Liveness Detection using Gaze Collinearity

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

This paper presents a liveness detection method based on tracking the gaze of the user of a face recognition system using a single camera. The user is required to follow a visual animation of a moving object on a display screen while his/her gaze is measured. The visual stimulus is designed to direct the gaze of the user to sets of collinear points on the screen. Features based on the measured collinearity of the observed gaze are then used to discriminate between live attempts at responding to this *challenge and those conducted by "impostors" holding* photographs and attempting to follow the stimulus. An initial set of experiments is reported that indicates the effectiveness of the proposed method in detecting this class of spoofing attacks

1. Introduction

Biometric systems have several advantages over most other security methods, such as PIN codes, passwords, keys, cards, IDs, tokens etc., which despite great success in recent years, they remain vulnerable to increasingly more sophisticated spoofing attacks using fake artifacts made from the biometric information of genuine users and presented to the system sensor(s). An impostor can present a photo or video of a genuine user to a face recognition system to gain access to unauthorised data or premises. To prevent such sensor-level spoofing, biometric systems need to establish the liveness of an acquired sample during the identity verification process.

Biometric technology surrounding facial recognition has developed rapidly in recent years as it is user friendly and convenient, and is used for many security purposes, but is vulnerable to abuse, such as photographic spoofing or video substitution and many others. However, by adding liveness detection the effectiveness of security systems can be substantially improved. The differences between a photograph or video of an individual and the real person can be used to establish liveness.

Photo spoofing can be averted by detecting smile, motion or eye blinks. However, this type of systems can be subverted by presenting a video of the genuine user to the face recognition system. To avoid video spoofing, more sophisticated methods have been suggested in the literature. Background clues and 3D facial images are exploited to avert video spoofing. Such techniques may also be subverted by controlling the video background or by wearing 3D masks.

An important source of liveness information is the direct user interactions with the system that are captured and assessed in real time. In this paper we present a novel challenge/response mechanism for a face-recognition system, using a single camera, based on tracking the gaze of the user moving in response to a visual stimulus. The stimulus is designed to facilitate the acquisition of distinguishing features based on the collinearity of sets of points along the gaze trajectory.

The paper is organized as follows. In Section 2 a brief overview of the previous work is offered. Section 3 presents the proposed technique. Section 4 reports on the experimental evaluation of the technique. Finally Section 5 offers conclusions and suggestions for further work

2. Related work

Various approaches have been presented in the literature to establish liveness in order to avert spoofing attempts. Liveness detection approaches can be grouped into two main categories, cooperative and non-cooperative. Cooperative approaches require user co-operation to enable the facial recognition system to estimate the liveness in the biometric samples captured at the sensor level. The non-cooperative approaches do not require user co-operation or even awareness but exploit involuntary physical movements, such as spontaneous eye blinks, and 3D properties of the image.

Systems based on challenge-response approach belong to the cooperative type, in which the user is asked to perform specific activities to ascertain liveness, such as uttering digits or changing the head pose.

Pan et al in [1] propose a liveness detection method by extracting the temporal information from the process of the eye blink. They used Conditional Random Fields to model and detect eye-blinks over a sequence of images. Jee et al's [2] method uses a single ordinary camera and

analyzes the sequence of the images captured. They locate the centre of both eyes in the facial image. If the variance is larger than a preset threshold, the image is considered as a live facial image. When the variance is smaller than the threshold the image is classifies as photograph.

Kollreider et al [3-5] 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. On the other hand, a translated photograph generates constant motion at various face regions. Wang et al [6] present a liveness detection method in which physiological motion is detected by estimating the eye blink and an eye contour extraction algorithm. They use the technique called active shape model with a random forest classifier trained to recognize the local appearance around each landmark. They also showed that if any motion in the face region is detected the sample is considered to be captured from an imposter.

Li et al [7] explore a technique based on the analysis of 2-D Fourier spectra of the face image. Their work is based on two principles. Firstly, they propose 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.

Frischholz et al [8] investigate a challenge-response approach to enhance the security of the face recognition system. The users are required to look in certain directions, which were chosen by the system randomly. The system estimates the head pose and compares the real time movement (response) to the instructions asked by the system (challenge) to verify the user authenticity.

Kollreider et al [9] propose a method to recognize the digits by lip-motion without audio information to assess its value for liveness detection.

