Biometric Presentation Attack Detection for Mobile Devices Using Gaze Information

A Thesis Submitted to the University of Kent
For the Degree of Doctor of Philosophy
In Electronic Engineering

By

Nawal Alsufyani

September 2018

Abstract

Facial recognition systems are among the most widely deployed in biometric applications. However, such systems are vulnerable to presentation attacks (spoofing), where a person tries to disguise as someone else by mimicking their biometric data and thereby gaining access to the system. Significant research attention has been directed toward developing robust strategies for detecting such attacks and thus assuring the security of these systems in real-world applications. This thesis is focused on presentation attack detection for face recognition systems using a gaze tracking approach.

The proposed challenge-response presentation attack detection system assesses the gaze of the user in response to a randomly moving stimulus on the screen. The user is required to track the moving stimulus with their gaze with natural head/eye movements. If the response is adequately similar to the challenge, the access attempt is seen as genuine. The attack scenarios considered in this work included the use of hand held displayed photos, 2D masks, and 3D masks. Due to the nature of the proposed challenge-response approaches for presentation attack detection, none of the existing public databases were appropriate and a new database has been collected. The Kent Gaze Dynamics Database (KGDD) consists of 2,400 sets of genuine and attack-based presentation attempts collected from 80 participants. The use of a mobile device were simulated on a desktop PC for two possible geometries corresponding to mobile phone and tablet devices. Three different types of challenge trajectories were used in this data collection exercise.

A number of novel gaze-based features were explored to develop the presentation attack detection algorithm. Initial experiments using the KGDD provided an encouraging indication of the potential of the proposed system for attack detection. In order to explore the feasibility of the scheme on a real hand held device, another database, the Mobile KGDD (MKGDD), was collected from 30 participants using a single mobile device (Google Nexus 6), to test the proposed features.

Comprehensive experimental analysis has been performed on the two collected databases for each of the proposed features. Performance evaluation results indicate that the proposed gaze-based features are effective in discriminating between genuine and presentation attack attempts.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors Prof Farzin Deravi and Dr Sanaul Hoque for the advice, encouragement and continuous supports received from them during my PhD study. Prof Deravi and Dr Hoque gave me this precious opportunity to pursue PhD under their supervision. Their guidance helped me in all the time of research and writing of this thesis.

I express grateful thank to the University of Kent. I would also like to thank Taif University in the Kingdom of Saudi Arabia for providing the funding which allowed me to undertake this study.

I would like to express appreciation to Dr Asad Ali who gave me tremendous help and guidance for my research. I have enormously benefited from our constructive and productive collaboration. I am also deeply appreciated other collaborators, colleagues and persons who have directly or indirectly helped me academically and personally.

Some special words of gratitude go to my friend, Rawabi Alsedais, who has always been a major source of support when things would get a bit discouraging. Thank you Rawabi for always being there for me.

I would like to express my deepest gratitude to my parents. This thesis would not have been possible without their warm love and endless support. I would also like to thank my sister, Majeedah, and my brothers, Musfer, Abdurrahman, Khaled, Majed and Dr Abdulmajeed, for their unconditional love, support and encouragement.

I have to thank my husband and love of my life, Dr Hamed Alsufyani, for keeping things going and for always showing how proud he is of me.

The last word goes for my sons, Hatim and Malik, who have been the light of my life and who have given me the extra strength and motivation to get things done. This thesis is dedicated to them.

Content

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Meaning
1D	One Dimensional
2D	Two Dimensional
3D	Three Dimensional
3DMAD	Three Dimensional Mask Attack Detection
ACER	Average Classification Error Rate
APCER	Attack Presentation Classification Error Rate
BPCER	Bona Fide Presentation Classification Error Rate
BSIF	Binarized Statistical Image Features.
CF	Colour Frequency
CNN	Convolutional Neural Networks
CoALBPs	Co-occurrence of Adjacent LBPs
CRF	Conditional Random Field
DFT	Discrete Fourier Transforms
DHFD	Dynamic High Frequency Descriptor
DMD	Dynamic Mode Decomposition
DoF	Depth of Field
DoG	Difference of Gaussian
EER	Equal Error Rate
FCN	Fully Convolutional Network
FFT	Fast Fourier Transform
FNR	False Negative Rate
FPR	False Positive Rate
GLCM	Gray Level Co-Occurrence Matrix
GLOH	Gradient Location And Orientation Histogram
GMM	Gaussian Mixture Model

GUI	Graphical User Interface
HFD	High Frequency Component
HMM	Hidden Markov Model
HOG	Histogram of Oriented Gradients.
HOOF	Oriented Optical Flows
HSC	Histograms of Shearlet Coefficients
HSV	Hue Saturation Value
HTER	Half Total Error Rate
IBG	International Biometric Group
ICA	Independent Component Analysis
ICAO	International Civil Aviation Organization
IDA	Image Distortion Analysis
IQA	Image Quality Assessment
IST LLFFSD	IST Lenslet Light Field Face Spoofing Database
KDA	Kernel Discriminant Analysis
KGDD	Kent Gaze Dynamics Database
LAPM	Modified Laplacian
LBP	Local Binary Pattern
LBP-TOP	LBP From Three Orthogonal Planes
LBPV	LBP Variance
LFALBP	light field angular LBP
LLR	Linear Logistic Regression
LPQ	Local Phase Quantization.
LSPs	Local Speed Patterns
MBSIF-TOP	Multiscale Dynamic Binaryized Statistical Image Features
MCT	Modified Census Transform
MKGDD	Mobile Kent Gaze Dynamics Database
MLPQTOP	Multiscale Dynamic Local Phase Quantization
MMD	Maximum Mean Discrepancy
PAD	Presentation Attack Detection
PAI	Presentation Attack Instrument
PLS	Partial Least Square

RGB	Red, Green and Blue
ROC	Receiver Operating Characteristics.
SAFE	Secure Authentication with Face and Eyes
SID	Scale-Invariant Descriptor
SML	Sum Modified Laplacian
SMQT	Successive Mean Quantization Transform
SNoW	Sparse Network of Winnows
SR-KDA	Spectral Regression Based on Kernel Discriminant Analysis
SVM	Support Vector Machine
SVM-RBF	Support Vector Machine-Radial Basis Function
TNR	True Negative Rate
TPR	True Positive Rate
TV	Total Variation
UVAD	Unicamp Video-Attack Database

Glossary

Term	Definition
Biometrics	The technology of confirming a person’s identity on the basis of their behavioural or physical characteristics.
Presentation attack	An impostor claims another user’s identity by copying their biometric trait/s and presenting it in front of a biometric recognition system.
Presentation attack detection	Systems or techniques for detecting presentation attacks.
Liveness detection	Indicate that presentation attacks do not manifest signs of liveness in the scene.
User	Any person who uses the verification system.
Real access and genuine access	Denote samples from a valid user who provides the correct identity to the verification system.
Presentation attack and spoofing attack	Used when artefacts are presented to the biometric system at sensor level to subvert its normal operation.
Presentation attack instrument	The biometric object used in a presentation attack.

List of publications

The research in this thesis have produced the following publications:

Articles in conference proceedings

1. Nawal Alsufyani, Asad Ali, Sanaul Hoque, and Farzin Deravi, "Biometric presentation attack detection using gaze alignment," in *IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA)*, 2018, pp. 1-8.

2. Asad Ali, Nawal Alsufyani, Sanaul Hoque, and Farzin Deravi, "Biometric Counter-spoofing for Mobile Devices using Gaze Information", In *7th International Conference on Pattern Recognition and Machine Intelligence*, Springer, 2017, pp. 11-18.

Chapter 1

Introduction

1.1 Biometric Systems

In contemporary society, there is an increasing need for secure large-scale systems that can accurately determine an individual’s identity in several contexts [1] such as online banking and border control. Traditionally, knowledge-based security such as a password or PIN (‘what you know’) or token-based security like an ID card or physical key (‘what you have’) have been used to validate an individual’s identity for access to secure systems. However, ID cards can be easily lost, shared, manipulated or stolen by an impostor to gain unauthorised access; simple passwords are easy for an impostor to guess, and difficult passwords may be hard for the authorized user to remember. Reliable identity management systems with robust authentication techniques based on ‘who you are’—that is, biometrics—help to overcome these difficulties [2].

Biometrics is the technology of confirming a person’s identity on the basis of their behavioural or physical characteristics rather than depending on ID cards or passwords [1–3]. Behavioural characteristics are attributes that relate to individual behavioural patterns such as gait [3-5], voice [6, 7], keystroke [8], or signature [9-11]. Physiological characteristics are traits related to the human body—for example, fingerprint, iris, face, palm print, finger vein, DNA, or retina. For a biometric system, the selection of traits should satisfy the following requirements [2, 12, 13]:

- **Universality:** each person should have the trait;
- **Distinctiveness**: the trait should be sufficient to discriminate among different persons;
- **Permanence:** the trait should be sufficiently invariant, with little or no change over a period of time;
- **Collectability:** the trait should be capable of being captured and quantised with the appropriate devices;
- **Performance:** the trait should achieve the desired accuracy and speed;
- **Acceptability:** users should be willing to present the trait to the system;
- **Circumvention:** the trait should be robust to fraudulent methods that could fool the system (e.g. fake samples).

Depending on the context, a biometric system can operate in either *verification* mode or *identification* mode. In verification mode (‘Am I who I claim to be?’), a person’s claimed identity is validated by comparing the captured sample against a previously collected biometric sample for that person. This mode performs one-to-one matching to determine whether the claim is true or not. In identification mode (‘Who am I?’), the system must recognise the individual’s identity by comparing their biometrics against a database of previously collected samples of N individuals. This one-to-many comparison can establish a person’s identity without requiring that person to claim an identity [12].

The rest of this chapter discusses the motivation for this research and is organised as follows. Section 1.1 presents a brief introduction to biometric systems. Section 1.2 provides the security of biometric systems. The motivation for this research and objectives are listed in Section 1.3, supported by a system block diagram. Section 1.4 outlines the study's contributions, and Section 1.5 describes the overall thesis structure.

1.2 Security of Biometric Systems

Despite the many advantages of biometric systems, they remain vulnerable to a wide range of attacks [14]. Between acquisition of the biometric trait and the final decision, the overall security of a biometric system can be compromised at various points [15]. Eight possible attack points have been identified [16] (see Figure 1.1), and these can be broadly classified as one of two types: indirect or direct. *Indirect attacks* are performed inside the biometric system by impostors such as cyber-criminal hackers, who may attempt to attack the feature extractor or the matcher (items 3 and 5 in Figure 1.1) or to alter the stored templates (item 6 in Figure 1.1). They might also attempt to replay digitally stored biometric data (item 2 in Figure 1.1) or exploit possible weaknesses in the communication channels to change the content of templates before they reach the matcher (items 4, 7, and 8 in Figure 1.1).

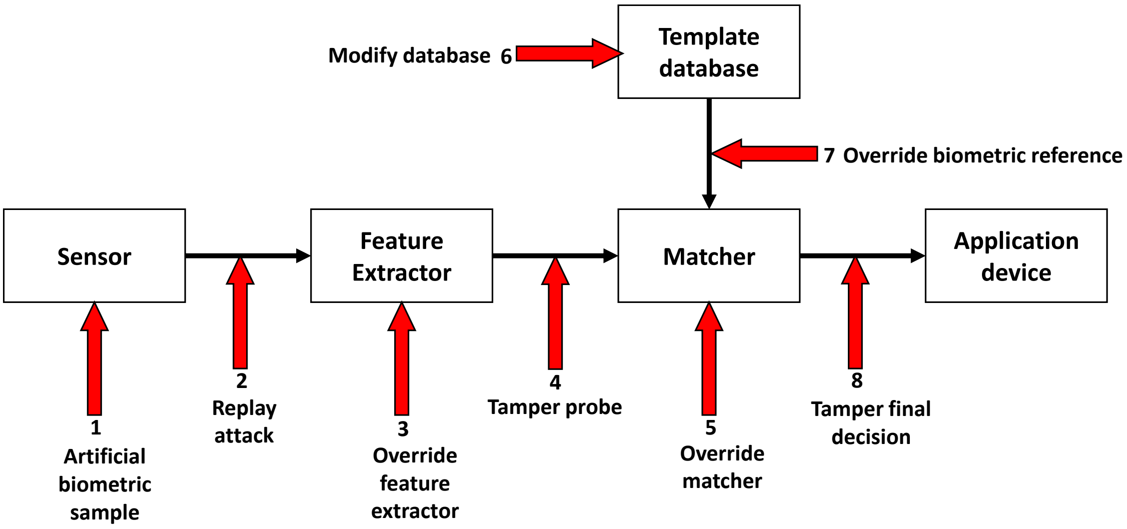


Figure 1.1: Possible attack points in a generic biometric system (after Ratha et al. [18])

Direct attacks (also known as *presentation attacks attacks* as defined in the current ISO/IEC 30107-3 standard [17]) are performed outside the biometric system, where the impostor presents synthetic biometric samples, (e.g. gummy finger, copy of a signature, or face mask) to the sensor (item 1 in Figure 1.1). In a presentation attack, an impostor claims another user's identity by copying their biometric trait/s and presenting it in front of a biometric recognition system [18]. Given the extensive deployment of such systems for security purposes, there is a need for ongoing research to guard against presentation attacks.

This study deals only with presentation attacks where the impostor requires no prior knowledge about the underlying working principles of the system, as end-users only have access to the sensor [8]. As any user can be considered a potential attacker, the most susceptible point for an attack is on presentation of the biometric trait [8]. Biometric traits (e.g. face, iris, fingerprint, DNA) are publicly available data and it is often easy to sample them; this is probably the most familiar drawback of biometric verification systems [19]. As it is not difficult to find step-by-step online tutorials on how to make artefacts such as face masks or artificial fingerprints, an artificial biometric sample can readily be produced. Because presentation attacks are performed outside the biometric system itself, manufacturers cannot implement digital protection mechanisms to prevent them. It follows that robust countermeasures are needed to deal with fake biometric samples.

1.3 Motivation and Objectives

Among biometric traits, the face modality has significant potential, as almost all consumer-level computing devices such as smartphones, tablets, and laptops are equipped with a front-facing camera [20]. According to the International Biometric Group (IBG), face is the second most-used modality after fingerprint [21] and is accepted in many large-scale applications, including International Civil Aviation Organization (ICAO)-compliant biometric passports [22] and various national ID cards. Face-based authentication and unlocking features have also been introduced on many smartphones (e.g. Face Unlock on Android devices) [11]. Some companies rely on the face modality for payment systems in supervised environments and on smartphone devices. However, mimicking face modality is straightforward as compared to other biometric modalities such as iris and fingerprint, as no special skills are required to create artefacts of the targeted face. As a face is normally visible, it is easy to capture a high-quality sample of the targeted face, with or without the individual’s cooperation. A genuine user’s facial image can easily be captured without their knowledge using distant cameras. Moreover, social image sharing and social networking websites make the facial images of many users accessible to the public. This means that an impostor can obtain images from a social network or similar source and present them to a biometric authentication system to gain access. Additionally, production of almost photo-realistic 3D masks has become affordable, and wearable masks manufactured from just two photographs of a person’s face can now be purchased in online stores [13].

Presentation attack detection (PAD) techniques can substantially improve the security of face recognition systems [23]. Along with high recognition performance, a practical face recognition system requires anti-spoofing capability [24].

The algorithms used in this study employ a challenge-response method to detect attacks on face recognition systems in mobile devices by recording the user’s gaze in response to a moving stimulus. Using facial landmarks, gaze-based features are extracted and analysed to determine whether or not the captured images are acquired from a genuine user. The proposed features are based on the assumption that spatial and temporal coordination of the eye movements following the challenge are significantly different when a genuine attempt is made as compared to attack attempts. The movements of the head and the gaze will be different when the imposter is attempting to follow the challenge by holding the photo. The hand movements are delayed until the eye is available for guiding the movement.

Many techniques have been proposed in the literature to examine the vulnerabilities of face recognition systems to direct attack, and multiple approaches have been proposed to secure them against this threat. However, investigating novel features that can detect all attack types is a challenging task.

The study objectives are as follows.

- Design a system using the challenge-response technique in which a stimulus moves in different trajectories and the camera captures the gaze of the user while following the moving stimulus.

- Investigate three presentation attack instruments, including displayed photo, 2D mask, and 3D mask.

- Collect two databases to evaluate the challenge-response system, the first one simulates mobile device use with two possible geometries corresponding to mobile phone and tablet formats, and the second database uses a mobile device to explore the feasibility of using gaze-based features on a real handheld device.

- Evaluate several gaze-based features for face-PAD system.

1.4 Contributions

This thesis has four main contributions which are summarised as follows.

First, a challenge-response software is designed for the proposed system (Chapter 3). The challenge is presented on a display screen to a user as a visual stimulus following three different randomized trajectories: Lines, Curves, and Points. The user is instructed to follow the challenge as it changes its location with their gaze through natural head/eye movements and the camera (sensor) captures the facial images at varying positions of the stimulus on the screen.

Second, a new database, Kent Gaze Dynamic Database (KGDD), was collected from 80 participants (Chapter 3). This database consists of 2,400 sets of genuine and attack attempts. The participants are assumed to be using a mobile device for biometric authentication with two possible screen sizes. The KGDD contains gaze information and associated evaluation protocols with three types of attack instruments: displayed photo, 2D mask, and 3D mask.

Third, three gaze-based features are exploited for face presentation attack detection (Chapter 4). The proposed features are based on the assumption that the eye\head movements following the challenge are significantly different when a genuine attempt is made compared with certain types of attack attempts. Features are extracted from facial images captured at each location of the challenge. The experimental results demonstrate that the proposed features are effective in detecting presentation attacks.

Fourth, a new database, Mobile Kent Gaze Dynamic Database (MKGDD), using a mobile device is collected from 30 participants (Chapter 5). An android application is developed to display the challenge and collect gaze information from the participants. The data collection application is loaded onto a Google Nexus 6. Data are collected from genuine attempts where users are tracking a moving visual target (‘challenge’) with natural head/eye movements and impostor attacks where users are holding a displayed photo, looking through a 2D mask, or holding a 3D mask of a genuine user and attempting to follow the stimulus. Several sets of experiments are conducted using the MKGDD to investigate the feasibility of the gaze-base features on a real handheld device.

1.5 Structure of the Thesis

The thesis is organised in six chapters, which can be briefly summarised as follows.

- **Chapter 1** includes a general introduction to biometric systems and their vulnerabilities. It also outlines the motivation, objectives, and the structure of the thesis.

- **Chapter 2** provides a comprehensive overview of face presentation attacks, including an introduction to face presentation attack instruments and a review of the relevant literature. Face presentation attacks are organised into two main categories and further divided into groups based on the cues used to distinguish between real access and spoofing attacks. The chapter also details publicly available databases used in the literature.

- **Chapter 3** describes the design and implementation of the proposed face presentation attack detection system, including further details of the hardware and software used to conduct the experiments. This chapter also provides details of the first database assembled to evaluate the system. Test protocols are provided, comprising several scenarios for thorough evaluation. The chapter also specifies the evaluation strategies used in this work [25].

- **Chapter 4** introduces three gaze-based features for face presentation attack detection: gaze correlation analysis, gaze alignment consistency, and gaze time dynamics features. Extensive experiments are conducted on these gaze-based features [26].

- **Chapter 5** details the second database for a mobile device and explores the feasibility of the system on a real handheld device. Results obtained from the experiments are presented and discussed.

- **Chapter 6** summarises the study’s contributions and outcomes and indicates possible directions for future work.

Chapter 2

Literature Review

2.1 Introduction

This chapter provides a comprehensive overview of face presentation attack, including an introduction to face presentation attack instruments. A literature review identifies two main categories of face PAD research, which are further subdivided on the basis of cues used to distinguish between real access and spoofing attacks. Finally, the chapter introduces benchmark datasets related to face presentation attack.

2.2 Face Presentation Attack Instruments

According to ISO/IEC 30107-1 [17], the *biometric object used in a presentation attack* is referred to as the Presentation Attack Instrument (PAI). The vast majority of face PAIs reported in the literature can be grouped into two categories of synthetic artefact (see Figure 2.1): i) 2D surfaces (e.g. photo, video) that have proved successful against 2D face recognition technologies and ii) 3D volumes (e.g. masks) that can be used to spoof 2D and 3D face systems [19]. To better understand the problem of face presentation attack and the rationale behind techniques for detecting these attacks, it is useful to take a closer look at the various types of face PAI and how they present to the sensor [27].

