Spoofing Attempt Detection using Gaze Colocation

Asad Ali, Farzin Deravi and Sanaul Hoque

School of Engineering and Digital Arts
University of Kent, Canterbury, Kent, CT2 7NT, United Kingdom
E-mail: {aa623, f.deravi, s.hoque}@kent.ac.uk

Abstract: Spoofing attacks on biometric systems are one of the major impediments to their use for secure unattended applications.This paper presents a novel method for face liveness detection by tracking the gaze of the user with an ordinary webcam. In the proposed system, an object appears randomly on the display screen which the user is required to look at while their gaze is measured. The visual stimulus appears in such a way that it repeatedly directs the gaze of the user to specific points on the screen. Features extracted from images captured at these sets of colocated points are used to estimate the liveness of the user. A scenario is investigated where genuine users track the challenge with head/eye movements whereas the impostors hold a photograph of the target user and attempt to follow the stimulus during simulated spoofing attacks. The results from the experiments indicate the effectiveness of the gaze colocation feature in detecting spoofing attack.

1 Introduction

Despite the successes in biometric recognition systems in recent decades, they still remain vulnerable to increasingly sophisticated spoofing attacks with the use of fake artifacts. These artifacts may be created from the biometric information of genuine users and presented at the system sensor(s). An impostor can present a fake biometric sample of a genuine user to a biometric recognition system to gain access to unauthorised data or premises. This type of spoofing is a direct attack on the sensor (also known as *presentation attack*); the impostor does not require any *a priori* knowledge about the internal operation of the biometric system. To prevent such sensor-level attacks, biometric systems need to establish the liveness of the source of an acquired sample.

Amongst biometric modalities, face recognition has emerged as being socially acceptable, accurate and convenient and is therefore used for a variety of security applications. But face recognition systems may be considered 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 [Tr11]. Therefore, by introducing a liveness detection mechanism, the security of biometric systems can be substantially improved.

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

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 face-recognition systems, using a standard webcam, by 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 colocation of sets of points along the gaze trajectory.

The paper is organized as follows. In Section 2 a brief overview of the state of the art is presented. Section 3 describes the proposed techniques while Section 4 reports on its experimental evaluation. Finally Section 5 provides conclusions and offers suggestions for further work.

2 Related Work

Various approaches have been presented in the literature to establish liveness and to detect presentation attacks. Liveness detection approaches can be grouped into two broad categories: active and passive. Active approaches require user engagement to enable the facial recognition 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, such as spontaneous eye blinks, and 3D properties of the image.

Passive anti-spooing techniques are usually based on the detection of signs of life, e.g. eye blink, facial expression, etc. For example Pan *et al* [PWL07] proposed 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* method [JJY06] uses a single ordinary camera and analyses the sequence of the images captured. They locate the centre of both eyes in the facial image. If the variance of each eye region is larger than a preset threshold, the image is considered as a live facial image; otherwise the image is classified as a photograph. Wang *et al* [WDF09] presented a liveness detection method in which physiological motion is detected by estimating the eye blink with an eye contour extraction algorithm. They use active shape models 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 impostor. Kollreider *et al* [Ko09, Ko08, Ko05] 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. They also proposed a method which uses lip-motion (without audio information) to assess liveness [Ko05].

Some anti-spoofing techiques are based on the analysis of skin reflectance, texture, noise signature etc. Li *et al* [Li04] explored a technique based on the analysis of 2-D Fourier spectra of the face image. Their work is based on two principles. 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* [Ki12] 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* [Pi12] used the noise signatures generated by the recaptured video to discriminate between live and fake attempts. They suggested noise was the artifact generated from video captured from other video (and not from real scenes). They used the Fourier spectrum, computation of the visual rhythm and extraction of the grey level co-occurrence matrices as feature descriptors.

Systems based on the challenge-response approach belong to the active category, where the user is asked to perform specific activities to ascertain liveness such as uttering digits or changing his or her head pose. For instance Frischholz *et al* [FW03] 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 the user authenticity. Ali *et al* [ADH12] presented a method of liveness detection based on gaze tracking. Users are required to follow a moving object with their head/gaze while a camera captures images of the user's face. The path of the object is designed in such a way that a number of collinear points are visited. Work has also been reported on using the gaze trajectory as a source of biometric information [DG11].