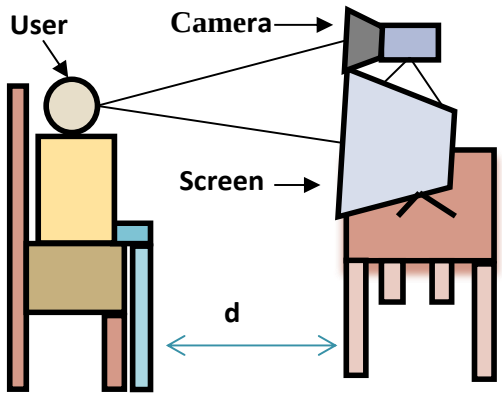

Figure 1. Experimental setup

3. Liveness detection through gaze tracking

The scenario considered in this paper is that of a face verification system using a single camera as its source. The spoofing attack would be through an imposter holding a photograph of a target to the camera and attempting authentication. A typical setting is depicted in Figure 1.

3.1. Visual stimulus

The user is presented with a video animation of a small moving object on the screen (challenge) and is required to follow it with his head/gaze movement. At each position of the stimulus, the camera captures an image of the user's face. The trajectory of the object is chosen so that it changes after every use in a random fashion to prevent predictive video attacks. The path of the object is chosen in such a way that a number of collinear points can be identified.

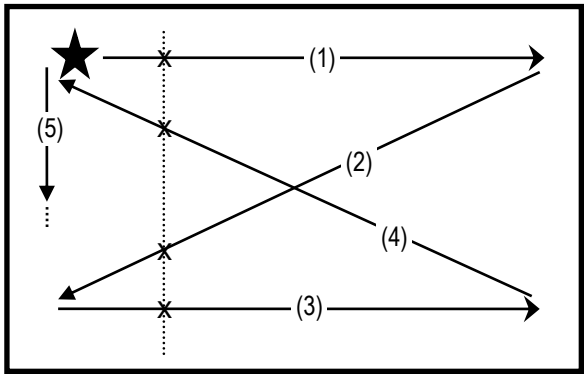

Figure 2. Trajectory of the challenge and collinear points

3.2. Face/Eyes detection

The images captured during the challenge were analysed to extract facial landmark points using STASM [10]. STASM uses the Viola-Jones detector to localize face and eyes in the input frames and returns 68 landmarks on the face region using active shape model technique. A subset of these 68 points (e.g., center of the pupils, corner of the eyes, etc.) will be used for feature extraction in the proposed scheme.

3.3. Collinearity features

For the observations reported here, only the centres of the pupils in the captured frames were used. For the vertically collinear points, the 'X'-coordinate values of the target are same. It can then be hypothesized that the x-coordinates of the pupil centres in the corresponding

frames should also be very similar. This should result in a very small variance in the observed x-coordinates (σ_x^2) of the pupil centre. Since there are many such sets of vertically collinear points, in order to reduce the feature dimensionality, the mean of these variances were used as the discriminatory feature. In a similar fashion, the mean of the variance of the y-coordinates (σ_y^2) for the horizontally collinear sets were included in the feature vector. Similar features can be extracted from other facial landmarks, but were not used in the results reported here.

4. Experiments

Experiments were carried out to verify the performance of the proposed algorithm in distinguishing genuine attempts from fakes. A setup similar to Figure 1 was used.

The setup consists of a webcam, a PC and a display monitor. The camera used is a Logitech Quick Cam Pro 5000, and is centrally mounted on the top of a 21.5" LCD screen, a commonly used monitor type, having a resolution of 1920×1080 pixels and 5ms response time. The computer has quad core processors with 3.2 GHz clock frequency, and 2 GB of RAM. The distance between the camera and the user is approximately 750 mm, it is not restricted to any particular distance, but must be near enough so that the facial features can be clearly acquired by the camera.

We collected data from 5 subjects in 3 sessions. Each person performed 3 fake and 3 genuine attempts in total, creating 15 sets of fake and 15 sets of genuine attempts. During the spoof attempts, the user held a high quality colour photo of a genuine user in front of the camera and tried moving the same to follow the stimulus. Each attempt acquired 358 image frames, and the resolution of the images is 352×288 pixels. This resolution gives good enough picture quality to recognize the facial landmarks. Increasing the resolution did not improve the accuracy of the proposed method while increasing the processing time. In total, 48 vertically collinear and 24 horizontal collinear point sets were extracted. There were a few frames where the pupil centre was not found by the STASM software and such frames (and associated collinear points) were excluded from feature extraction process.

4.1. Experimental Results

Figure 3 illustrates the distribution of the genuine and impostor attempts in the feature space. As expected, genuine users showed much smaller variances compared to those of the fake attempts in most of the cases.