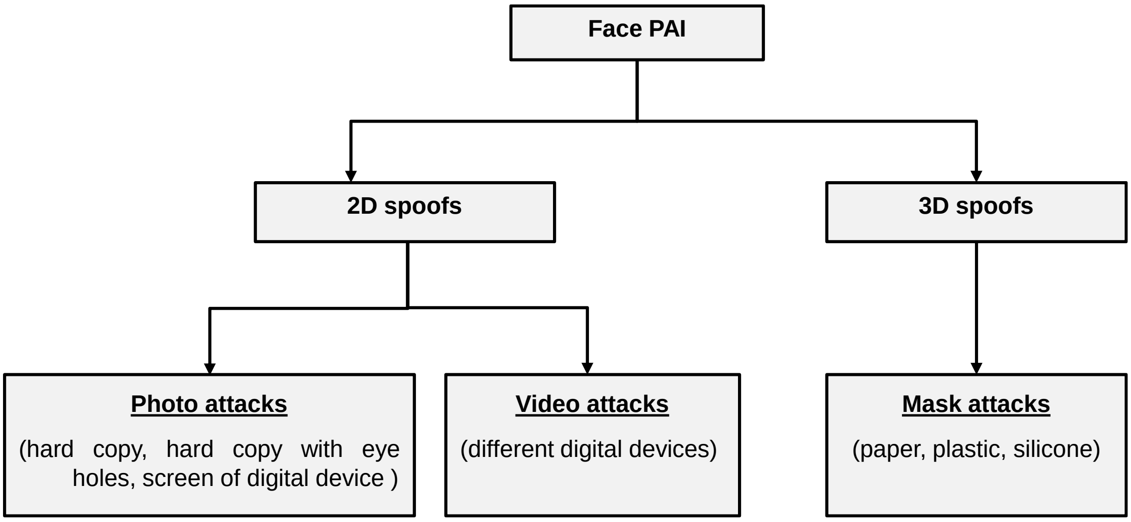

Face PAI
2D spoofs
3D spoofs
Photo attacks
(hard copy, hard copy with eye holes, screen of digital device)
Video attacks
(different digital devices)
Mask attacks
(paper, plastic, silicone)

Figure 2.1: General classification of face PAIs referred to in the literature

- **Photo Attacks**. The impostor presents a photograph of the genuine user to the recognition system. The photo may have been taken by the attacker using a digital camera or retrieved from the Internet after the genuine user uploaded it to one of the popular social networks [28]. The image can then be printed on paper or displayed on a digital device such as a mobile phone or tablet [29]. Motion analysis-based methods are very effective for detecting this type of attack, as photographs present only the appearance of the face. However, the attacker may try to simulate facial and head movements by rotating or warping a printed photo (see Figure 2.2) [3]. Another attack method involves photographic masks; these are high-resolution printed photos with eyes and mouth cut out. From behind the

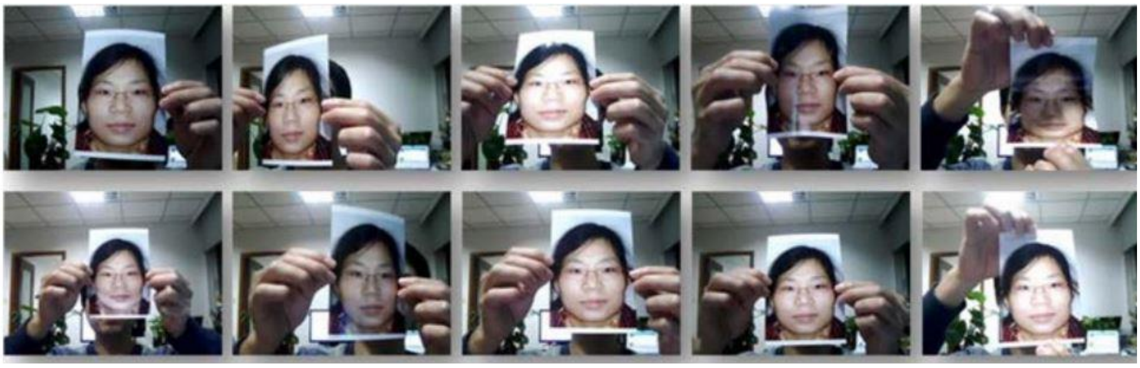


Figure 2.2: Attack samples from the NUAA Photo Imposter Database. Photo attacks are simulated by translating, rotating and bending (warping) a photograph

mask, the attacker attempts facial movements such as eye blinking and mouth movements [19].

- **Video Attacks.** These are sometimes called *replay attacks*. Using a digital device such as a laptop (see Figure 2.3), mobile phone or tablet, the impostor replays a video of the genuine user [30]. This type of attack is more advanced than the photo-based approach, simulating both the 2D texture of the face and its dynamics.

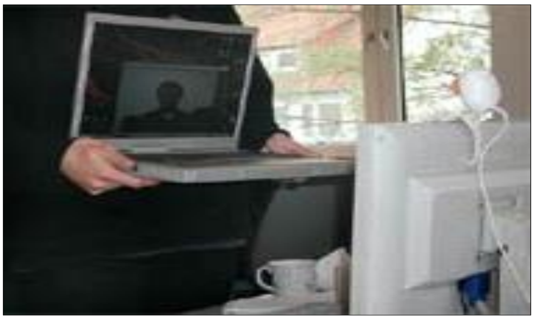


Figure 2.3: Example of a replay attack using a laptop

- **3D Mask Attacks.** Here, the attack artefact is a 3D mask of the genuine user's face (see Figure 2.4), so overcoming the depth cues that guard against the above attack types [31]. The new generation of affordable 3D acquisition sensors and dedicated scanning software and the affordability of 3D printing makes it increasingly feasible to manufacture 3D face masks [2].

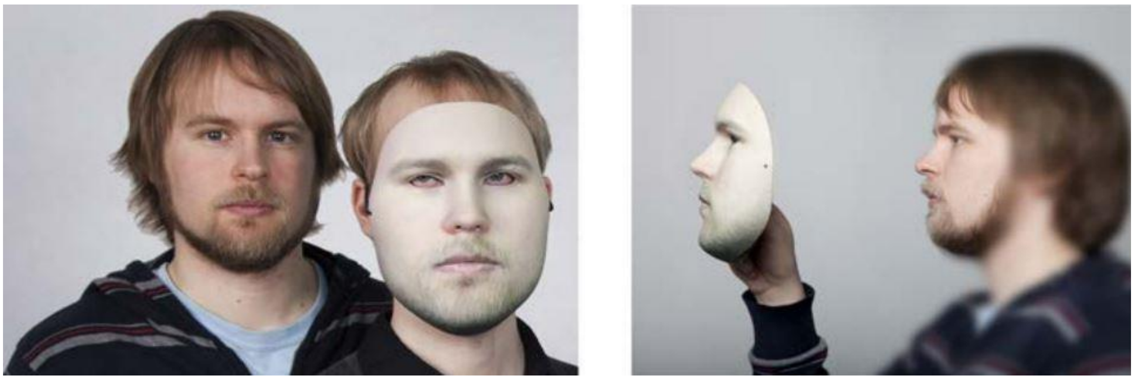


Figure 2.4: A wearable 3D mask ordered from http://www.thatsmyface.com/

2.3 Face Presentation Attack Detection Techniques

In the current state of the art, face PADs can be grouped into two categories: (1) passive approaches that do not require user cooperation or awareness and (2) active methods that require user engagement (see Figure 2.5) [2].

2.3.1 Passive Approaches

Passive methods of presentation attack detection do not require user cooperation or even awareness but instead exploit involuntary physical movements and 3D properties of the image. There are four main categories of passive face PAD, the first of which is based on texture and frequency analysis. The second uses motion analysis and cues from the scene. The third approach analyses 3D facial information and the presence or absence of depth features [2]. Finally, deep learning-based methods of PAD use deep neural networks to learn feature representations.

2.3.1.1 Texture and Frequency Analysis

Texture analysis is based on detectable patterns such as image blurring and printing defects [32]. These methods work on the assumption that fake faces are printed on paper and presented to the camera for verification. The paper and printing process produce texture features which can be distinguished from real facial images [33].

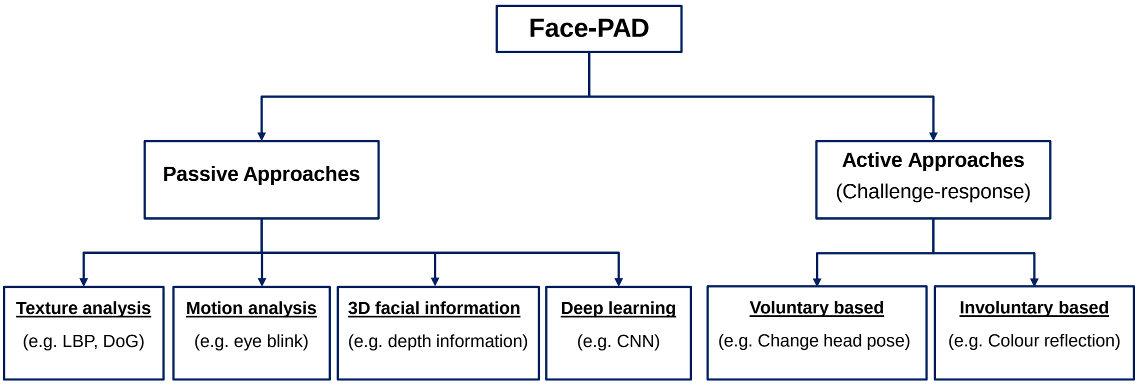


Figure 2.5: Categorisation of PAD methodologies referred to in the literature

This analysis uses image processing tools such as the Fourier Spectrum [34], multiple Difference of Gaussian (DoG) filters to extract frequency information [30, 35], partial least squares [36], and Local Binary Patterns (LBP) [32, 37] that are usually combined with other texture descriptors such as Gabor Wavelets, and shape-related information extracted using Histogram Oriented Gradients (HOG) [38]. For example, Määttä et al. [39] proposed an approach based on differences between images of a real face and a printed face, using texture-based (LBP and Gabor wavelet) and gradient-based (HOG) features to detect a presentation attack. LBP and Gabor filters are used to extract textural features (micro-texture pattern and macroscopic information), and local shape description using HOG provides additional facial information. Using this method, the face was first detected, cropped and normalised into an M × M pixel image. The facial images were then divided into several local regions, and the three descriptors (LBP, Gabor Wavelet and HOG) were extracted from each block. A homogeneous kernel map was applied to each resulting feature vector, and the resulting representation was inputted to a Support Vector Machine (SVM) classifier. Finally, weighted score level fusion was used to combine the outputs of individual SVMs to determine whether the input image corresponded to a live face (see Figure 2.6). Extensive experiments were performed on three publicly available databases containing several real and fake faces: the NUAA Photograph Imposter Database (NUAA) [40], the Yale Recaptured Database [41] and the IDIAP Research Institute’s PRINT-ATTACK Database (PRINT-ATTACK) [42].

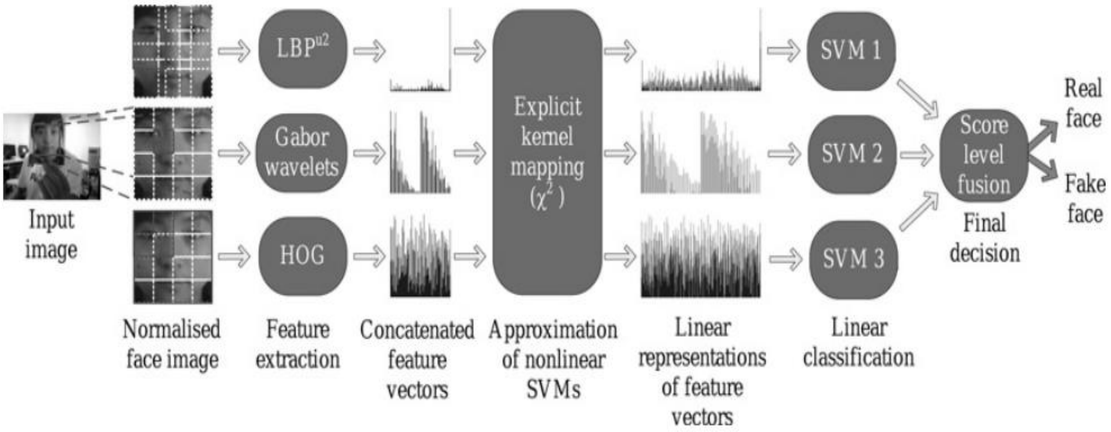


Figure 2.6: The proposed approach in [39]

In [43], multi-scale LBP helped to detect a face print attack. First, the face was detected, cropped and normalized. The LBP operator was then applied to the normalized face image, and the resulting LBP face image was divided into 3 x 3 overlapping regions. The local histograms from each region were computed and combined into a single histogram, and two other histograms were then computed from the whole face image using $LBP_{8,2}^{u2}$ and $LBP_{16,2}^{u2}$ operators. All histograms were concatenated to form the final enhanced feature histogram (see Figure 2.7). A nonlinear SVM classifier was used to distinguish between the live and fake face, and extensive experiments were conducted using the NUAA database.

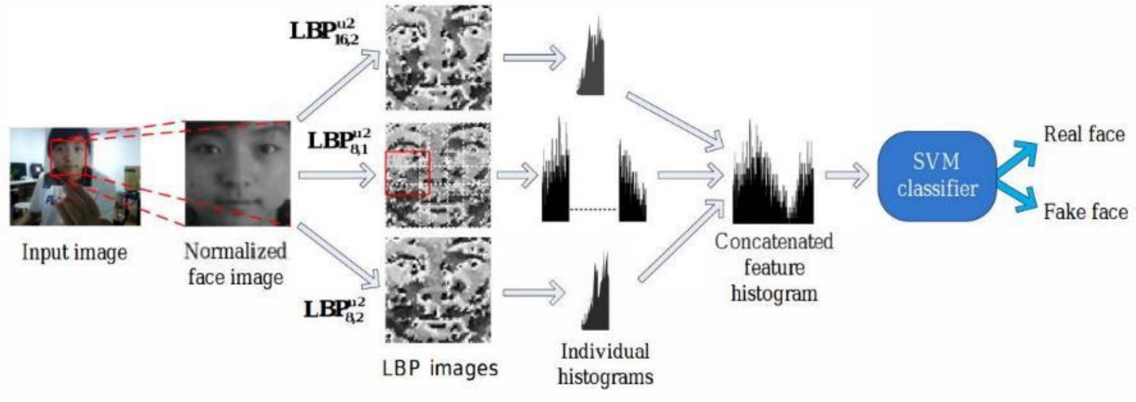


Figure 2.7: Block Diagram of the proposed approach in [43]

Schwartz et al. [36] proposed an anti-spoofing solution for photo-based attacks that used different properties of the face (texture, colour and shape) to obtain a holistic representation (see Figure 2.8). Using only the face region, they generated a feature vector formed by combining low-level feature descriptors for each frame of the video, such as HOG, Colour Frequency (CF) [44], Gray Level Co-occurrence Matrix (GLCM) [45], and Histograms of Shearlet Coefficients (HSC) [46]. The feature vectors were then combined into one feature vector containing rich spatiotemporal information about the biometric sample and were subjected to partial least square (PLS) [47] classification. The authors validated the method using the NUAA database. The experimental results confirmed that improvements were achieved by combining feature descriptors.

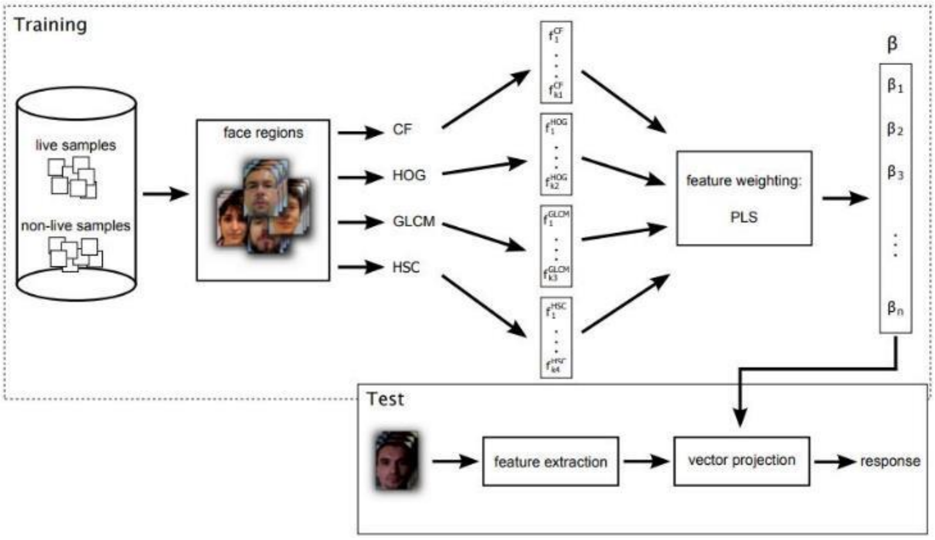


Figure 2.8: Block diagram of the proposed method in [36]

Pinto *et al.* [48] described an algorithm for video-based spoofing attack detection by analysing noise signatures generated by the video acquisition process. Using a Fourier transform to capture noise properties for each frame of the video, a compact representation called a visual rhythm was extracted to detect temporal information in the Fourier spectrum. Features were extracted from the visual rhythms using GLCM, LBP and HOG descriptors; the GLCM method proved the most discriminative and compact feature representation for this purpose. Figure 2.9 summarises the steps involved. SVM and PLS analyses were used to determine whether a sample was fake. The authors evaluated the proposed method using the Unicamp Video-Attack Database (UVAD).

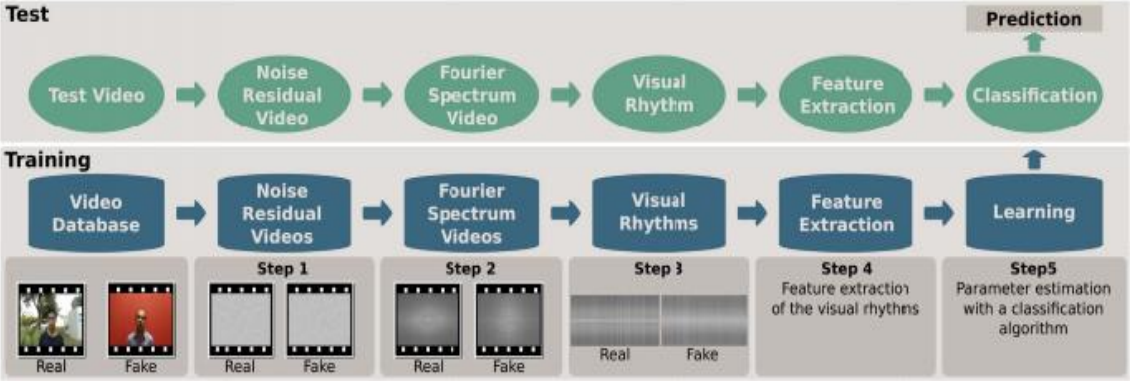


Figure 2.9: Proposed method using visual rhythm in [48]

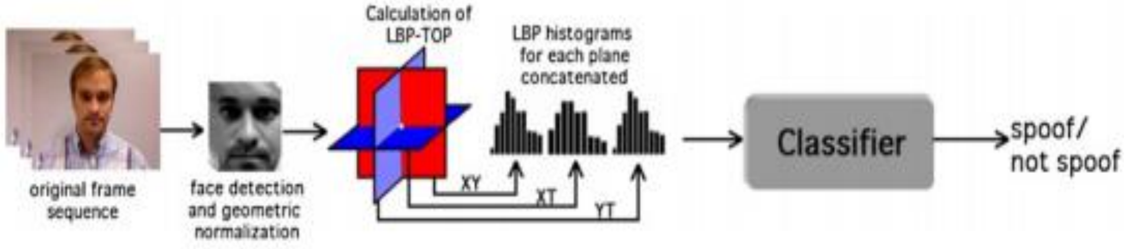

original frame sequence
face detection and geometric normalization
Calculation of LBP-TOP
LBP histograms for each plane concatenated
XY
XT
YT
Classifier
spoof/ not spoof

Figure 2.10: Proposed countermeasure using dynamic texture in [49]

Pereira et al. [49] proposed an anti-spoofing solution based on a dynamic texture—specifically, a spatiotemporal version of the original LBP. The authors explored the utility of LBP from three orthogonal planes (termed as LBP-TOP); the key idea was to detect the structure and dynamics of the microtextures that characterise real faces. LBP-TOP uses temporal information, computing LBP histograms in the XT and YT planes, along with spatial information in the XY plane. As shown in Figure 2.10, each frame of the original sequence was grey-scaled and passed through a face detector using modified census transform (MCT) features. Only detected faces of more than 50 pixels width and height were used, and these were geometrically normalized to 64 × 64 pixels. Following face detection, the LBP operators were applied to each plane (XY, XT and YT); the LBP histogram of each plane was computed, and the histograms were concatenated. After feature extraction, the feature vector was passed to a binary classifier to distinguish spoofing attacks from authorized access attempts. Using two publicly available databases to evaluate the approach—the IDIAP Research Institute's Replay-Attack Database (Replay-Attack) [32] and the CASIA Face Anti-Spoofing Database (CASIA-FASD) [30]—LBP-based dynamic texture description was found to be more effective than the original LBP.

