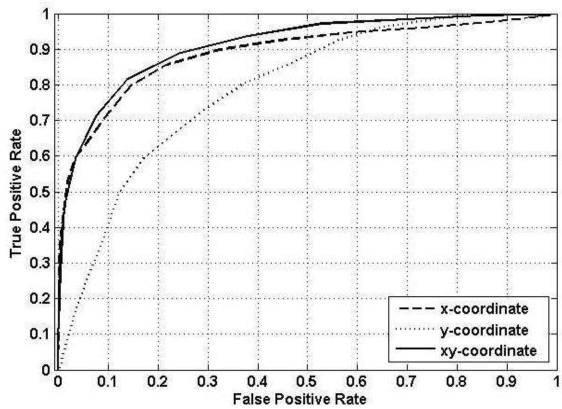


Fig. 3. ROC curves showing the performances of the three proposed scheme

Table I shows the performance of the system using various feature subsets, and fusion schemes. The x-coordinate features gave better performance compared to the y-coordinate features. But combining the x and y-coordinate improve the performance slightly. The sum score fusion rule gave the most promising result.

TABLE I. PERFORMANCE COMPARISON OF DIFFERENT FEATURES

| TPR | | | | |
|---|---|---|---|---|
| | **Ruled base fusion** | **@FPR =0.02** | **@FPR =0.05** | **@FPR =0.10** |
| x-coordinate from both eyes | Product | 0.53 | .062 | 0.70 |
| y-coordinate from both eyes | Product | 0.07 | 0.19 | 0.35 |
| x-y-coordinate from both eyes | Product | 0.50 | 0.63 | 0.72 |
| x-coordinate from both eyes | Sum | 0.52 | 0.63 | 0.72 |
| y-coordinate from both eyes | Sum | 0.10 | 0.23 | 0.41 |
| x-y-coordinate from both eyes | Sum | 0.54 | 0.64 | 0.75 |
| x-coordinate from both eyes | Majority vote | 0.21 | 0.53 | 0.64 |
| y-coordinate from both eyes | Majority vote | 0.139 | 0.16 | 0.31 |
| x-y-coordinate from both eyes | Majority vote | 0.06 | 0.52 | 0.72 |

This is an interesting result that requires further investigation for the identification of its possible causes. It may be suggested that human beings can move and position their head/eye more easily and with more accuracy in the horizontal direction. This effect may also be due to the nature of the display screen used for the challenge i.e. the width of the screen is greater than the height.

## V. CONCLUSION

This paper presents a face liveness detection technique that may be used for a range of biometric applications. We proposed a challenge-response approach using a visual stimulus to measure the gaze of the user for the purpose of establishing the presence of photographic spoofing attacks. Collinearity features are then used to provide a measure to discriminate between live and fake attempts. We analysed the performance of the system using x- and y-coordinates of the pupil centre. The features based on x-coordinates of eye centre locations were found to be more effective for liveness detection. Given that the acquisition time will have to be bounded, this implies that the set of challenge points should be chosen to have more vertically collinear sets of points.

These preliminary results will have to be confirmed using a substantially bigger database with data from a large set of users taken on multiple occasions. Future work will be focused on the design of the stimulus so that the duration of the challenge is minimized while maintaining a high spoof detection rate. The conjecture that human beings can move and position their head/eye more easily and with more accuracy in the horizontal direction could also be further investigated with a series of tests using alternative stimuli and screen arrangements.

## REFERENCES


[1] R. Tronci, D. Muntoni, G. Fadda, M. Pili, N, Sirena, G. Murgia, M. Ristori and F. Roli, “Fusion of multiple clues for photo-attack detection in face recognition systems,” International Joint Conference on Biometrics (IJCB), pp. 1-6, October 2011.

[2] H. K. Jee, S. U. Jung, and J. H. Yoo, “Liveness detection for embedded face recognition system,” International Journal of Biological and Medical Sciences, vol. 1(4), pp. 235-238, 2006.

[3] L. Wang, X. Ding, and C. Fang, “Face Live Detection Method Based on Physiological Motion Analysis” Tsinghua Science and Technology, vol. 14(6), pp. 685-690, 2009.

[4] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, “Face Liveness Detection Using 3D Structure Recovered from a Single Camera,” International Conference on Biometrics,Madrid, Spain, 2013.

[5] J. Komulainen, A. Hadid, M. Pietikäinen A. Anjos, and S. Marcel, “Complementary countermeasures for detecting scenic face spoofing attacks,” International Conference on Biometrics,Madrid, Spain, 2013.

[6] K. Kollreider, H. Fronthaler, and J. Bigun, “Non-intrusive liveness detection by face images,” Image and Vision Computing, vol. 27(3), pp. 233–244, 2009.

[7] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, “Real-Time Face Detection and Motion Analysis with application in ‘Liveness’ Assessment,” IEEE Transaction on Information Forensics and Security, vol. 2(3), pp. 548-558, 2007.

[8] K. Kollreider H. Fronthaler, and J. Bigun, “Verifying liveness by multiple experts in face biometrics,” in Proc of IEEE Computer Vision and Pattern Recognition Workshop on Biometrics, Anchorage, AK, USA, Pp. 331-338, June 2008.

[9] J. Li Y. Wang, T. Tan, and A.K. Jain, “Live face detection based on the analysis of Fourier spectra,” in Proc of Biometric Technology for Human Identification, Orlando, FL, USA. (SPIE 5404), pp. 296-303, April 2004.

[10] G. Kim, S.Eum, J. K. Suhr, D. I. Kim, K. R. Park, and J, Kim, “Face liveness detection based on texture and frequency analyses,” 5th IAPR International Conference on Biometrics (ICB), New Delhi, India. pp. 67-72, March 2012.

[11] A. D. S. Pinto, H. Pedrini, W. Schwartz, and A. Rocha, “Video-Based Face Spoofing Detection through Visual Rhythm Analysis,” 25th SIBGRAPI Conference on Graphics, Patterns and Images, Ouro Perto, Brazil, pp. 221-228 August 2012.

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

[13] K. Kollreider H. Fronthaler, and J. Bigun, “Evaluating liveness by face images and the structure tensor,” in Proc of 4th IEEE Workshop on Automatic Identification Advanced Technologies, Washington DC, USA, pp. 75-80, October 2005.

[14] A. Ali F. Deravi, and S. Hoque, “Liveness Detection using Gaze Collinearity,” In Proc of 3rd Intl Conference on Emerging Security Technologies, Lisbon, Portugal, pp. 62-65, September. 2012.

[15] S. Milborrow, and F. Nicolls, “Locating facial features with an extended active shape model,” In Proc. of the 10th European Conference on Computer Vision (ECCV), Marseille, France, pp. 504-513, October 2008.

[16] R. O. Duda, P. E. Hart, and D. G. Stort, “ Pattern Classification”, Jone Wiiley and Sons, 2001.

[17] J. Kittler, M.Hatef, R. P. W. Duin, and J. Matas, “On combining classifiers,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20(3), pp. 226-239, March 1998.