For a small number of fake attempts, features very similar to genuine users are evident. This phenomena is not unusual, for example, if the impostor holds the photo still (i.e., no attempt to respond to the challenge) thus producing zero or very small variances. This may also happen when the genuine user is non-cooperative or non-responsive. On the other hand, the genuine user response to the challenge involves independent head and eye movements and also due to the fact that users have a wide field of view, some may produce large variances too. Such behaviour can be treated as suspect (denoted by outliers) and be flagged as fake attempts. Figure 4 shows the distribution when these outliers are excluded. It is evident that in this case the separation between genuine and impostor features has become more prominent.

The criteria used for the identification of outliers are shown in Table 1. Rule 1 excludes the cases where the net variance in X and Y-coordinates are smaller than certain thresholds whereas Rule 2 filters those with very large net variances. The actual thresholds were decided empirically.

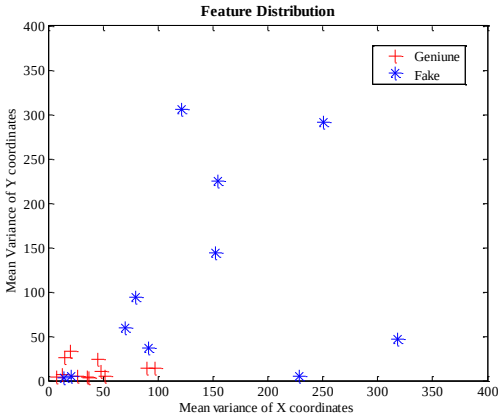

Figure 3. Feature distribution

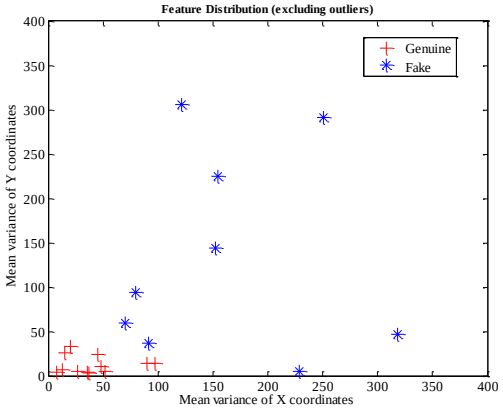

Figure 4. Feature distribution with outlier exclusion

Table 1. Outlier filter criteria

Rule 1	$\sum\sigma_x^2 < 130$ AND $\sum\sigma_y^2 < 13$
Rule 2	$\sum\sigma_x^2 > 1000$ AND $\sum\sigma_y^2 > 300$

The liveness detection scheme proposed here is a two phase process. In the first phase, the scheme applies the outliers rule and if true, identifies the attempt as inconclusive and more data would be needed to establish liveness of the user In the second phase, a linear discriminant classifier using the collinearity features is employed to decide the liveness of the user. Table 2 shows the accuracy of the proposed method both with and without the outlier detection phase.

Table 2. Performance of the proposed method

	FAR	FRR
Outliers not excluded	13.3%	0%
Outliers removed	0%	0%

Here, FAR represents cases where an impostor holding a photo is accepted as a live genuine user whereas FRR represents the cases where a genuine user is rejected as an impostor. The results show that in both configurations (with or without the outlier detection phase) the FRR is 0%. Exclusion of the outliers reduced the FAR to 0% too. The initial set of experimental results, therefore, indicates the potential of the proposed method in detecting spoofing attacks. Table 3 presents a comparative performance analysis for the proposed technique.

Table 3. Comparative performance analysis

Method	FAR	FRR
Wang [6]	0%	2.5%
Kollreider [9]	1.5%	19%
Proposed method	0%	0%

5. Conclusion

This paper presents a novel technique for liveness detection in the presence of photo spoofing attacks on face verification systems. A challenge-response approach using a visual stimulus to direct the gaze of the users is combined with gaze collinearity features to provide a measure of discrimination between genuine and fake attempts.

Initial experiments indicate the potential viability of this approach. Future work will expand the experiments to include more users and attempts and will also explore additional features for improving the anti-spoofing capabilities of the system in response to more sophisticated attacks.

6. References

[1] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera", in Proc. of 11th IEEE Intl Conference on Computer Vision, Rio de Janeiro, Brazil, pp.1-8, 2007.

[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] 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.

[4] K. Kollreider, H. Fronthaler, M. 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.

[5] 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. pages 331–338, June 2008.

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

[7] 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.

[8] 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.

[9] 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, pp.75-80, Washington DC, USA, October 2005.

[10] 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, October 2008.