In [50], the authors described a novel approach for detecting face liveness from a single image. Their key idea was that the difference in surface properties between fake and live faces can be efficiently estimated using diffusion speed (as shown in Figure 2.11). Specifically, they calculated diffusion speed by using a total variation (TV) flow scheme and extracting anti-spoofing features based on local patterns of diffusion speed, referred to as Local Speed Patterns (LSPs). These features were subsequently inputted to a linear SVM classifier to determine the liveness of the given face image.

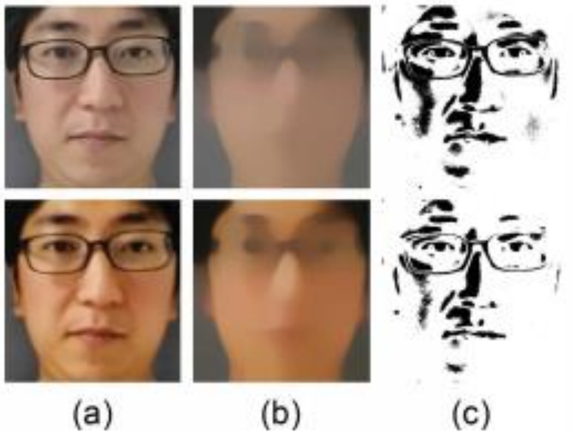


Figure 2.11: Diffusion speed maps for live faces (top) and fake faces (bottom): (a) original images; (b) diffused images; (c) binarized diffusion speed values [50]

The authors validated the method using three benchmark datasets: the NUAA dataset, the authors' own dataset of real-world scenarios in indoor and outdoor environments, and the Replay-Attack dataset.

In [51], the authors proposed a technique for face PAD using histograms of dynamic texture descriptors on three orthogonal planes. For this purpose, two effective spatiotemporal texture descriptors were used: histograms of multiscale dynamic binarized statistical image features (MBSIF-TOP) [52] and multiscale dynamic local phase quantization (MLPQTOP) [53, 54]. Having obtained the histograms of MBSIF-TOP and MLPQTOP, these were projected onto a subspace to separate genuine faces from attack attempts. This discriminative subspace was constructed using efficient kernel discriminant analysis based on spectral regression (SR-KDA); by avoiding costly eigen-analysis, this proved faster than ordinary KDA. To further improve system performance, the two kernels corresponding to the two representations were combined by means of a sum rule. The architecture of the proposed system for face spoofing detection is shown in Figure 2.12. The proposed system was evaluated on three publicly available databases: CASIA-FASD, Replay-Attack and the NUAA database.

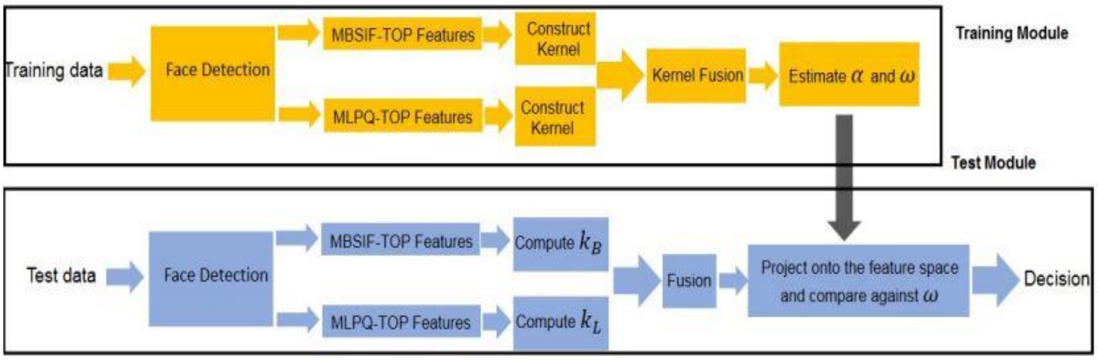


Figure 2.12: Architecture of the proposed system in [51] (where ω is projection of the mean, α is the coefficient, and K_B and K_L correspond to the kernel matrices constructed using the MBSIF-TOP and MLPQ-TOP descriptors, respectively).

In [55], the authors proposed an algorithm to search for Moiré patterns caused by the overlap of the digital grids. Figure 2.13 (a) shows a test image, and Figure 2.13 (b) is a photograph of (a). Figures 2.13 (c) and (d) show details of Figures 2.13 (a) and (b), respectively, demonstrating the patterns that occur after an image is recaptured from a screen. The detection of Moiré patterns at the spatial domain is very complex. However, in the frequency domain, the analysis can be simplified. Figures 2.13 (e) and (f) show the absolute values of the Discrete Fourier Transforms (DFT) of Figures 2.13 (a) and (b), respectively. Figure 2.13 (f) shows distinctive peaks at mid and high frequencies. These peaks resulted from the overlapping pixel grids of screen and camera. Based on the assumption that most of the energy in non-face spoofing images is at low frequencies while face spoofing images contain unusual peaks at higher frequencies, the algorithm worked as follows. A peak detector was applied to the absolute DFT value of each image of a detected face; where any strong peak was detected, the image was considered face-spoofing. The effectiveness of the proposed algorithm was verified by running tests on a database comprising 50 images of individuals under 13 different conditions, using images from the MIT-CBCL Face Recognition Database, the Extended Yale Face Database B and the Frontal Face dataset from the Computational Group at Caltech [56] [57] [58]. Results showed that, under the right conditions, face spoofing could be detected with great accuracy.

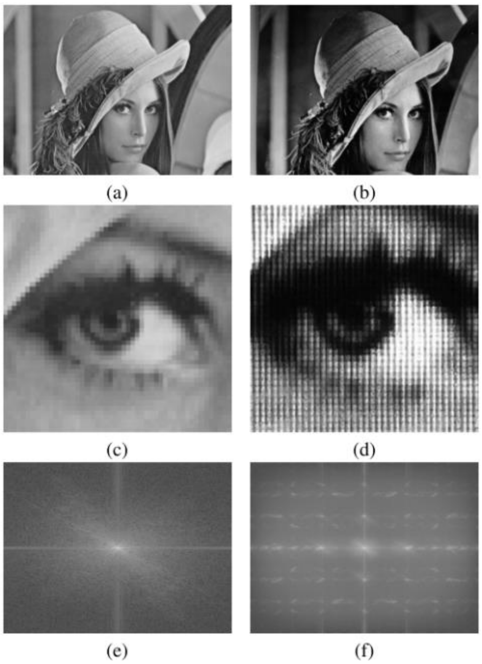


Figure 2.13: Example of Moiré patterns caused by the overlapping of digital grids: (a) portion of the Lena test image; (b) photograph of (a); (c)-(d): details of (a)-(b), respectively; (e)-(f): absolute values of the discrete Fourier transforms of (a)-(b), respectively, after logarithmic scaling for viewing purposes [55]

Jain et al. [59] proposed an efficient and robust approach to face spoof detection based on Image Distortion Analysis (IDA). Four features (specular reflection, blurriness, chromatic moment, and colour diversity) were extracted from the normalized face image and concatenated to generate the IDA vector. An ensemble classifier, comprising multiple SVM classifiers trained for different spoof attacks (e.g. printed photo, replayed video), was used to distinguish between real and fake faces. Using a voting scheme, the proposed method was extended to multi-frame face spoof detection in videos. Figure 2.14 shows the system diagram of the proposed IDA-based spoof detection algorithm.

The authors also assembled a face spoof database called the MSU Mobile Face Spoofing Database (MSU MFSD) that included three types of spoof attack (printed photo, replayed video with iPhone 5S and iPad Air), using two mobile devices (Google Nexus 5 and MacBook Air). Using two public domain databases (Replay-Attack and CASIA-FASD) and the MSU MFSD database, experimental results were reported for both intra-database and cross-database scenarios. The results confirmed high performance and highlighted the difficulty of separating genuine and spoof faces, especially in cross-database scenarios.

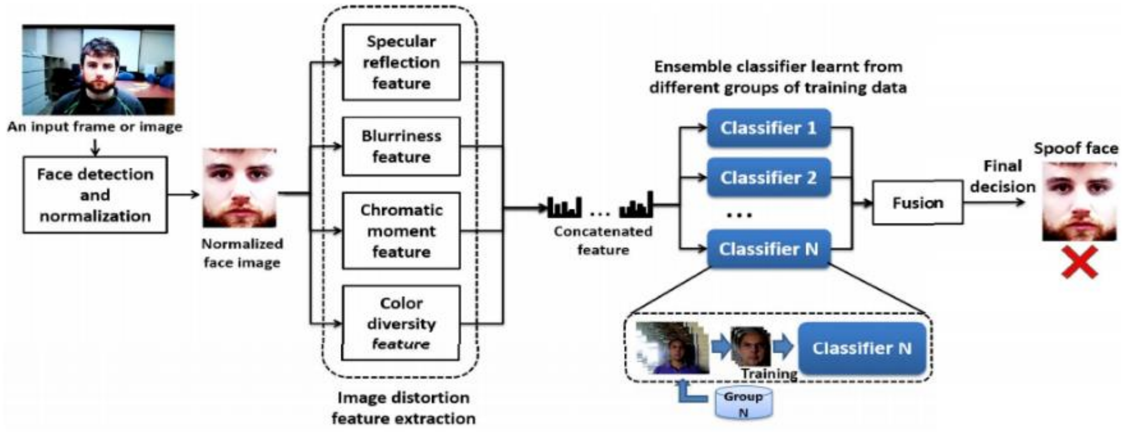


Figure 2.14: Proposed face spoof detection algorithm based on Image Distortion Analysis in [59]

In [60], the authors argued that texture features such as LBP, DoG and HOG are capable of differentiating real faces from artefacts in fake faces. As shown in Figure 2.15, they first expanded the detected face to obtain one so-called Holistic-Face (H-Face). This was then divided into six components, including contour, facial, left eye, right eye, mouth, and nose regions. The authors further divided the contour and facial regions into 2 × 2 grids. For all twelve components, low-level features such as LBP, Local Phase Quantization (LPQ), and HOG were extracted. Using the extracted local features, component-based coding was performed, based on local codes obtained from an offline trained codebook. The local codes were then concatenated into a high-level descriptor, with weights derived from Fisher criterion analysis. Finally, features were inputted to an SVM classifier. The proposed methods achieved significant success when using the NUAA database, the PRINT-ATTACK database, and CASIA-FASD.

Kim et al. [61] proposed a single image-based method of differentiating live faces from 2-D paper masks based on frequency [62] and texture analyses. They explored two key observations: 1) differences in 3D shapes lead to a difference in low frequency regions; and 2) differences between real and fake faces generate disparities in high frequency information. They opted to use texture information because the richness of texture details tends to diminish in printed faces. To extract frequency information, a 2D DFT was used to transform the facial image into a frequency domain, which they then divided into several groups of concentric rings, each representing a corresponding region in the

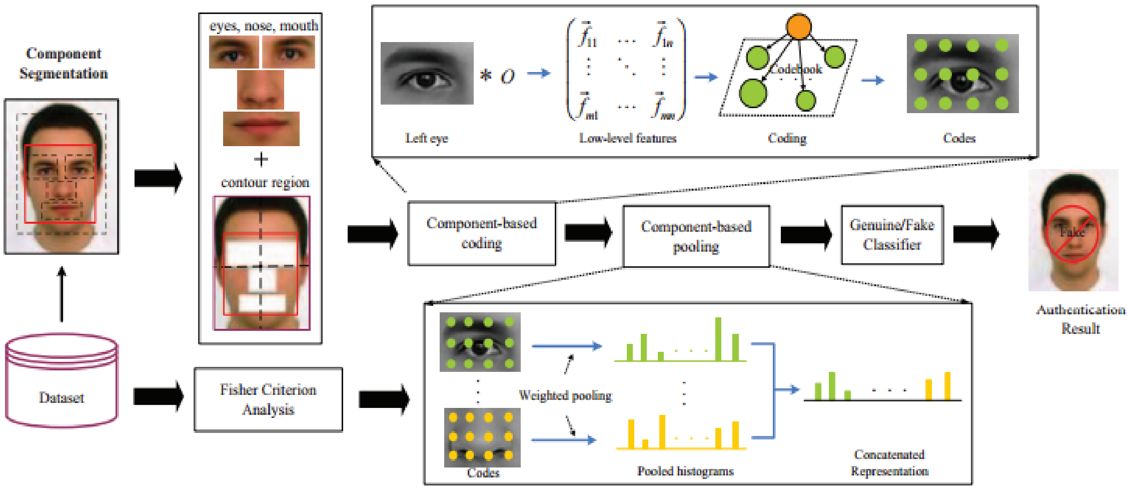


Figure 2.15: Flowchart of proposed framework (*O* represents a feature operator) in [60]

frequency band. Finally, a 1D feature vector was generated by combining the average energy values of all concentric rings. Figure 2.16 shows this frequency-based feature extraction process. Figure 2.16(c) shows the resulting frequency feature which was exploited for the classification. For texture-based feature extraction, the LBP was used to describe texture information. Figure 2.17 shows the LBP-based feature extraction process. Figure 2.17(c) shows the histogram of Figure 2.17(b) which was exploited as the feature vector for the classification. An SVM classifier was used for liveness detection. The decision value of SVM classifier trained by power spectrum-based feature vectors and that of SVM classifier trained by LBP-based feature vectors were combined to detect the fake. The experiments drew on two databases: the BERC Webcam Database and the BERC ATM Database. The images in the webcam database were captured under three different illumination conditions, and the fake faces were captured from printed paper and magazines.

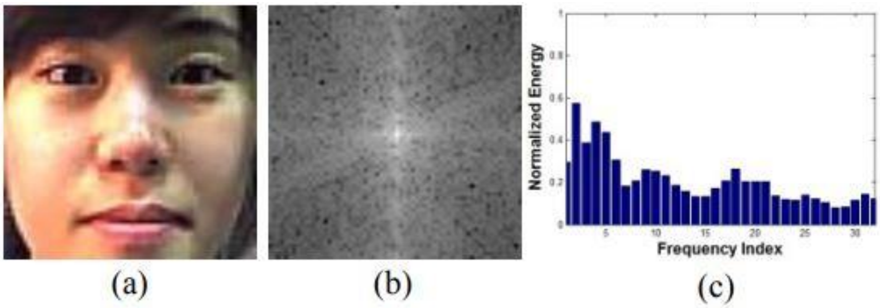


Figure 2.16: Frequency-based feature extraction: (a) original facial image; (b) log-scale magnitude of the Fourier-transformed image (power spectrum); (c) 1D frequency feature vector extracted from the normalized power spectrum [61]

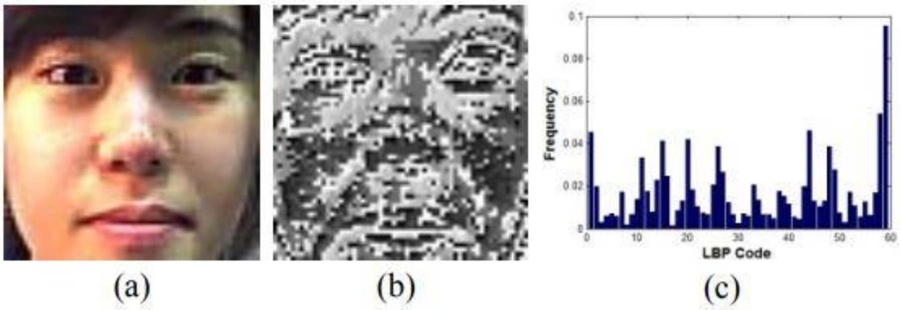


Figure 2.17: Feature vector extraction process based on LBP: (a) original facial image; (b) LBP-coded image; (c) histogram of the LBP-coded image [61]

Raghavendra et al. [63] described a PAD algorithm that explored both global (i.e. face) and local (i.e. periocular or eye) regions to accurately identify 2D and 3D face masks. Local and global features were extracted using Binarized Statistical Image Features (BSIF) [64] and LBP features. A linear SVM classifier was trained separately on local and global features, and scores were combined using the weighted sum rule to obtain comparison scores. Figure 2.18 shows the proposed method for identifying 2D and 3D face masks. The main purpose of extracting the eye region was to accurately capture any variations and discontinuities caused by the mask. The presence of a 2D or 3D mask would hide rich eye region details such as eyelashes and eyelids. In addition, a mask would introduce strong edges at the eye opening. Extensive experiments were carried out on two publicly available databases: CASIAFASD (2D) and 3DMAD (3D). Algorithm performance was specified in terms of Average Classification Error Rate (ACER), returning scores of 5.74% and 4.78% on CASIA-FASD and 3DMAD, respectively.

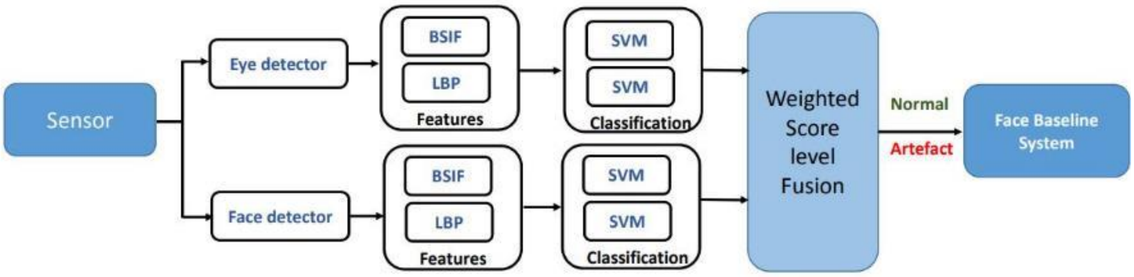


Figure 2.18: Block diagram of the proposed PAD scheme in [63]

In another study of image-based attacks, Lee et al. [65] proposed a frequency entropy analysis for spoofing detection. By splitting colour video of facial regions into RGB (red, green and blue) channels, they captured time sequences for each colour channel. To eliminate cross-channel image noise, three RGB sequences were then analysed by means of an Independent Component Analysis (ICA) algorithm [66]. The Fast Fourier Transform (FFT) was also applied to these signals to capture the power spectra of each RGB channel. Finally, the power spectra were verified by entropy calculation to validate liveness. The experimental results indicated better than 95% accuracy for a private collected data set of 21 participants (15 males and 6 females). The approach is illustrated in Figure 2.19.

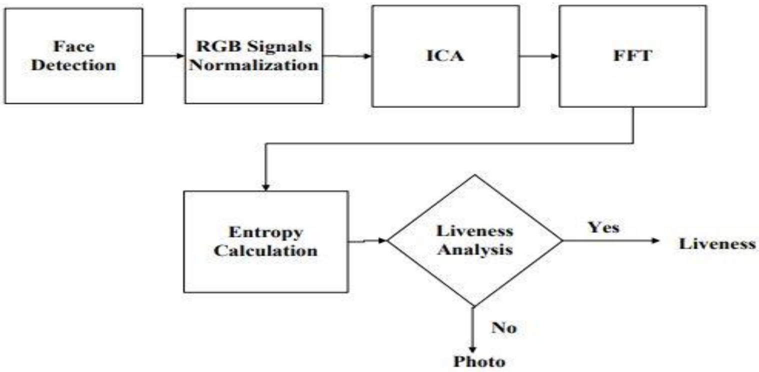


Figure 2.19: The proposed method in [65]

Lifang et al. [67] described another liveness detection scheme based on texture and colour analysis, combining Fourier statistics and LBP. After using face detection to obtain an input image, that image was pre-processed using gamma correction and DoG filtering to reduce illumination variation and to preserve key information. Next, LBP histogram and Fourier statistics were extracted to form the feature vectors. Finally, an SVM classifier was used to discriminate between live and fake faces. Figure 2.20 shows how LBP and Fourier spectrum features were combined using SVM. Using the NUAA database, the experimental results confirmed that the proposed scheme was efficient and robust, with an accuracy of 96.16%.