The work presented here explores a new feature set, hereby referred to as the *gaze colocation feature* set, for the detection of presentation attacks. Although a similar setup to the one in [ADH12] has been used the novel features proposed here establish the ability of the natural gaze to return to the same location consistently. Here the users gaze is directed to some pre-selected random positions on the display and features are extracted from sets of gazes at these colocated targets.The underlying hypothsis is that the variance in gazes for colocated positions should be small in genuine user attempts. This phenomenon is then exploited to differentiate between a photo spoof attack and a genuine user input. Video spoofing presents an even greater challenge. A video camera is required and, as reported, sophisticated methods such as video background control, 3D masks, 3D facial images and placing fidcucial points in the background are all being employed to prevent video spoofing [Tr11, Pa11]. In this paper, however, we do not report results of tests on the proposed system under video spoofing attacks.

3 Liveness Detection through Gaze Tracking

The scenario considered in this paper is that of a face verification system using an ordinary camera (webcam). A block diagram of the proposed system is shown in Figure 1. An object appears on the display and the camera (sensor) captures the frames as the position of the object on the display changes. The gaze colocation features are extracted from the pupil centres in the captured frames which are then classified as genuine or fake.

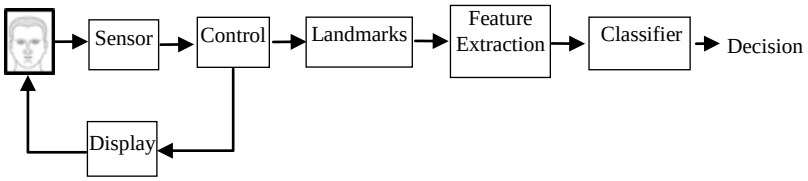
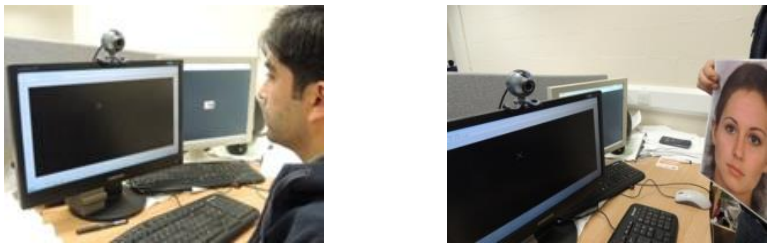


Figure 1: System block diagram

3.1 Visual Stimulus

A small object appears at random locations on the screen and the user is required to find and follow it with head/gaze movement. It is not necessary to space these targets uniformly but ideally these should not be too close to one another and each should be visited multiple times. At each appearance of the stimulus, the camera captures an image of the user's face. The presentation of the challenge takes approximately 130 seconds to complete, capturing 90 still images at each location of the challenge. The object appears in a random sequence to prevent predictive video attacks. The object visits each position at least three times. In this way a number of colocated sets of gaze can be identified. In Figure 2(a) a genuine user is seen tracking the challenge to establish liveness, while in Figure 2(b) the impostor is responding to the challenge by carefully shifting a high quality printed photo to gain access to the system.



(a) (b)

Figure 2: Example of (a) Genuine attempt, and (b) Spoof attempt

3.2 Facial Landmark Detection

The images captured during the challenge-response were analysed using STASM [MN08] to extract facial landmark points. STASM returns 68 different landmarks on the face region using an active shape model technique. The coordinates of the center of the pupils were used for feature extraction in the proposed scheme.

3.3 Gaze Colocation Features

The gaze colocation features are extracted from images when the stimulus is at a given location. The ‘x’ and ‘y’coordinates of the object on the display are same when they reappear at a given place at different times during this exercise. It can therefore be assumed that the ‘x’ and ‘y’ coordinates of the pupil centres in the corresponding frames should also be very close. This should result in a very small variance in the observed x- and y-coordinates of the pupil centres in genuine attempts. A feature vector is thus formed from the variances of pupil centre coordinates for all the occasions where the stimulus is colocated.

Similar features can be extracted from other facial landmarks, but were not used in the results reported here.

4 Experiments

The system setup was similar to the one shown in Figure 3. 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 distance between the camera and the user was approximately 750 mm. This distance was not a tight constraint but had to be such that the facial features could be clearly acquired by the camera.

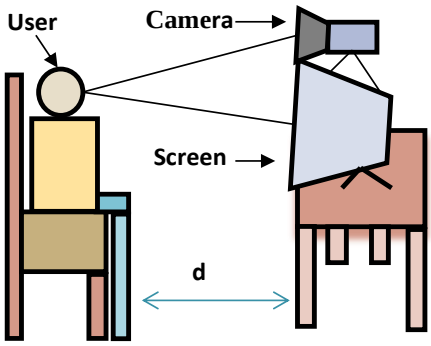
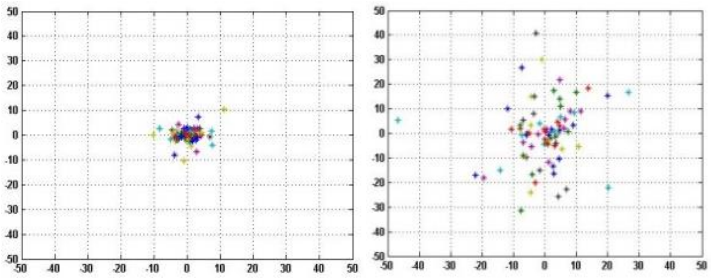


Figure 3: System Setup

Data was collected from 8 subjects in 3 sessions. Each subject provided data for both genuine and impostor attempts, creating 26 sets of each. During spoofing attacks a high quality colour photo of a genuine user was held in front of the camera while attempting to follow the stimulus. Each attempt acquired 90 image frames of resolution 352×288 pixels. This resolution provided a good enough picture quality to recognize the facial landmarks. In total, 30 sets of x-y coordinates of the pupil centres from collocated gaze targets were extracted resulting in a feature vector of size 60 for each eye. There were a small number frames where the pupil centres were not detected by STASM and such frames (and associated colocation points) were excluded from the feature extraction process.

For this data, the (x,y) coordinates of the pupil centres from frames captured while users are looking at the central stiumulus location are plotted in Figure 4 displaying deviations from their mean for all the genuine and fake attempts respectively. It can be observed that the range of the points in genuine attempts is much smaller compared to that of the spoof attempts. This is because the impostor, relying on hand-eye coordination, is unable to align the photo back to the same spot as accurately as a genuine user.



(a) (b)

Figure 4: *Pupil centre deviations capture during (a) genuine attempt and (b) spoof attempt for the central location of the target*

4.1 Evaluation Framework

Face liveness detection is a two-class problem and there are four possible outcomes of the classification process: true positive, true negative, false negative and false positive. When a genuine (live/non-spoof) 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. FP and FN are the errorenous outcomes of the process and the rates of their occurance is reported as False Positive Rate (FPR) and False Negative Rate (FNR) in this report in order to facilitate the assessment and comparison of system performance. The term True Positive Rate (TPR) is also used and is equal to 1-FNR. Total Error Rate (TER) can be defined as the proportion of misclassified attempts out ot all the attempts, including both genuine and fake.

For the experiments reported here, the database was divided into two disjoint sets 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 attempts were chosen. The experiments were repeated 50 times, and on each occasion the system used randomly selected samples for testing and training. The mean error rates are reported here.

4.2 Experimental Results

Error rates were calculated for a range of system parameters and are reported in this section. True Positive Rates at a set of predefined FPR values were obtained and used for comparison. The ROC curve of the proposed scheme using features from the single eye is presented in Figure 7. It is apparent that the system did not perform very accurately when the entire feature vectors are used. However, the performance improved significantly when subsets of the available features were used for training and testing. The forward feature selection method was used to rank the features [BL97]. the best results were achieved when a subset of the best 15 features was used (as shown in Figure 7). At 10% FPR, the TPR was above 90% which was only around 40% when using the entire feature set.