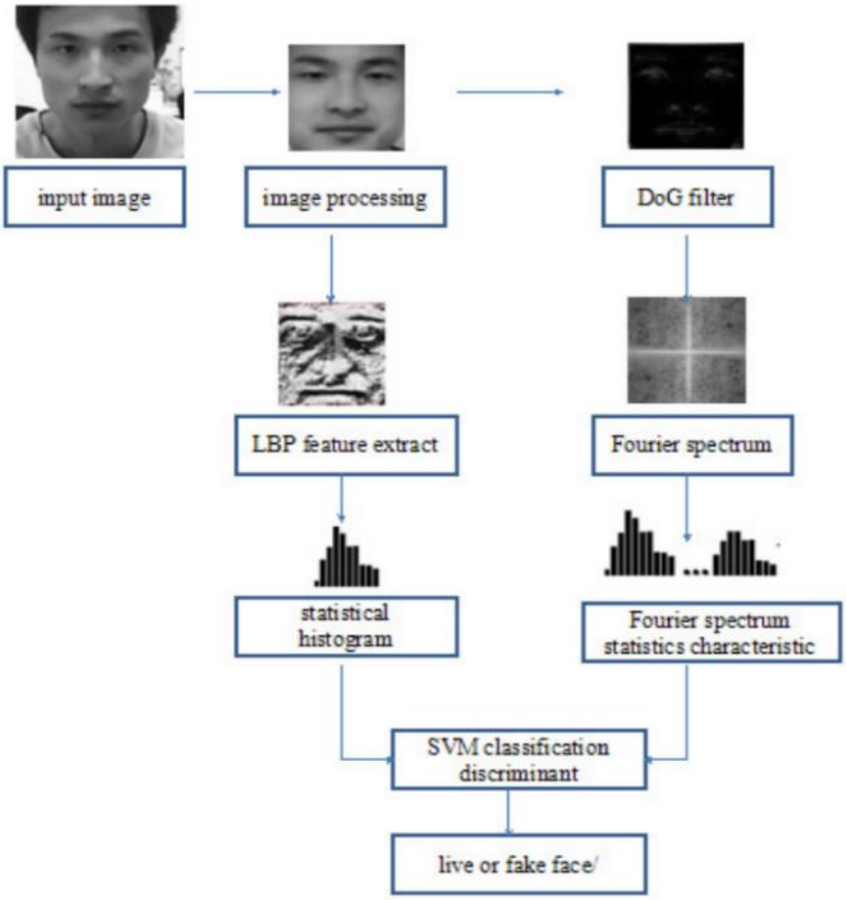


Figure 2.20: The proposed system using LBP and Fourier spectrum in [67]

Boulkenafet et al. [68, 69] reported a novel approach for detecting face presentation attack using colour texture analysis. Here, joint colour texture information from face image luminance and chrominance channels was exploited by extracting complementary low-level feature descriptions from different colour spaces (RGB, HSV and YCbCr). Colour texture was analysed using five descriptors: LBPs, Co-occurrence of Adjacent LBPs (CoALBPs), Local Phase Quantization (LPQ), Binarized Statistical Image Features (BSIF) and Scale-Invariant Descriptor (SID). The proposed algorithm is described in Figure 2.21. First, the face in the image was detected, cropped and normalised into an

$M \times N$ pixel image. Colour texture descriptions were then extracted from each colour channel, and the resulting feature vectors were concatenated into an enhanced feature vector. This final feature vector was passed to a linear SVM classifier, and the output score value determined whether the person in front of the sensor was real or fake. This approach was evaluated using the three latest face anti-spoofing databases (CASIA FASD, Replay-Attack Database and MSU MFSD) and returned excellent results (0.4% error rate on CASIA FASD).

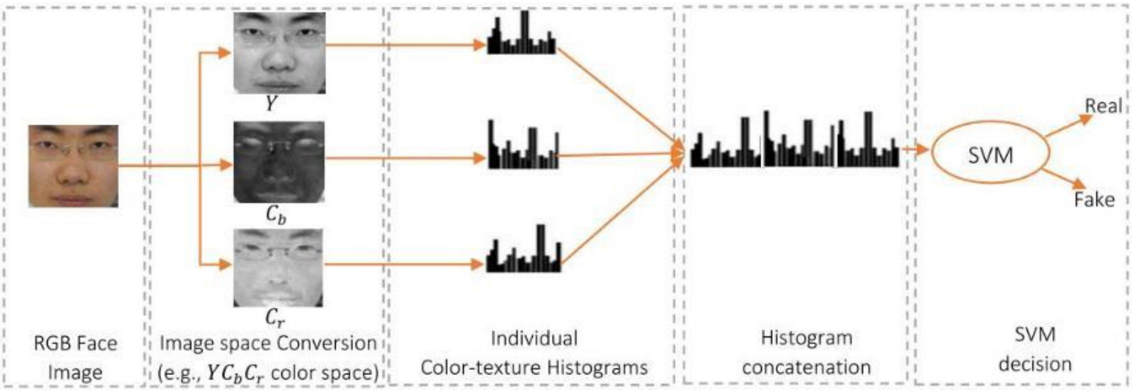

Y
C_b
C_r
Real
SVM
Fake
RGB Face Image
Image space Conversion (e.g., YC_bC_r color space)
Individual Color-texture Histograms
Histogram concatenation
SVM decision

Figure 2.21: The proposed face anti-spoofing approach in [68]

In [70], a new multiscale space was proposed for representation of face images prior to texture feature extraction, using three types of multiscale filtering: Gaussian scale space [71], DoG scale space and Multiscale Retinex [72]. The pipeline of the proposed approach is illustrated in Figure 2.22. The LBP descriptor was separately applied to each scale image, and the resulting histograms were concatenated to form the final feature vector. Finally, the face image texture representation was fed into a SVM classifier to determine whether the captured biometric sample was from a genuine or a fake face.

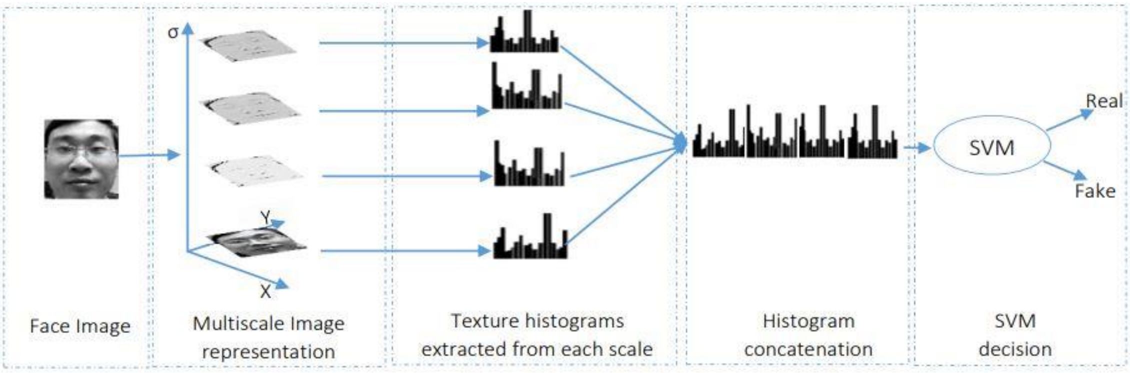

σ
Y
X
Face Image
Multiscale Image representation
Texture histograms extracted from each scale
Histogram concatenation
SVM
Real
Fake
SVM decision

Figure 2.22: The proposed approach in [70]

The availability of richer imaging sensors such as light-field cameras has created new possibilities for face PAD solutions [73], attracting increasing attention from the biometrics and forensics communities for both face recognition [74-76] and face PAD [77-80].

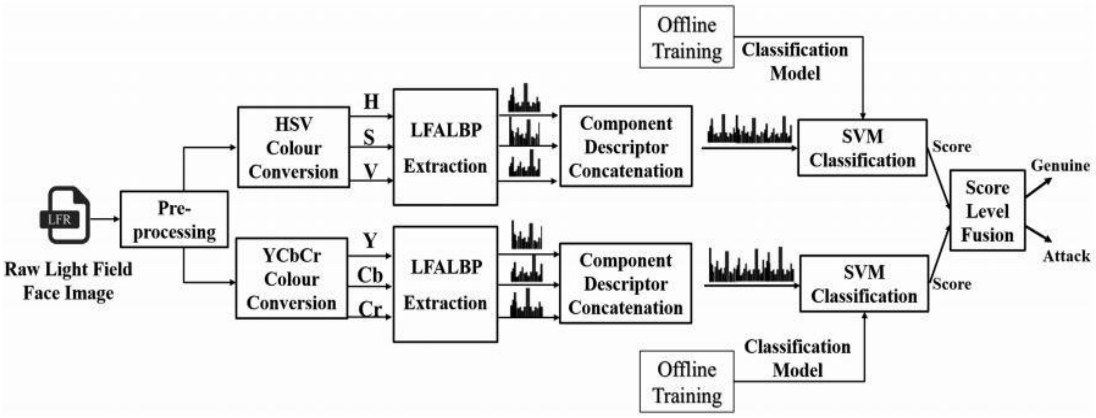


Figure 2.23: The proposed face anti-spoofing approach in [80]

More recently, Sepas-Moghaddam et al. [80] proposed light field angular LBP (LFALBP), a face spoofing detection solution based on an extension of the LBP descriptor, which captures disparity information in light-field images. The proposed algorithm is illustrated in Figure 2.23. Face images were first converted to the HSV and YCbCr colour spaces, and LFALBP histograms were extracted from each colour component. The extracted histograms were concatenated, resulting in a three-component descriptor vector for each colour space. The LFALBP descriptor vectors were then fed to the SVM classifier, and individual SVM classifier outputs for the two colour spaces were fused using score level fusion, applying a sum rule to determine whether the input image was a genuine face or an attempted attack. A database referred to as the IST Lenslet Light Field Face Spoofing Database (IST LLFFSD) was assembled, comprising 50 subjects captured with a Lytro ILLUM lenslet light-field camera. Extensive experiments were performed on the IST LLFFSD database.

Raghavendra et al. [81] described a PAD solution based on extraction of statistical features to capture micro-texture variation using Binarized Statistical Image Features (BSIF) [64] and Cepstral features reflecting micro changes in frequency using 2D Cepstrum analysis [82]. These features were fused to form a single feature vector, and a decision was derived using a linear SVM classifier. Figure 2.24 shows qualitative results for 2D Cepstrum features obtained from a real face. Extensive experiments were performed on the CASIA face spoof.

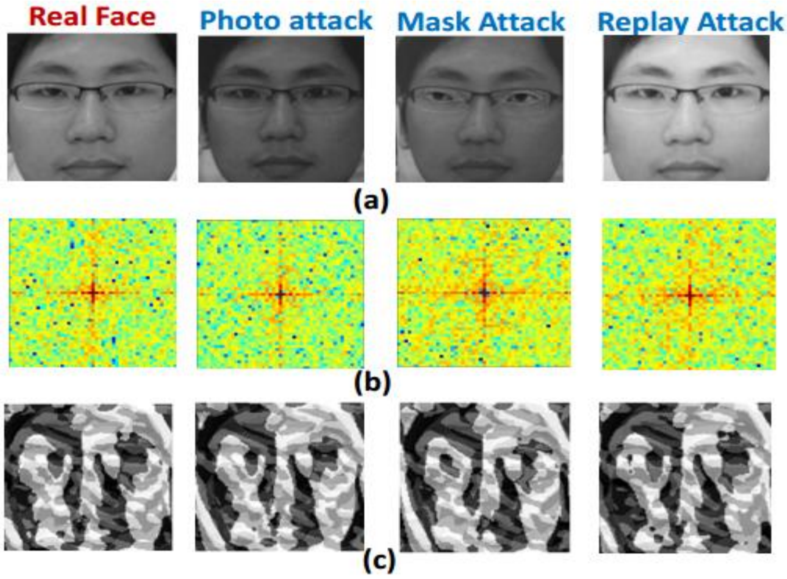


Figure 2.24: Qualitative results for the proposed PAD algorithm: (a) raw images; (b) 2D Cepstrum results; (c) BSIF results [81]

2.3.1.2 Motion Analysis and Cues from the Scene

In early work, motion analysis of the face region was used to detect liveness, and this approach remains popular in guarding against print attacks. Typical cues include eye blinking [83], mouth movement [84] and face and head gestures such as nodding, smiling and looking in different directions. Exploring the role of eye opening and closing in detecting attacks performed with print photos, Sun et al. [85] proposed an algorithm for face liveness detection that used an undirected conditional random field framework (CRF) to model blinking, represented by an image sequence including both open and closed eyes. As a half-open state proved difficult to define because of variations in eye size across individuals, two state labels were used: C for closed and NC for not-closed (including half-open and open) (see Figure 2.25). Tests were performed on a database comprising 80 videos of 20 individuals (four clips of each user). The first clip was a frontal view without glasses; the second clip was a frontal view with thinly rimmed glasses; the third clip was a frontal view with black framed glasses; and the last clip was an upward view without glasses. When compared to discriminative models like Adaboost [86] and generative models like the Hidden Markov Model (HMM) [87], the CRF model achieved a 98.3% imposter detection rate.

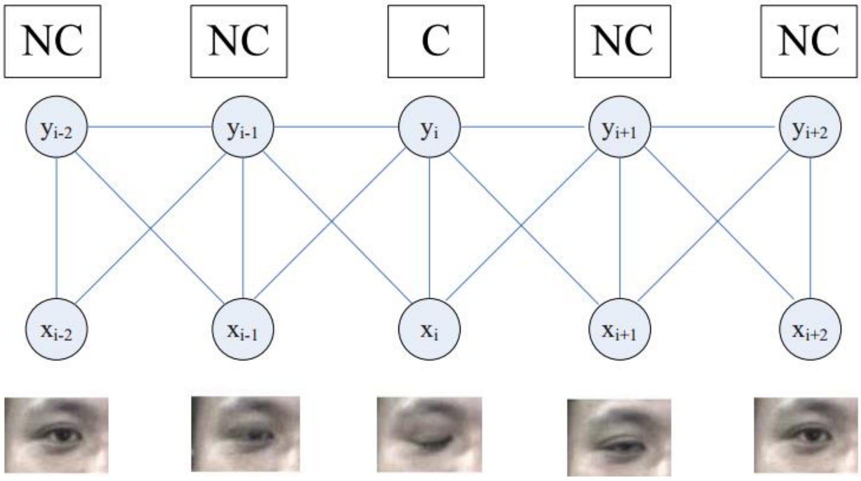


Figure 2.25: CRF-based blinking model, *C* for closed and *NC* for not-closed [85]

This work was further developed by Pan et al. [88]. As well as the main eye states (opening and closing), they included an ambiguous blinking state (from open to closed or closed to open). Temporal information was extracted from the eye-blink process as consecutive stages of open, half-closed and closed, followed by half-open and fully open. By defining this three-state set (α: open, γ: close, β: ambiguous), a typical blink can be described as the state change pattern α → β → γ → β → α as shown in Figure 2.26. The eye closeness values can be used as an effective features. The bigger the closeness value, the higher the degree of eye blinking. The database used in [42] was again employed to evaluate performance, comparing their method against cascaded Adaboost and HMM.

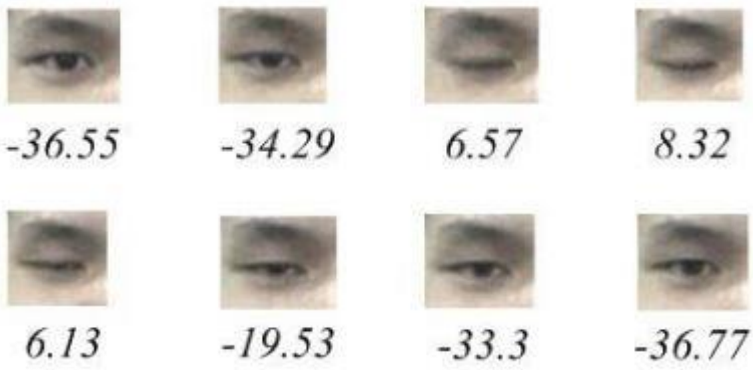


Figure 2.26: The blinking activity sequence, with closeness values below the corresponding frame (the higher the value, the higher the degree of closeness) [88]

In further investigations to enhance the face liveness detection system discussed in [85] and [88], Pan et al. [89] successfully combined eye-blink cues with contextual information from the scene to address spoofing involving photographs, videos and 3D models of a genuine user. Assuming that the face recognition system camera was fixed while operating, the first frame captured the scene without anyone in front of the camera. This frame was then used as the reference, as a spoofing video would show a different scene. Contextual cues were extracted from the right and left of the detected face region to calculate LBP around a set of key points, and the resulting histograms were compared to previously calculated reference patterns (see Figure 2.27). The eye blink prevented photo and 3D model spoofing, and the scene context was used to prevent video replay. This liveness detection system is illustrated in Figure 2.27. To evaluate the method, the authors built two data sets: a photo-impostor video database and a live video database. In extensive experiments combining contextual information and eye-blink cues, the liveness detection rate was 99.5% (195 out of 196).

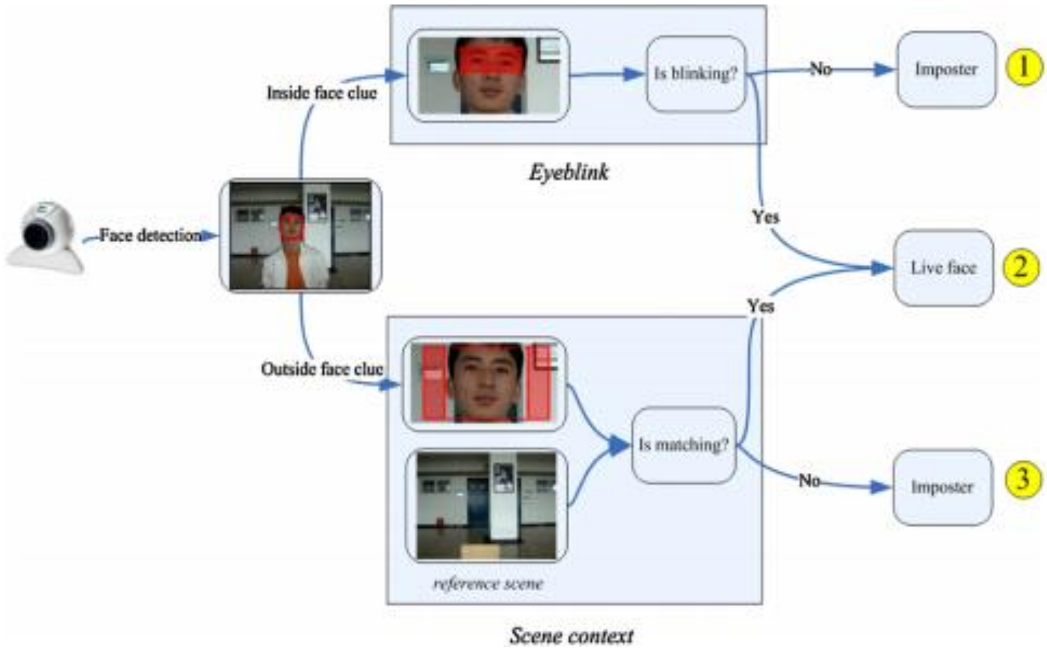


Figure 2.27: Illustration of liveness detection system using a combination of eye blinks and scene context in [89]

Yan et al. [90] introduced the technique of exploring multiple scenic cues, including non-rigid motion, face-background consistency and imaging banding. Non-rigid motion cues included facial motions such as blinking, using low-rank matrix decomposition-based image alignment to extract these features. The face-background consistency cue assumed that the motion of face and background was consistently high for fake images but low for genuine faces. A motion detection method based on the Gaussian Mixture Model (GMM) [91] was used to describe motion in the scene. The image-banding cue took account of defects in the quality of reproduced images, which were detected using wavelet decomposition [92]. Non-rigid motion analysis was used to capture liveness cues in facial expressions; face-background consistency analysis can capture liveness cues in videos with complex background while the non-rigid motion and the banding effect cues were used to capture the liveness cues in videos with a clean background. These three cues were integrated to determine whether the captured sample was real or fake. The method was validated using PRINT-ATTACK and other assembled image sets, achieving 100% accuracy on PRINT-ATTACK and high performance on the self-collected images.

Anjos et al. [42] proposed a motion-based algorithm that detects correlations between head movements and scene context. In the event of an attack, it should be possible to observe a high correlation between total movements in these two regions of interest (RoI). A measure of motion was calculated for each video frame, and a one-dimensional signal described by the extraction of five measures was used to form a feature vector. The method was validated using PRINT-ATTACK and achieved a 10% EER.

Tirunagari et al. [93] proposed a dynamic mode decomposition (DMD) method to model video content such as blinking eyes and moving lips. The classification pipeline consisted of DMD, LBPs and SVMs. Using DMD to represent temporal information from an image sequence as a single image, a single dynamic mode image was selected, corresponding to the eigenvalue with a phase angle equal to or closest to 0. LBP histogram features were then computed for the dynamic mode image. Finally, the LBP code was inputted to a trained SVM classifier. The effectiveness of this approach was demonstrated using three publicly available databases: 1) PRINT-ATTACK; 2) REPLAY-ATTACK and 3) CASIA-FASD, achieving comparable results on all of these. Figure 2.28 shows the proposed method.