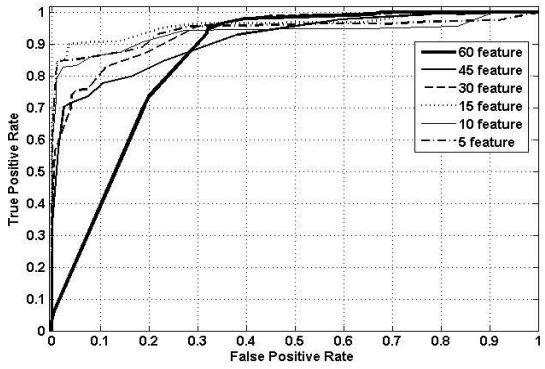


Figure 7: Performance with features extracted from a single eye

4.3 Feature Combination Schemes

While the colocation features from each eye may be used in isolation it is interesting to explore if there is complementarity in these feature sets and if a greater accuracy can be achieved by their combination. Therefore, both feature and score fusion schemes were explored to find if there can be gain in accuracy by combing information from features extracts from the two eyes. The following sub-sections will cover each of these fusion schemes in turn.

4.3.1 Feature Fusion

The features extracted from both the eyes were concatenated to form a larger feature vector which was then used for training and testing. All 60 features from the left eye and the 60 features from the right eye were combined in a feature-level fusion scheme. The scheme is illustrated in Figure 8. A feature selection method was incorporated to find the optimum feature subsets for this scheme.

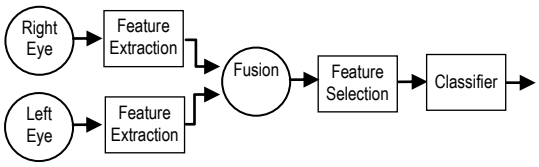


Figure 8: Feature fusion using left and right eye

Figure 9 shows the ROC curves for different feature dimensions. The TPR of the system was found to be lower than the instances when only one eye was used. Reducing the number of features improved the performance but the best TPR (at 10% FPR) of the system was about 80% while for the single-eye system it was above 90%.

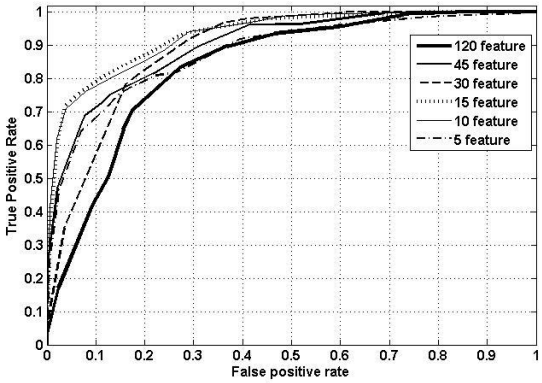


Figure 9: Feature fusion performance

4.3.2 Score fusion

An alternative to the feature fusion strategy, a score fusion scheme is often implemented. In the score fusion scheme, these features were extracted from the right and left eyes and independent classifiers were used to obtain classification scores for each eye. In this multi-classifier system two k-NN classifiers were used for each eye. The a posteriori probabilities from the two classifiers were combined using the 'product rule' for liveness

detection [Ki98]. Figure 10 illustrates the scheme and Figure 11 shows the corresponding ROC curves. The scheme achieved a TPR of 99% at FPR of 10%.