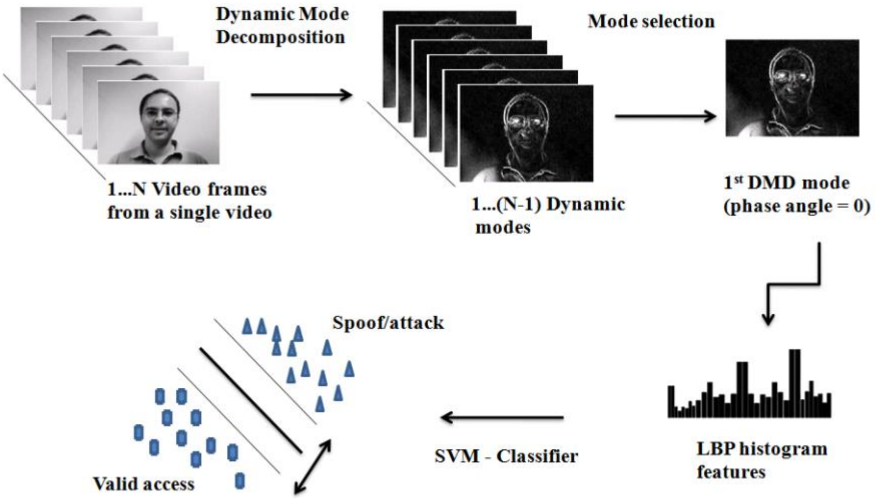


Figure 2.28: Steps of the methodological pipeline in [93]

Exploring the fusion of motion and microtexture, Komulainen et al. [94, 95] examined different visual cues and showed that individual methods can be significantly improved by performing fusion at score level. To combine motion and microtexture analysis techniques, the video sequences were divided into overlapping windows of *N* frames with an overlap of *N-1* frames. Each observation generated a score independent of the rest of the video sequence. For simplicity, LBP face description was calculated only for the last frame, with five frames extracted over the entire time course to evaluate motion correlation. Using linear logistic regression (LLR), fusion of the two visual cues was then performed at score level [96]. Using the REPLAY-ATTACK database, HTER of the best individual countermeasure declined from 11.2% to 5.1%. Figure 2.29 shows the fusion strategy.

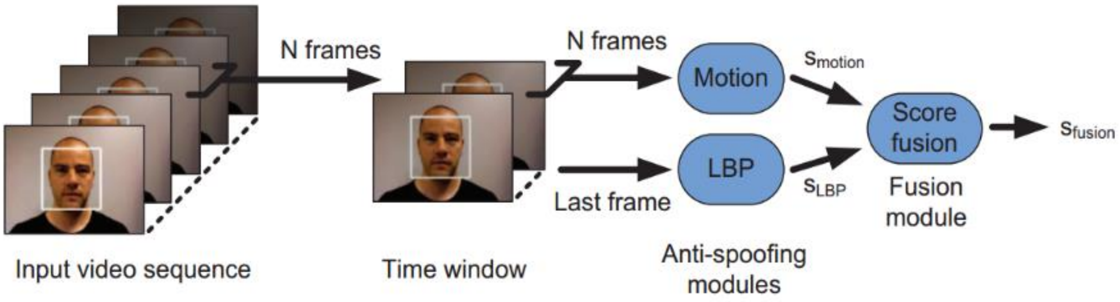


Figure 2.29: Block diagram of the used fusion strategy in [94], where *N* frames is the number of frames, S_{motion} and S_{LBP} are the scores of motion and LBP respectively, combining in score fusion S_{fusion}.

A new approach to spoofing detection in face videos in [97] used motion magnification supported by two powerful feature extraction techniques: Multi-scale LBP and Histogram of Oriented Optical Flows (HOOF) [98]. The proposed algorithm first used Eulerian motion magnification to enhance facial expressions in the captured video, and face spoof detection was then performed in two ways: (1) texture-based detection using motion magnification and (2) motion-based detection using motion magnification.

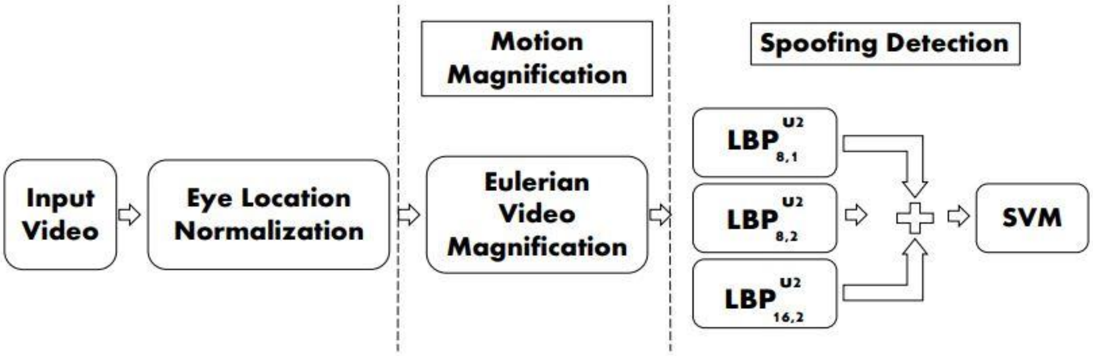


Figure 2.30: Proposed texture-based spoofing detection approach using motion magnification [97]

The motion magnification approach with LBP is shown in Figure 2.30. Here, eye location in the input video was normalized before applying Eulerian Motion Magnification to enhance facial movements. Using three LBP operators with different scales, three global histograms were then extracted and concatenated to generate a feature vector, which was inputted to a SVM classifier to distinguish between real and fake faces.

The second approach, using motion magnification and HOOF, is illustrated in Figure 2.31.

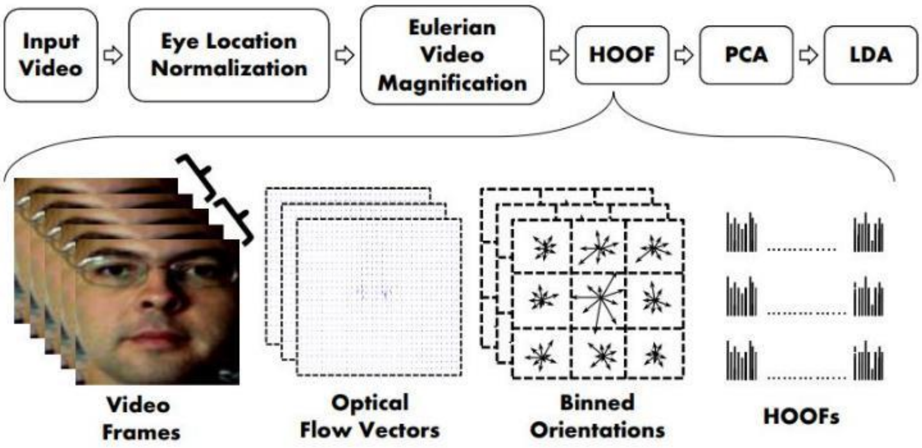


Figure 2.31: Motion magnification and HOOF [97]

Here, after applying Eulerian Motion Magnification to enhance facial movements in the video, optical flow between frames was calculated at a fixed interval. A histogram of optical flow orientation angle was then extracted across local blocks and concatenated to form a single vector (HOOF), using PCA to reduce the dimensionality of the feature vector. Finally, LDA was applied to the reduced feature vector, and classification was performed using thresholding and the nearest neighbour approach. The experiments were conducted on the PRINT-ATTACK and REPLAY-ATTACK databases.

In [99], the proposed face spoofing detection method was based on motion analysis and a cross-database voting strategy (see Figure 2.32). First, motion analysis was performed on optical flow and extracted motion information map (MIM) from neighbouring frames. Convolutional neural networks (CNN) were then applied to extract features, and each MIM was classified as real or fake. Finally, a cross-database voting threshold was adaptively generated to assign the corresponding video to the category of real or fake.

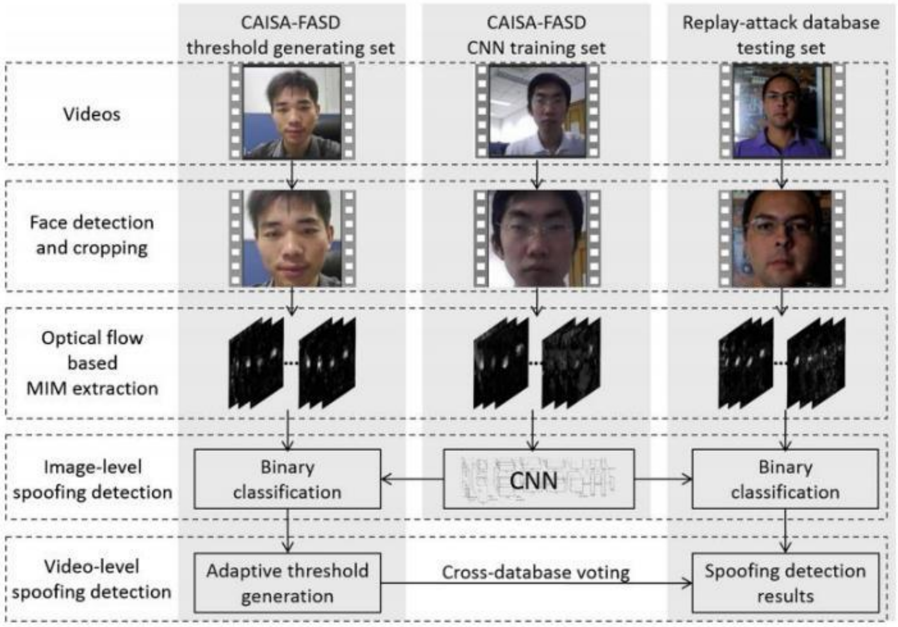

CAISA-FASD threshold generating set
CAISA-FASD CNN training set
Replay-attack database testing set
Videos
Face detection and cropping
Optical flow based MIM extraction
Image-level spoofing detection
Binary classification
CNN
Binary classification
Video-level spoofing detection
Adaptive threshold generation
Cross-database voting
Spoofing detection results

Figure 2.32: Proposed face spoofing detection method using motion analysis and CNN-based cross-database voting in [99]

To verify the effectiveness of the proposed algorithm in face spoofing detection, experiments were performed on publicly available databases, including REPLAY-ATTACK and CASIA-FASD.

2.3.1.3 3D Shape Information

The most obvious difference between a real face and a fake face is the presence or absence of depth information—in other words, human faces have curves while photos are flat, and researchers have exploited this feature to detect spoofing attacks. Lagorio et al. [100] proposed a novel liveness detection method based on 3D facial structure. They suggested that the proposed approach could be implemented in different scenarios with 2D or 3D face recognition systems for early detection of spoofing attacks. To decide if a face presented to the acquisition camera was human, 3D features of the captured face were computed. The findings showed that lack of surface variation and low surface curvature were key indicators of a 2D source. The mean curvature of the surface was computed to compare the two 3D scans using an approximation of the actual curvature value at each point, computed from the principal components of the Cartesian coordinates within a given neighbourhood. The mean curvature of 3D points on the face was then computed.

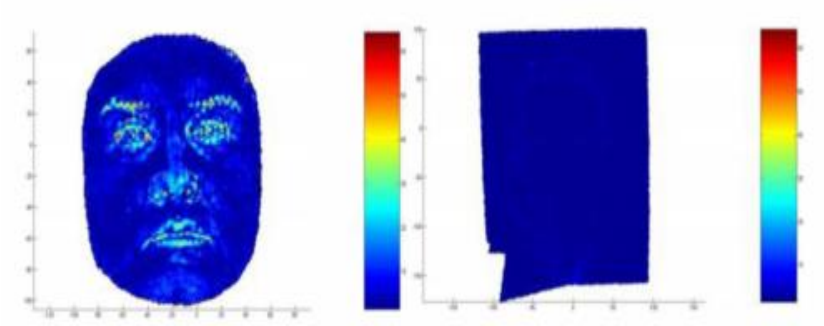


Figure 2.33: Curvature values based on 3D data captured from a real human face (left) and a printed picture of the same user (right) [100]

Experiments using a database assembled from 3D scans of real human faces and 2D pictures collected using the Vectra system confirmed the effectiveness and robustness of the proposed approach. Figure 2.33 compares 3D data acquired from a photograph and from a real face.

In another novel approach to face liveness detection, using 3D structure recovered from a single camera, Wang et al. [101] proposed to counter spoofing attacks by recovering sparse 3D facial structure. While genuine faces usually contain sufficient 3D structure information, photographs are usually planar. Using a face video or a sequence of images captured from more than two viewpoints, facial landmarks were detected, and key frames were then selected using a constrained local models (CLM) algorithm [102]. Sparse 3D facial structures were then recovered from the selected frames, and an SVM classifier was trained to differentiate the data. Using three databases employing cameras of differing quality, the approach achieved 100% accuracy in classification and face liveness detection. Examples of genuine and fake attempts are shown in Figure 2.34.

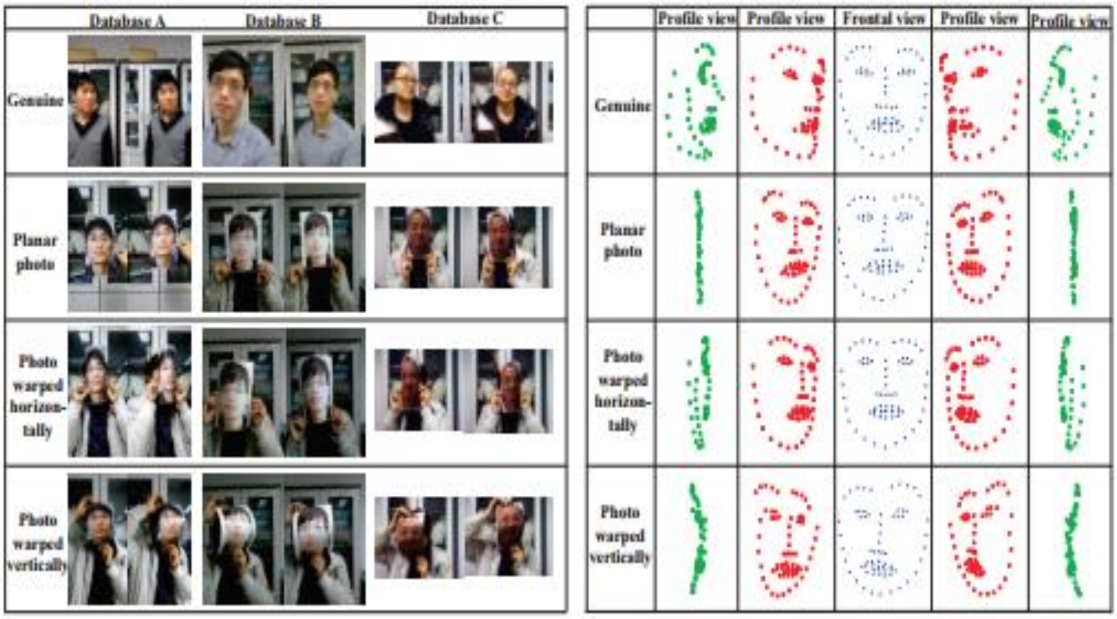


Figure 2.34: Examples of genuine and fake attempts [101]

Some methods use defocusing to estimate the depth in an image [103-106]. Depth of field (DoF) determines the degree of focus and range of focus between nearest and farthest objects in a given focal plane. Kim et al. [107, 108] implemented a focus-based method of face liveness detection. The key idea was to use the variation in pixel values between two sequential images that differed in focus (one of the camera’s functions). Assuming no great difference in movement, they tried to identify the difference in focal values between real and fake faces. In real faces, the focal regions were clear while others were blurred. However, in a printed copy of the face, there was little difference between images at different focuses (Figure 2.35). In two sequential pictures, the nose was closest to the camera lens while the ears were furthest away, with a sufficient difference in depth to register a 3D effect. To detect forged faces, features are extracted from normalized images; in the proposed method, three feature descriptors were extracted using modified Laplacian (LAPM) [109], power histogram and gradient location and orientation histogram (GLOH) [110]. These extracted features were concatenated and inputted to a SVM-radial basis function (SVM-RBF) [111]. Figure 2.36 shows this feature-level fusion approach using the assembled database. The results showed that when DoF was very small, false accept rate (FAR) was 2.86%, and false reject rate (FRR) was 0.00%; when DoF was large, however, average FAR and FRR increased.

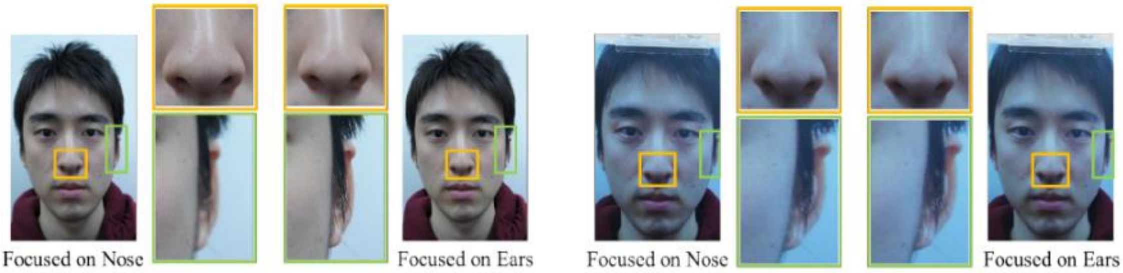


Figure 2.35: Partially focused images of (a) real and (b) fake faces [107]

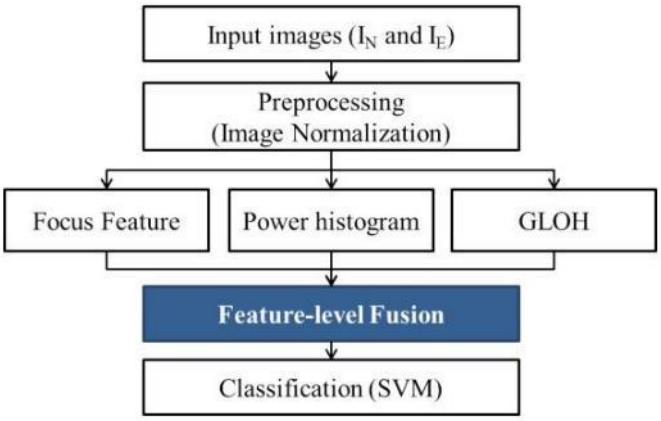


Figure 2.36: Feature-level fusion approach [107]

2.3.1.4 Deep Learning-based Methods

Recent work using CNN for biometric PAD purposes [112-116] has employed deep learning-based methods for feature representation [117]. This approach offers good generalization but depends on large and representative training databases [118]. Li et al. [119] proposed one such approach to PAD based on deep learning and domain generalization, using spatial and temporal information for feature representation. A 3D CNN network was first applied to extract spatial and temporal information, and generalization was regularized by minimising the Maximum Mean Discrepancy (MMD) among different domains to further improve generalization performance. Finally, SVM with linear kernel was used for classifier training and generation of a final detection result as shown in Figure 2.37. The experiments were performed on four databases: 1) Idiap REPLAY-ATTACK; 2) CASIA Face Anti Spoofing; 3) MSU mobile face spoofing database; and 4) RoseYoutu Face Liveness Detection database. The results show improved generalization ability as compared with state-of-the-art methods.

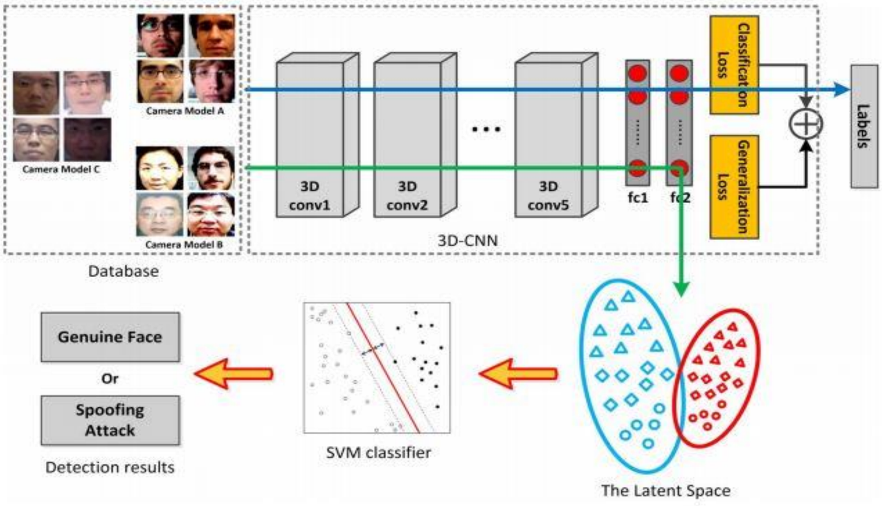


Figure 2.37: Pipeline of the proposed scheme in [119]

Li et al. [120] proposed a new form of anti-spoofing that leverages hybrid CNN for facial parts. First, the face was divided into several parts before training the corresponding CNN model for each part. The last layer of the hybrid model was concatenated to constitute the features, which were inputted to an SVM classifier. The main architecture of the hybrid CNN is shown in Figure 2.38. Effectiveness was tested using the REPLAY-ATTACK and CASIA databases and achieved satisfactory results when compared to state-of-the-art methods.

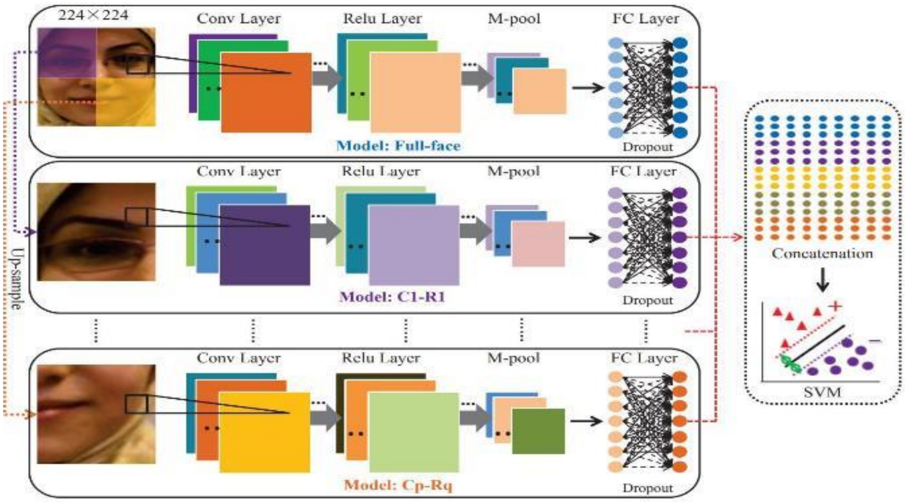


Figure 2.38: Primary architecture of the proposed hybrid CNN for face anti-spoofing [120]

Atoum et al. [121] proposed a two-stream CNN-based approach for face PAD, including patch-based CNN and depth-based CNN. Figure 2.39 shows both streams, along with a fusion strategy for combining them. For the patch-based stream, a deep neural network was trained end-to-end to learn rich appearance features, which were extracted from random patches within the face region. For the depth-based CNN stream, a fully convolutional network (FCN) was trained to estimate the depth of a face image by assuming that print or replay presentation attacks have a flat depth map while live faces are of normal depth. The patch-based CNN assigned a score to each patch randomly extracted from a face image, which was assigned the average score. The depth-based CNN provided a liveness score based on the face image's estimated depth map. Fusion of the scores for both CNNs yielded the final estimated class of live vs. spoof. Extensive experiments were conducted on three databases (CASIA-FASD, MSU-USSA, and REPLAY-ATTACK) for comparison to the state-of-the-art.

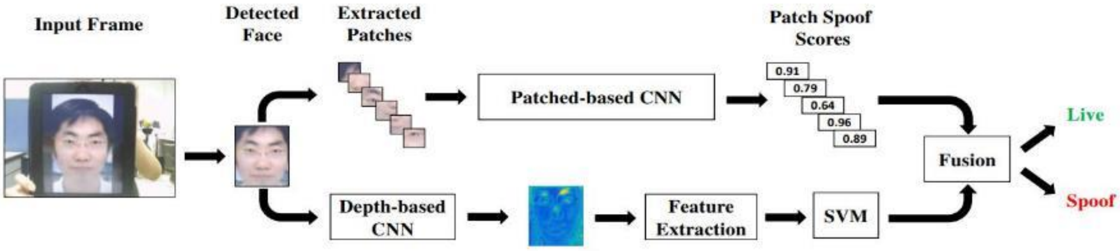


Figure 2.39: Architecture of the proposed face anti-spoofing approach in [121]

2.3.2 Active Approaches

Active anti-spoofing methods are often based on a challenge-response technique, where a system poses challenges that can only be overcome by a real person. Such systems can be assigned to one of two categories: voluntary (where a user is asked to respond to specific requests from a system) or involuntary (where the system checks liveness automatically).

2.3.2.1 Voluntary Methods

In methods of this type, the user is asked to perform specific activities to ascertain liveness, such as uttering digits [122] or altering their head pose [123-125]. For instance, Frischholz et al. [126] proposed a challenge-response method to enhance the security of their face recognition system, asking users to look in certain randomly chosen directions (see Figure 2.40). After estimating head pose, the system compared real-time movement (response) to the instructions asked by the system (challenge) in order to verify user authenticity. After responding to these challenges, the user was required to look straight into the camera, and the final image was captured for face recognition.

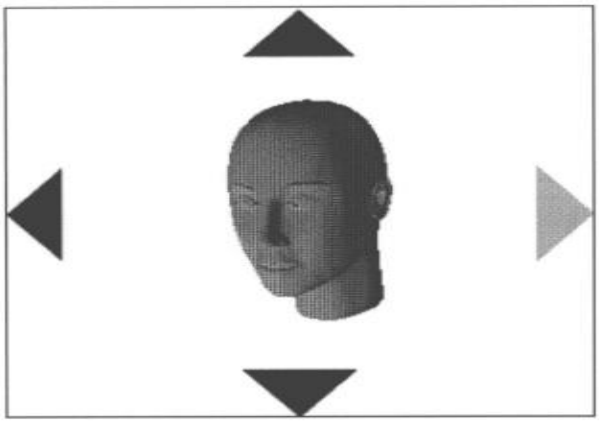
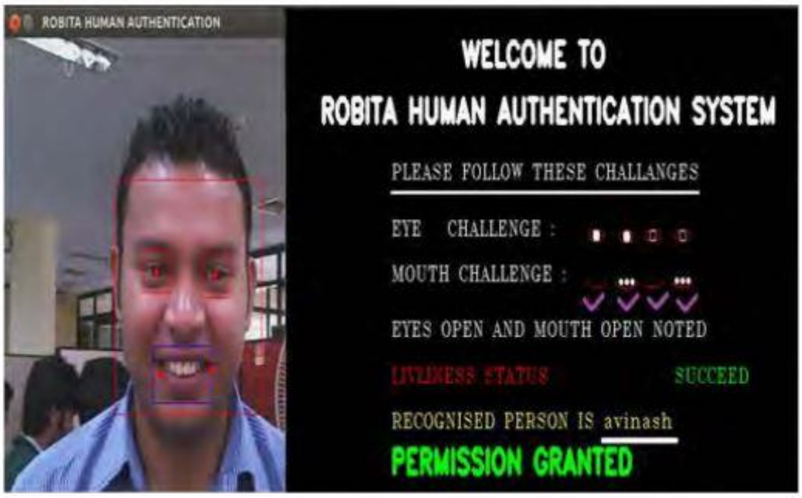


Figure 2.40: Random challenge directions [126]

Ali et al. [127-130] were first to report a system that used gaze as a cue for anti-spoofing. The user was required to follow a moving point shown randomly on the computer screen while their gaze was measured. The visual stimulus repeatedly directed the user’s gaze to certain points on the screen, and features extracted from images captured at these collocated points were used to estimate user liveness. The experiments were conducted using a collected database, and the results showed that the method was effective for face spoofing detection.

Singh et al. [131] proposed a liveness detection method based on a challenge-and-response scheme, using eye and mouth movements to generate random challenges that facilitated observation of the user's response. The challenges were generated randomly in such a way that only a live person could respond, and responses were extracted by counting eye and mouth movements. The system calculated facial movement by measuring the Hue Saturation Value (HSV) of the teeth. The presence of open or closed eyes was calculated by means of a search feature in the eye region, and the presence of an open or closed mouth was calculated by searching for teeth. If the sum of responses was equal to the threshold (the number of challenges posed by the system) then the system identified the person as real. The experimental setup was in two parts: a) generation of attack and b) liveness detection test. The authors designed five types of attack: 1) photo impostor; 2) eye impostor; 3) mouth impostor; 4) eye and mouth impostor; 5) video impostor. The system successfully prevented attacks in all these conditions, with no false rejections. To pass the liveness test, the attacker had to forge both eye and mouth regions of the genuine user, resulting in major facial changes. If the attacker passed the liveness test, the system’s face recognition module still blocked it. Figure 2.41 and Figure 2.42 show a spoof attack and a genuine attempt detection.



WELCOME TO
ROBITA HUMAN AUTHENTICATION SYSTEM

PLEASE FOLLOW THESE CHALLANGES

EYE CHALLENGE :

MOUTH CHALLENGE :

✓ ✓ ✓ ✓

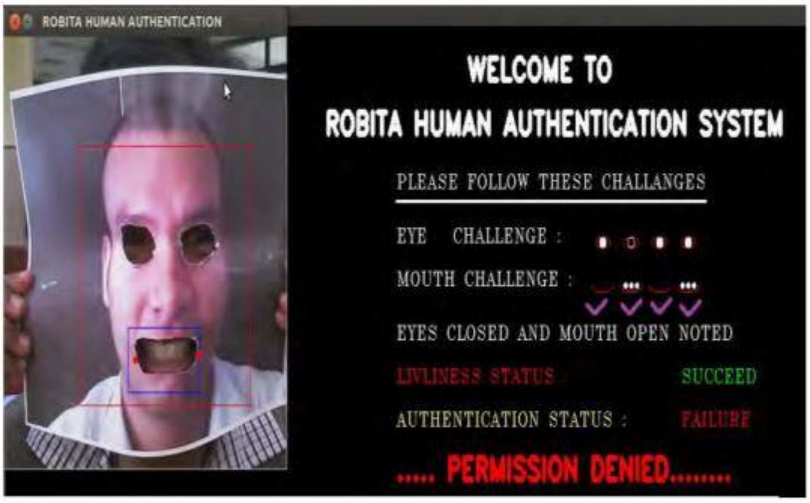
EYES OPEN AND MOUTH OPEN NOTED

LIVLINESS STATUS SUCCEED

RECOGNISED PERSON IS avinash

PERMISSION GRANTED

Figure 2.41: A genuine user recognized by the system in [131]



WELCOME TO
ROBITA HUMAN AUTHENTICATION SYSTEM

PLEASE FOLLOW THESE CHALLANGES

EYE CHALLENGE :

MOUTH CHALLENGE :

EYES CLOSED AND MOUTH OPEN NOTED

LIVLINESS STATUS SUCCEED

AUTHENTICATION STATUS : FAILURE

..... PERMISSION DENIED........

Figure 2.42: Spoof attack detected by the system in [131]

Boehm et al. [132] proposed a face authentication system that included a secrecy challenge called Secure Authentication with Face and Eyes (SAFE), an improved approach that used a gaze tracker. The user was required not only to show his face but also to track a secret icon moving across the screen. The system incorporated an enrolment phase and an authentication phase. The enrolment phase involved taking pictures of the user for the face recognition system and selecting a set of secret icons in a multi-phase challenge-response protocol. To log in successfully, the user had to identify his secret icon among a set and then track it visually. In the authentication phase, the SAFE user interface was displayed if the user's face was recognized, and a window opened with *n* icons populating the sides of the screen. One of these *n* icons was a member of the user's set of secret icons. Each icon sat on a line indicating its path of movement as shown in Figure 2.43; the icons moved with uniform speed along their paths, and the user was required to visually track his secret icon with his eyes. After the icons moved off the screen, another set of icons might appear. This second phase was similar to the first, but the set of icons on the screen was different. The user followed their next secret icon, repeating the process until it completed *p* total phases. The number of icons per phase *n* and the number of phase *p* depended on system configuration, which depended on security requirements.

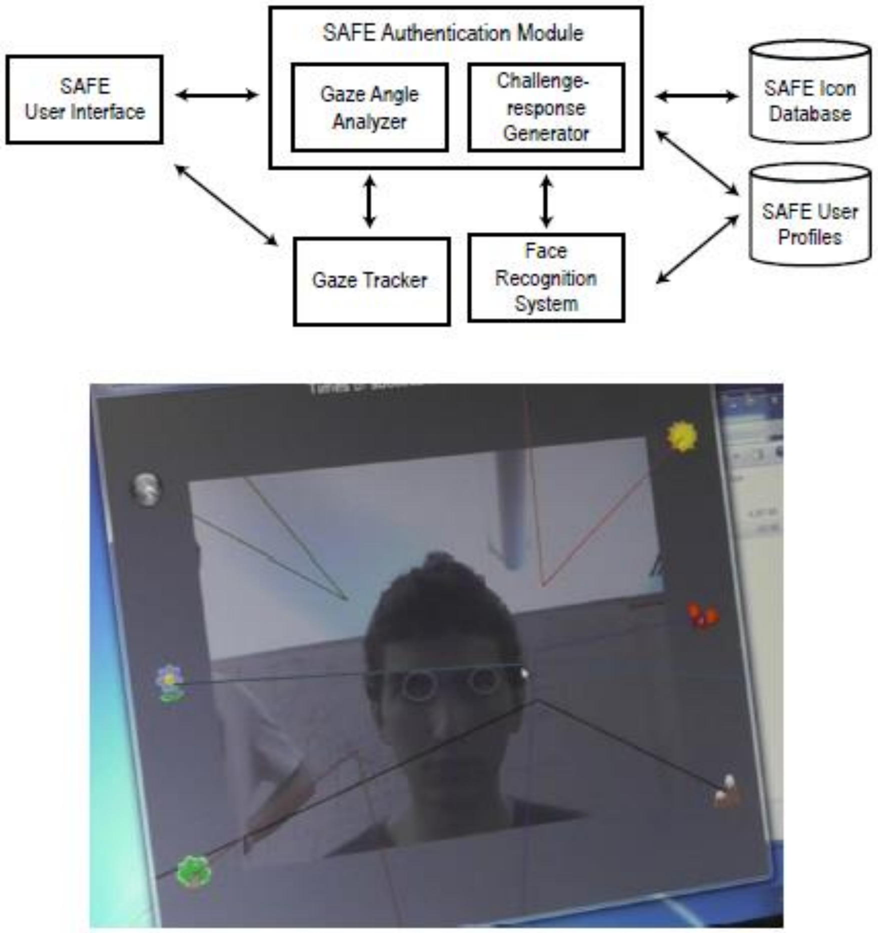


Figure 2.43: SAFE system architecture and SAFE user [132]

2.3.2.2 Involuntary Methods

Here, the system checked liveness by performing actions to which the user responds spontaneously [133]. Cai et al. [134] described face spoofing detection based on gaze estimation, which required no additional device or user cooperation. The proposed method comprised two elements: (1) gaze estimation and (2) a spoofing detection test.

The first phase established the position of the user's gaze on the computer screen; in the second phase, spoofing detection based on gaze estimation involved three stages: gaze tracking, gaze quantification and encoding, and liveness estimation.

First, a three- to five-second video clip of a test user was recorded, and gaze trajectory (i.e. gaze location in each frame of the video clip) was estimated using the gaze estimation model. The gaze histogram was then extracted by quantifying and encoding the gaze trajectory. Finally, the gaze histogram's information entropy was calculated to assess the gaze trajectory's uncertainty level and to estimate user liveness. The authors constructed two databases by combining self-collected data with parts of CASIA-FASD and REPLAY-ATTACK to form two new databases: Gaze-CASIA and Gaze-REPLAY-ATTACK. The experimental results confirmed that the proposed method effectively distinguished attacks from genuine access attempts. Figure 2.44 shows the flowchart of the proposed method.

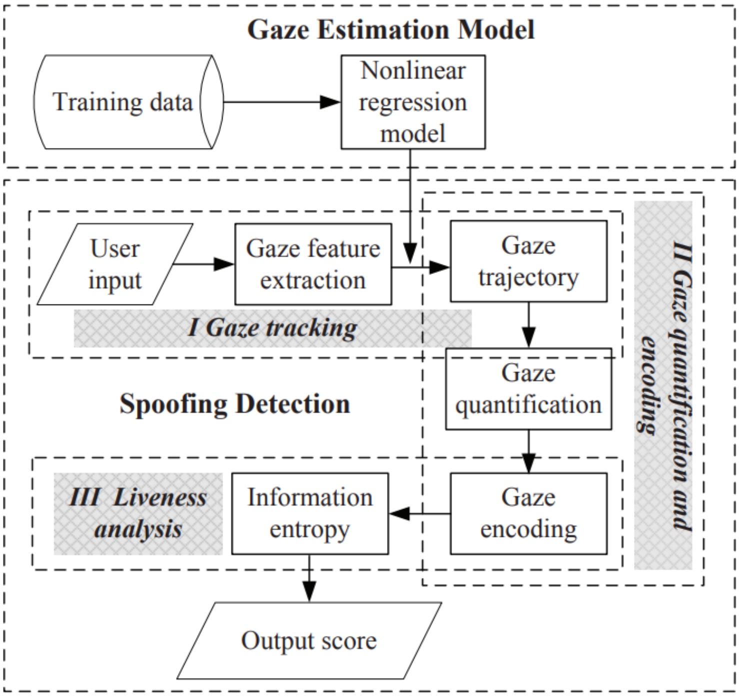


Figure 2.44: The proposed system architecture in [134].

Smith et al. [135] proposed a method of countering replay attacks on smart devices using a challenge-and-response technique. The image on the screen created the challenge, and the dynamic reflection of the face of the person looking at the screen generated the response. The sequence of screen images and associated reflections digitally watermarked the video. By extracting features from the reflection, it was possible to determine whether it matched the sequence of images displayed on the screen. Illuminating the face with a single bright colour (e.g. white) provided an excellent reflection response but supplied very little entropy (or uncertainty) for a challenge-response system. The use of multiple colours provided the necessary uncertainty, especially when different colours were used in a sequence. Entropy increased with more colours and longer sequences. The sequence of colours displayed provided the challenge, and the response was computed by analysing the face to determine the reflected colour. If the calculated response matched the sequence of displayed colours, there was higher confidence that the video had been captured at that moment in time and on that specific device rather than being replayed from a previously captured video. Figure 2.45 depicts the screen-face reflection-generic camera setup for a tablet computer. Results indicated that, under ideal conditions, face reflection sequences can be classified with a high degree of confidence.

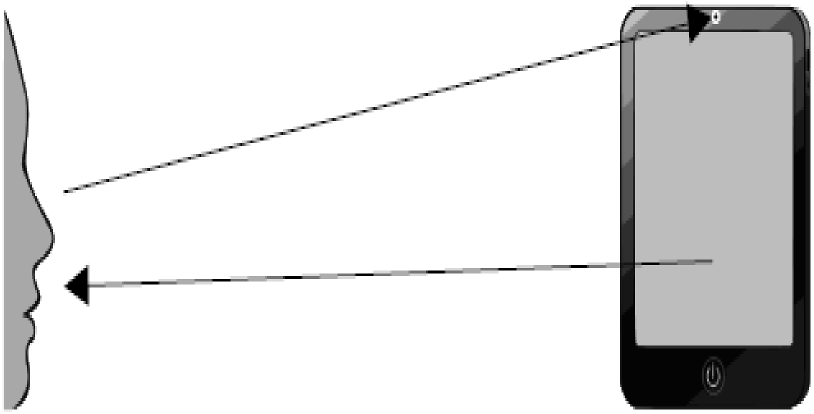


Figure 2.45: Analysing face reflections from images on the screen displayed as a colour image captured by the camera [135]

Some proposed PAD methods to overcome spoofing attacks have involved the use of an additional device, such as a light source or flash [136-140] to explore the differences between real and fake faces displayed in 2D media. For example, a face liveness detection method proposed in [141] was based on brightness difference, in which two images were captured from an object with and without flashlight. Rectangular regions of the face and background were defined by pixels in the upper right and lower left corners. The face region was first detected using the Sparse Network of Winnows (SNoW) classifier and the Successive Mean Quantization Transform (SMQT). Brightness values for the face and the background were then extracted from the images, and differences were computed separately as input features for liveness face detection. Figure 2.46 shows an example of a live person and spoofing video with and without flashlight. A collected data set of 21 subjects was used to evaluate the proposed method, and an acceptable level of accuracy was achieved.