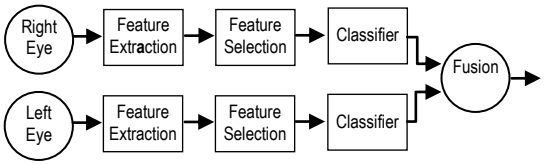


Figure 10: Score fusion scheme

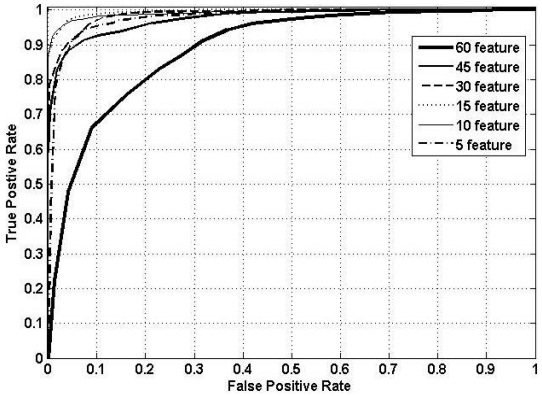


Figure 11: Score fusion performance

In order to establish the tradeoff between the feature dimensionality and liveness detection accuracy experiments were performed to establish the performance of the system as the number of dimensions was steadily reduced. Figure 12 illustrates total error rates for different feature dimensions selected using the forward feature selection method. It can be seen that the lowest total error rate was observed when the feature dimension was reduced to around 15. The total error rate started increasing when the feature set was further reduced. The system produced higher total error rates when the feature dimension was large. The reason for this might be that only a small amount of data was available for training given the size of the feature set.

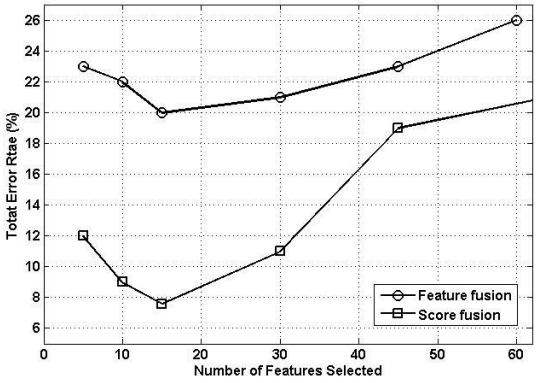


Figure 12: Variation in accuracy with feature dimension

Table 1 presents a comparative performance of the proposed methods at various levels of FPR. The feature fusion scheme gave the highest error rates in all cases. While using features from only one eye, the system TPR was up to 91%. This improved vastly when the score fusion approach was implemented. At 1% FPR, a TPR of 93% was achieved using score fusion. At 10% FPR, this rose to 99%.

Table 1: Performance comparison of the three schemes

	TPR			
	@FPR =0.01	@FPR =0.02	@FPR =0.05	@FPR =0.10
Single Eye	84%	86%	90%	91%
Feature Fusion	47%	62%	74%	78%
Score Fusion	93%	94%	97%	99%

Table 2 shows a comparative analysis of our experimental observations with the performances reported for similar spoof attacks published in the literature. Although the results are from different databases they suggest possible comparative ranking of these various methods and indicate that the proposed method compares favourably with these schemes.

Table 2: Comparative performance analysis

Method	FPR	FNR	TER
Ali *et al* [ADH12]	13.3%	0.0%	6.7%
Kollreider *et al* [Ko07]	1.5%	19.0%	10.3%
Tan *et al* [PMR11]	9.3%	17.6%	13.5%
Peixoto *et al* [PMR11]	6.3%	7.0%	7.0%
Proposed method	1.0%	7%	4.0%

5 Conclusion

This paper presents a novel feature set for liveness detection in the presence of photo spoofing for face verification systems. A challenge-response approach is described which uses a visual stimulus to direct the gaze. The test scenario did not constrain the users to move either their head or eyes exclusively. However, the proposed gaze colocation features provided a robust measure for discriminating between live and fake attempts.

Initial experiments prove the potential viability of this approach, however, more data is required to establish the performance of the proposed approach with confidence. Although video attacks are excluded in the tests it is expected that within the proposed challenge response framework they would be difficult to mount due to the need for synchronisation with the challenge sequence. Future work will expand the experiments to include a larger database of users and will also explore incorporation of additional features for improving the anti-spoofing capabilities of the system in response to more sophisticated attacks. In particular the relative position of eye centres within the face will be a subject of further study.

References

[ADH12] Ali, A.; Deravi, F.; Hoque, S.: Liveness detection using gaze collinearity. In Proc of 3rd Intl Conference on Emerging Security Technologies, Lisbon, Portugal. Pages 62-65, Sept. 2012.