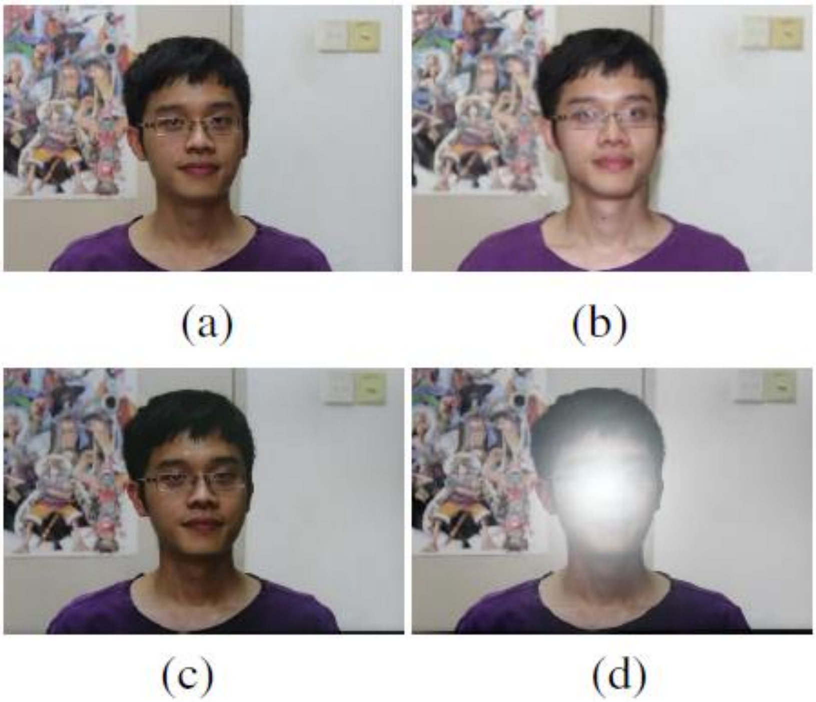


Figure 2.46: Examples of live person and spoofing video with/without flashlight: (a) live person without flashlight; (b) live person under flashlight; (c) video screen without flashlight; (d) video screen under flashlight [141]

2.4 Existing Face PAD Databases

While many published algorithms have been designed and tested on proprietary spoof databases [142] [143] [101] [144] to evaluate their effectiveness, only a few face spoof databases have been made publicly available. This section provides a brief summary of some public domain databases, including the NUAA Photograph Imposter database [40], the IDIAP PRINT-ATTACK database [42], the IDIAP REPLAY-ATTACK database [32] and the CASIA Face Anti-Spoofing Database [30]. Recently, the MSU Mobile Face

Spoof Database (MSU MFSD) [59], along with the Replay-Mobile [145] and Oulu-NPU [146] databases enabled mobile authentication of public face PAD benchmark databases. Other public databases such as the DaFEx database (8 subjects) and the VidTIMIT Audio-Video database [147] are less attractive for use in experimental evaluations because of their limited size and diversity of presentation attack. This section provides an overview of current public databases.

2.4.1 NUAA PI Database

Released in 2010, the NUAA Photograph Imposter database [40] is publicly available from the Pattern Recognition and Neural Computing Group at Nanjing University of Aeronautics and Astronautics. This database contains images of both genuine and attack attempts involving 15 participants. Images were captured using a generic webcam (model not specified) with a resolution of 640x480 pixels. Only hand-held printed photo attacks were included in this database, which was assembled under uncontrolled illumination conditions in three sessions two weeks apart. The amount of data from these sessions is unbalanced as not all subjects participated in all three sessions. The database comprises a training set (samples from the two first sessions) and a test set (samples from the last session), with no overlap between the two sets in terms of samples, although some subjects appear in both sets. Examples from the NUAA database are shown in Figure 2.47.

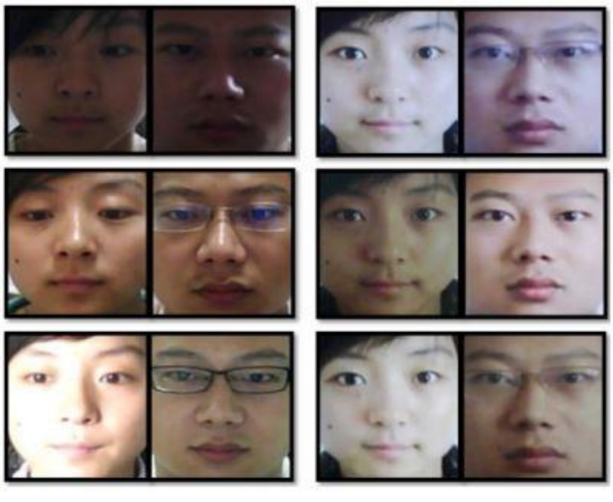


Figure 2.47: Sample images from the NUAA database [40], including real faces (left pairs) and photos (right pairs)

2.4.2 YALE-RECAPTURED Database

Released in 2011, the YALE-RECAPTURED database [41] is publicly available from the University of Campinas. It contains 640 images of genuine attempts and 1,920 attack attempts involving 10 participants. The genuine subset is extracted from the existing Yale Face DB-B [148], with data collected under 64 different illumination conditions. The attempted attacks were generated by displaying real images on three different LCD monitors: i) an LG Flatron L196WTQ Wide 19"; ii) a CTL 171Lx 1700 TFT; and iii) a DELL Inspiron 1545 notebook. The images were recaptured using two cameras: a Kodak C813 (resolution 8.2 megapixels) and a Samsung Omnia i900 (resolution 5 megapixels). Examples of images from the database are shown in Figure 2.48.

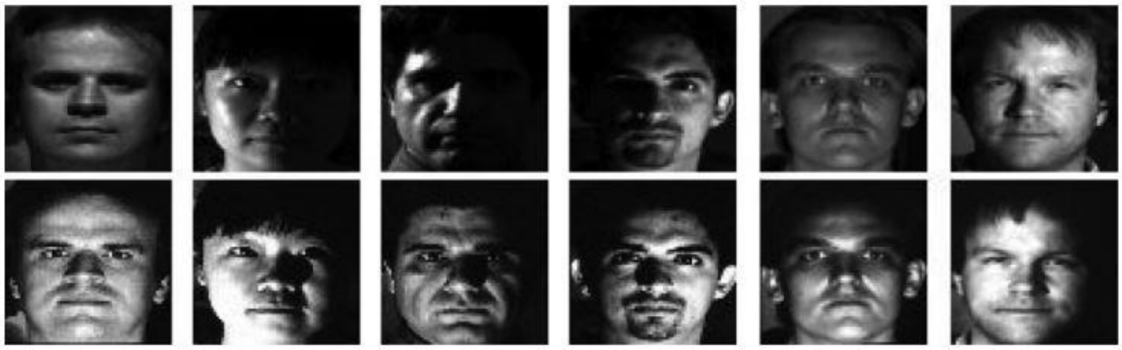


Figure 2.48: Example images from the Yale Recaptured Database: real faces (upper) and printed photos (below) [41]

2.4.3 PRINT-ATTACK Database

Released in 2011, the PRINT-ATTACK data set [42] is publicly available from the IDIAP Research Institute website. It includes 400 videos of genuine and print attack attempts by 50 participants, organised as training, development and test sets. Genuine and attack attempts were acquired using the 320X240 pixel (QVGA) camera of a MacBook laptop at 25 frames per second, with an average duration of about 10 seconds. Videos were recorded under two different background and illumination conditions: *i)* controlled and *ii)* adverse (as shown in Figure 2.49).

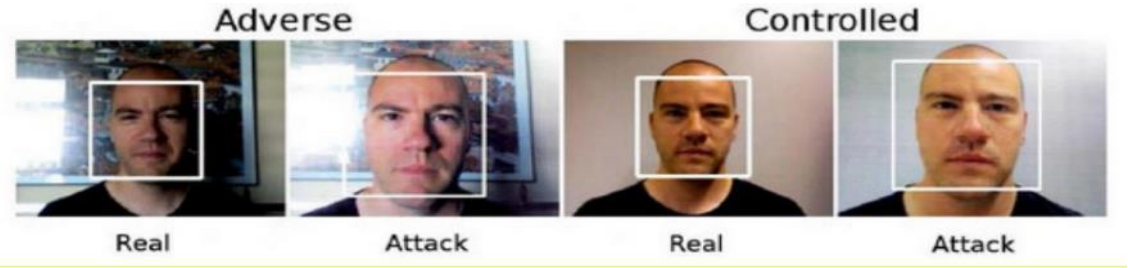


Figure 2.49: Sample images from Print-Attack database [42]

2.4.4 REPLAY-ATTACK Database

Released in 2012, the REPLAY-ATTACK face spoofing database [32] is publicly available at the IDIAP Research Institute website. It consists of 1,300 videos of 50 participants, of which 100 genuine videos were used for enrolment in face recognition experiments. The remaining 1,200 were separated into three non-overlapping subsets: training, tuning and test sets. The attempt videos were recorded under two different illumination conditions: i) controlled (with uniform background and artificial lighting) and ii) adverse (with natural illumination and non-uniform background). The three types of attack include printed photos, displayed photos and replayed videos. The videos were acquired using a 13″ MacBook (resolution 320×240 pixels). Examples of images from the database are shown in Figure 2.50.

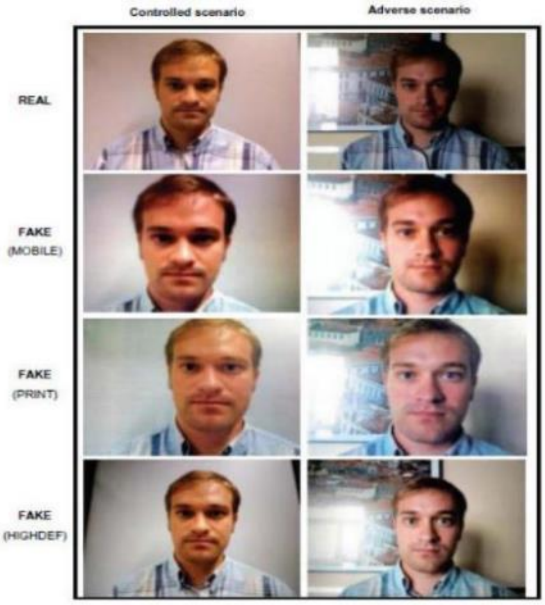


Figure 2.50: Typical examples of real and fake face images from Replay-Attack database in [32]

2.4.5 CASIA FAS Database

Released in 2012, the CASIA Face Anti-Spoofing database [30] is publicly available from the Chinese Academy of Sciences (CASIA) Centre for Biometrics and Security Research (CASIA-CBSR). It consists of 600 video recordings of genuine and attack attempts by 50 participants, divided into training and test sets, with no overlap between them in terms of subjects and samples. Three different attacks were considered: warped photos, cut photos and replayed videos. Samples were captured at three different resolutions: i) low resolution (using an old 640×480 pixels USB web camera); ii) normal resolution (using a modern 480×640 USB web camera); and iii) high resolution (using a 1920×1080 pixels Sony NEX-5 high definition camera. Example images from the CASIA Face Anti-Spoofing database are shown in Figure 2.51.

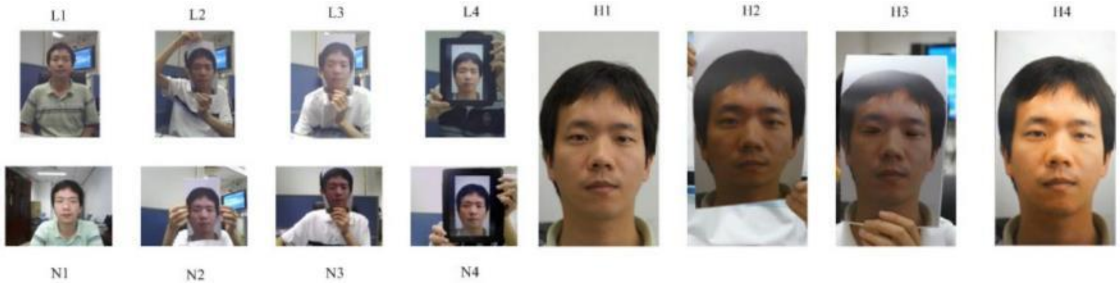


Figure 2.51: One complete video set for an individual subject in the CASIA-FAS database [30]

2.4.6 3D MASK-ATTACK Database

Released in 2012, the 3D MASK-ATTACK database (3DMAD) [149] is publicly available at the IDIAP Research Institute website. This dataset consists of genuine and 3D mask attack access attempts by 17 different participants, performed wearing real-size 3D masks of genuine users. The 3D masks were generated using ThatsMyFace.com, which requires only frontal and profile face images of each user. The database was captured in three different sessions; the first two captured genuine access samples, and the third session captured 3D mask attacks. In each session and for each subject, five videos of 10 seconds were recorded using the Microsoft Kinect for Xbox 360 at a resolution of 640×480 pixels.

2.4.7 MSU-MFSD Database

Released in 2015, the MSU-MFSD [59] is publicly available from the Pattern Recognition and Image Processing (PRIP) Lab at Michigan State University. The database contains 440 genuine and attack attempts by 35 participants. Videos of genuine attempts (averaging 12 seconds) were recorded using two devices: a 13″ MacBook Air (built-in camera with a resolution of 640×480 pixels) and a Google Nexus 5 (Android 4.4.2 with a camera resolution of 720×480 pixels). The database includes three kinds of attack: printed photo attacks, video replays on a smartphone (iPhone 5s) and high-definition video replays (captured on a Canon 550D SLR and played back on an iPad Air). Examples from the MSU MFSD database are shown in Figure 2.52.

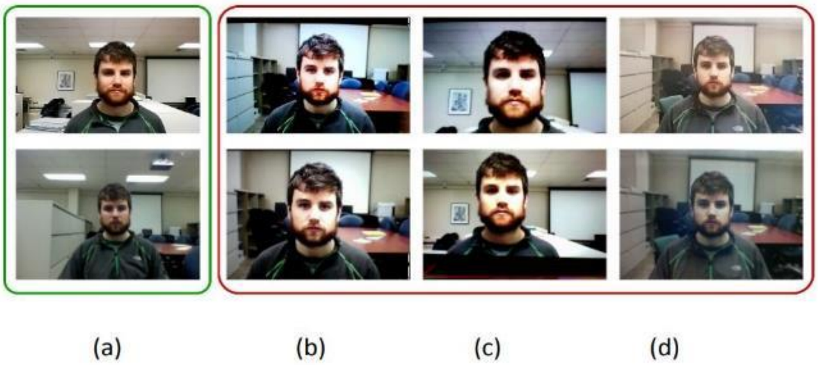


Figure 2.52: Example images from the MSU MFSD database: (a) genuine faces; (b) video replay attack by iPad; (c) video replay attack by iPhone; (d) photo attack [59]

2.4.8 Replay-Mobile Database

Released in 2016, the Replay-Mobile Database [145] is publicly available at the IDIAP Research Institute website and consists of 1,200 videos of genuine and attack attempts by 40 subjects. Two acquisition devices (a tablet and a smartphone) were used to record genuine and attack attempts. The attacks include printed photos and replayed videos. The data were collected in two sessions at an interval of two weeks under two different illumination conditions: *lighton* (electric lights on) and *lightoff* (electric lights off). The database includes five different mobile scenarios: i) *controlled* (uniform background and light switched on); ii) *adverse* (uniform background and light switched off); iii) *direct* (complex background and user facing window with direct sunlight). The database is divided into 3 subsets (training, development and test) with no overlap between them. Examples from the Replay-Mobile database are shown in Figure 2.53.

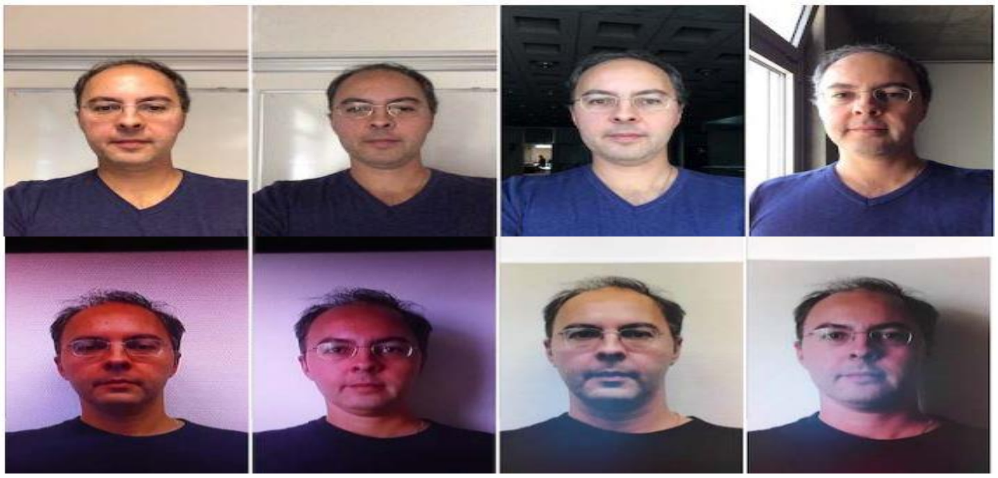


Figure 2.53: Example images from the Replay-Mobile database (top: real access; bottom: attack access) [145]

2.4.9 OULU-NPU Database

Released in 2017, the OULU-NPU face presentation attack detection database [146] is publicly available from the University of Oulu in Finland and from the North-Western Polytechnical University in China. The database consists of 4,950 genuine and attack attempt videos of 55 different participants, recorded using the front cameras of six mobile devices with full HD resolution (1920×1080 pixels). The data were collected in three sessions with differing illumination conditions and background scenes. The presentation attack types are printed photos and replayed video, created using two printers and two display devices. Examples from the database are shown in Figure 2.54.

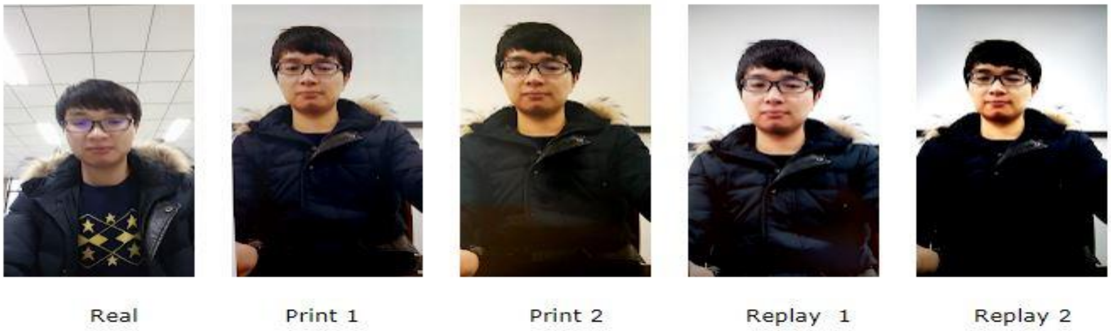


Figure 2.54: Sample images of real and attack videos from the OULU-NPU database [146]

2.4.10 IST Lenslet Light Field Face Spoofing Database

Released in 2017, the IST Lenslet Light Field Face Spoofing Database (IST LLFFSD) [80] consists of 700 images of genuine and attack attempts collected from 50 participants, captured with a Lytro ILLUM lenslet light-field camera. The six types of simulated presentation attack include printed paper, wrapped printed paper, laptop, tablet and two different mobile phones. The data were collected in two separate acquisition sessions, at a time interval of between 1 and 6 months, under controlled conditions and with a uniform background. Examples from the database are shown in Figure 2.55.

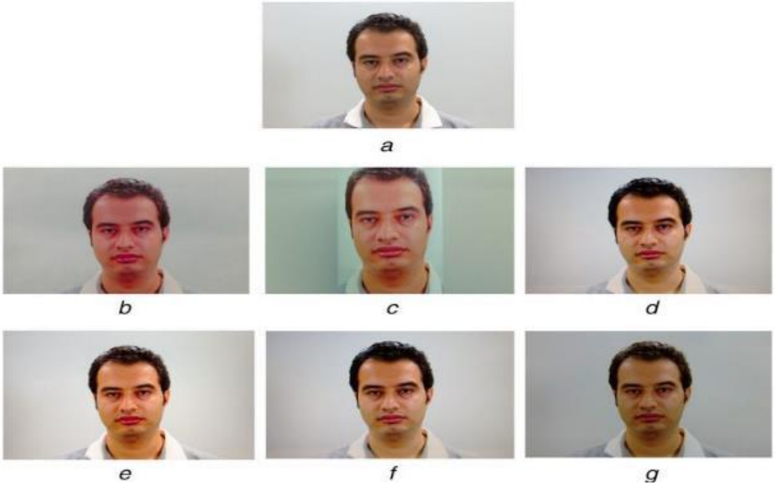


Figure 2.55: Example images from the IST LLFFSD: (a) genuine face; (b) print paper attack; (c) wrapped print paper attack; (d) laptop attack; (e) tablet attack; (f) mobile attack 1; (g) mobile attack 2 [80]

2.5 Summary

As discussed here, extensive research has examined the vulnerabilities of face recognition systems to direct attack, and multiple approaches have been proposed to secure them against this threat. The chapter included a categorization of techniques and types of liveness indicators or cues used for face PAD, along with an overview of current public databases encompassing most attack scenarios.