[BL97] Blum, A. L.; Langley, P.: Selection of relevant features and examples in machine learning. Artificial intelligence, 97(1):245-271. December 1997.

[DG11] Deravi, F., Guness, S. P., (2011). Gaze Trajectory as a Biometric Modality. In Proceedings of the BIOSIGNALS Conference, Rome, Italy. Pages 335-341, January 2011.

[FW03] Frischholz, R. W.; Werner, A.: 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. Pages 234-235, 2003.

[JJY06] Jee, H. K.; Jung, S. U.; Yoo, J. H.: Liveness detection for embedded face recognition system. International Journal of Biological and Medical Sciences, 1(4):235-238, 2006.

[Ki12] Kim, G.; Eum, S.; Suhr, J. K.; Kim, D. I., Park, K. R.; Kim, J.: Face liveness detection based on texture and frequency analyses. 5th IAPR International Conference on Biometrics (ICB), New Delhi, India. Pages 67-72, 2012.

[Ki98] Kittler, J.; Hatef, M.; Duin, R. P. W.; Matas, J.: On combining classifiers., IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(3):226-239, March 1998.

[Ko09] Kollreider, K.; Fronthaler, H.; Bigun, J.: Non-intrusive liveness detection by face images. Image and Vision Computing, 27(3):233–244, 2009.

[Ko08] Kollreider, K.; Fronthaler, H.; Bigun, J.: 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.

[Ko05] Kollreider, K.; Fronthaler, H.; Bigun, J.: Evaluating liveness by face images and the structure tensor. in Proc of 4th IEEE Workshop on Automatic Identification Advanced Technologies, Washington DC, USA. Pages 75 80, October 2005.

[Ko07] Kollreider, K.; Fronthaler, H.; Faraj, M. I.; Bigun, J.: Real-Time Face Detection and Motion Analysis with application in 'Liveness' Assessment. IEEE Transaction on Information Forensics and Security, 2(3): 548-558, 2007.

[Li04] Li, J.; Wang, Y.; Tan, T.; Jain, A. K.: Live face detection based on the analysis of Fourier spectra. in Proc of Biometric Technology for Human Identification, Orlando, FL, USA. (SPIE 5404), pages 296-303, April 2004.

[MN08] Milborrow, S.; Nicolls, F.: Locating facial features with an extended active shape model. In Proc. of the 10th European Conference on Computer Vision (ECCV), Marseille, France. October 2008.

[Pi12] Pinto, A. D. S.; Pedrini, H.; Schwartz, W.; Rocha A.: Video-Based Face Spoofing Detection through Visual Rhythm Analysis. 25th SIBGRAPI Conference on Graphics, Patterns and Images, Ouro Perto, Brazil. 2012.

[PMR11] Peixoto, B.; Michelassi C.; Rocha, A.: Face liveness detection under bad illumination conditions. 18th IEEE Intl Conf on Image Processing (ICIP), Brussels, Belgium. pp. 3557-3560, 2011.

[Pa11] Pan, G.; Sun, L.; Wu, Z.; Wang, Y.: Monocular camera-based face liveness detection bycombining eyeblink and scene context. Telecommunication Systems. 47(3-4):215-225, August 2011.

[PWL07] Pan, G.; Lin S.; Wu, Z..; Lao, S.: Eyeblink-based anti-spoofing in face recognition from a generic webcamera. in Proc. of 11th IEEE Intl Conf. on Computer Vision, Rio de Janeiro, Brazil 2007. Pages 1-8.

[Tr11] Tronci, R.; Muntoni M.; Fadda, G.; Pili, M.; Sirena, N.; Murgia, G,; Ristori M.; Roli, F.: Fusion of multiple clues for photo-attack detection in face recognition systems. International Joint Conference on Biometrics, (IJCB), pages 1-6, October 2011.

[WDF09] Wang, L,; Ding, X.; Fang, C.: Face Live Detection Method Based on PhysiologicalMotion Analysis. Tsinghua Science & Technology, 14(6):685-690, 2009.