Chapter 3

Experimental Framework

3.1 Introduction

This chapter presents details of the experimental framework developed for the evaluation of the proposed system. It also describes the database that is developed for the purpose of evaluation of face PAD methods. There are already some databases available for evaluation of PAD systems. However, due to the specific nature of the proposed challenge-response scheme for PAD, none of the existing public databases are appropriate, and therefore, a new database has been collected to evaluate the proposed method. The challenge was designed to direct the user gaze and the system collects the user response (gaze).

In this chapter, we introduce the new face PAD database, the KGDD, which covers a diverse range of potential attack instruments and simulates mobile device use. The KGDD consists of 2400 sets of genuine and attack attempts collected from 80 participants. The challenge design used for collecting the data is also explained. Three challenge types are implemented, which include Lines, Curves and Points challenges. Three presentation attack instruments are implemented: displayed photo attack, 2D mask attack and 3D mask attack. The usability of the proposed scheme has also been investigated via a 'participant survey'. Test protocols are provided, consisting of 18 scenarios for a thorough evaluation. In this chapter, we also introduce the software for detecting facial landmarks used in feature extraction. This chapter also covers the performance measures used to assess detection rates.

The remainder of the chapter is organised as follows. Section 3.2 introduces the proposed system. Section 3.3 presents the introduction of the KGDD, covering data collection setup, the demographics of the participants, presentation attack instruments, and challenge design. Section 3.4 provides the test protocols. Section 3.5 presents the objective evaluation methods and metrics used in this study, and Section 3.6 presents a brief conclusion.

3.2 Proposed System

Figure 3.1 shows the proposed algorithm based on a challenge-response mechanism. A visual stimulus appearing on a display screen and moving in different directions serves as the challenge, and the user's gaze as they follow the stimulus on the screen serves as the response. The user is required to track the moving stimulus across the screen with their gaze, using natural head and eye movements. The camera captures the facial images at each location of the challenge. The placement of the target stimulus on the screen and the image acquisition are synchronised using a control mechanism. As the system is based on gaze features, the facial landmarks in the captured frames are extracted. From these landmarks, various features are computed and then used to classify the presentation attempt as a genuine attempt (i.e. coming from a live sample) or a presentation attack attempt (i.e. coming from an impostor presenting a displayed photo, 2D mask or 3D mask attack). The scores from the separate classifiers are combined using a score fusion rule to make a final decision.

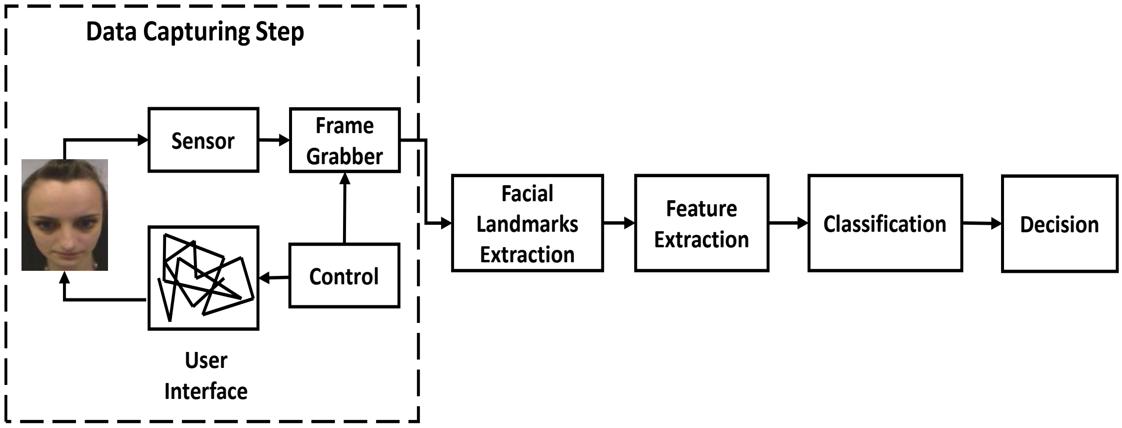


Figure 3.1: System block diagram

3.2.1 Facial Landmark Extraction

The images thus captured during the challenge-response operation were processed using Chehra version 3.0 [150] to extract facial landmark points. Chehra is a fully-automatic real-time face and eyes landmark detection and tracking software capable of handling faces in uncontrolled natural settings. This software, which is written in C++, returns 49 different facial landmark points and 10 eye landmark points, as shown in Figure 3.2. It also estimates the pitch, roll and yaw angles of the 3D head-pose. The coordinates of these landmarks were used for feature extraction in the proposed scheme for PAD, and the landmarks were returned in an integer array.

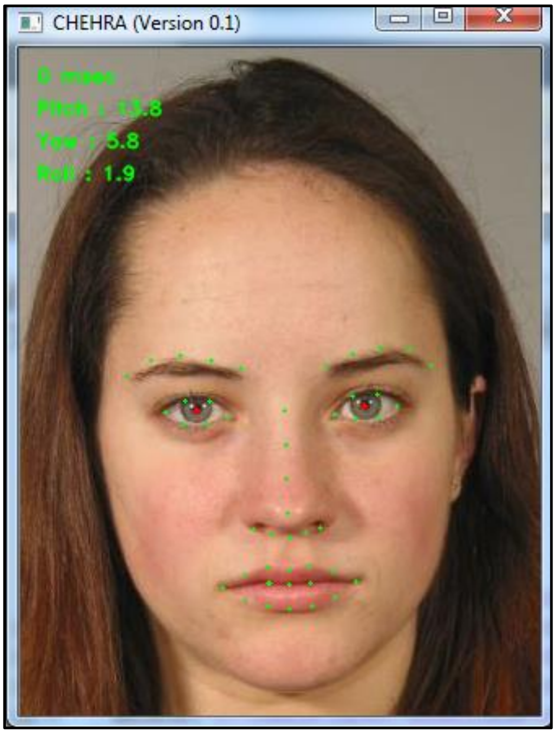


Figure 3.2: Landmarks extracted using Chehra [150]

3.2.2 Feature Extraction

After facial landmark extraction, many features can be computed from these landmarks, which can be used to distinguish between genuine and attack attempts. In the proposed system, three gaze-based features were extracted to measure the similarity between the gaze and challenge trajectories. These features include gaze correlation analysis, gaze alignment consistency and gaze time dynamics features and will be described in Chapter 4.

3.2.3 Classification

In this work, the k-NN classifier from the PRtools package [151]automatically determined an optimum k value (the number of training samples which is closest in distance to a new sample) for each experiment. Each experiment was repeated 100 times with random sets of data for training and testing, resulting in a different optimum k value for each run. In most cases, the optimal k value was equal to 1. In our experiments, we found that by training separate classifiers, the ensemble classifier performed better than training a single classifier on the whole database.

In the testing stage, the input feature vectors are fed to all classifiers and their outputs are fused to get the final result, as shown in Figure 3.3. We have evaluated three types of score-level fusion schemes: sum, product and majority-vote rules [152]. In most cases, the product rule performs better as the classifiers are trained with different data, and so it is used in all our experiments in this thesis. Product rule is mathematically defined as [152]:

$$f = \prod_{m=1}^{M} X_m \tag{3.1}$$

where f is the fused score, X_m is the score of the m^{th} classifier, m =1,2,…, M.

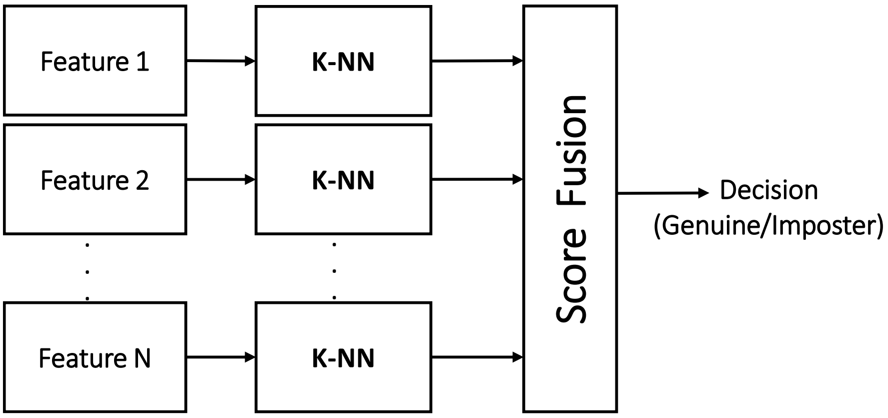


Figure 3.3: The classification structure.

3.3 Kent Gaze Dynamics Database

This section describes the database that has been developed for the purpose of evaluation of the proposed face PAD methods. There are already some existing public databases available for the evaluation of various PAD systems. However, none of these databases are appropriate due to the specific nature of the proposed challenge-response scheme for PAD. Therefore, a new database has been collected to evaluate the proposed system.

The Kent Gaze Dynamics Database (KGDD) covers a diverse range of potential attack instruments and simulates the mobile device use. It consists of 2400 sets of genuine and attack attempts collected from 80 participants. Three presentation attack instruments were simulated: displayed photo attack, 2D mask attack and 3D mask attack. Test protocols are also recommended, consisting of 18 scenarios for a thorough evaluation.

3.3.1 Challenge Design and Response Acquisition

A challenge-response user interface was implemented using a graphical user interface (GUI) in MATLAB. A small shape (stimulus) is presented to the users on the screen whilst users are seated in front of the computer and asked to follow the stimulus with natural head/eye movements as it changes its location. Two screen sizes (7.45 × 13.23 cm and 15.87 × 21.18 cm, for a mobile device and for a tablet device, respectively) were chosen to display the challenge. Figure 3.4 shows the tablet format, and Figure 3.5 presents the phone format. Three different types of challenge trajectories (Lines, Curves and Points) were shown to the user for each attempt. The challenge interface shows three buttons, namely Lines, Curves, and Points, as shown in Figure 3.4 and Figure 3.5. When the user clicks Lines, Curves, or Points, the challenge is held for one second to allow users' eyes to fixate on the challenge before it moves and the camera starts capturing images at each location. The record ends when the stimulus finishes its movements.

In the case of the **Lines** challenge (Figure 3.6), the stimulus moves along a set of connected straight lines in various random directions. Each session contains several lines and the stimulus moves along each line for an average of 1 s per line. Data was captured for the whole attempt (approximately 1 min), including all the lines shown to the participants. Subsets of the data captured were used for training and testing the system.

The **Curves** (Figure 3.7) challenge has a very similar design to the Lines challenge, except that the stimulus moves along randomised curved paths. The curved trajectories are generated using a number of random control points used for fitting a spline function. Total duration of the Curves challenge presentation is approximately 1 min.

For the **Points** challenge (Figure 3.8), the stimulus appears at random locations on the screen. In this implementation, there are 30 locations (i.e. D = 30) where the stimulus can appear up to three times at random locations (thus, M = 90), and a facial image is captured at each of these locations. In this work, a 0.5 s delay was used between each presentation of the stimulus shape to provide ample time for the users to fixate their gaze. The total duration of the challenge presentation was about 45 s. This type of challenge was inspired by the work in [128] where the challenge was appeared on the PC screen at 30 predefined locations visiting each position 3 times and the duration of the challenge was about 2 minutes.

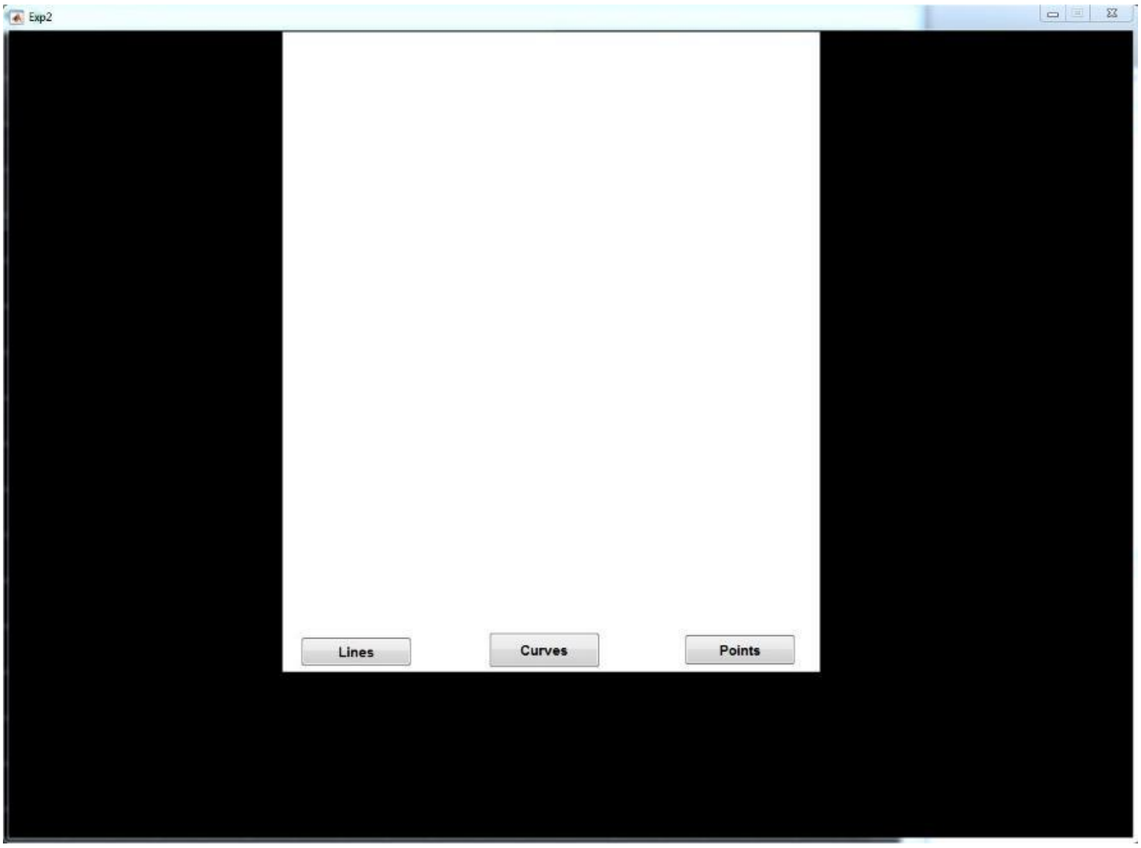


Figure 3.4: Tablet format

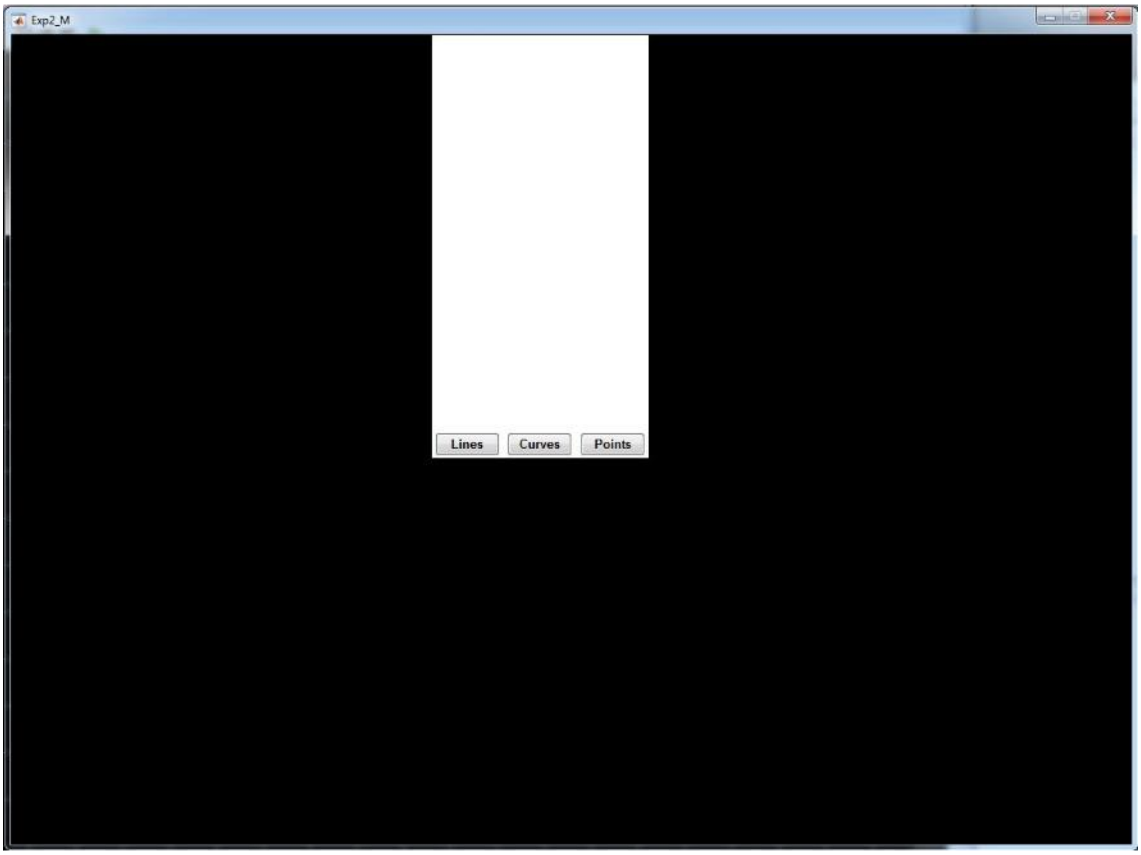


Figure 3.5: Phone format

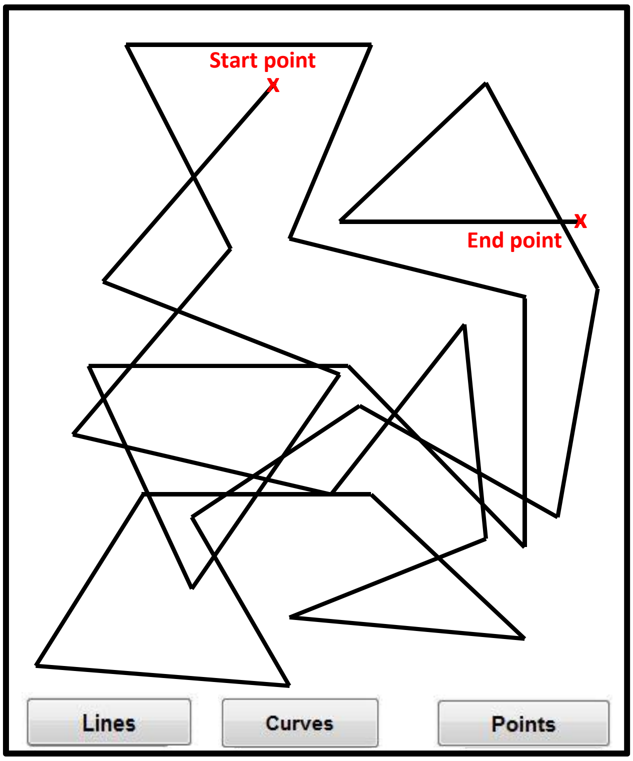


Figure 3.6: Lines challenge trajectory

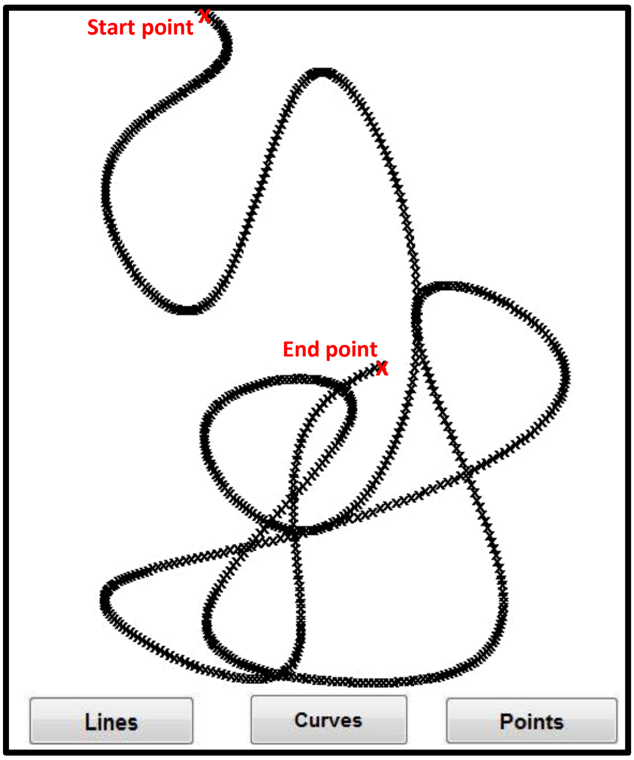
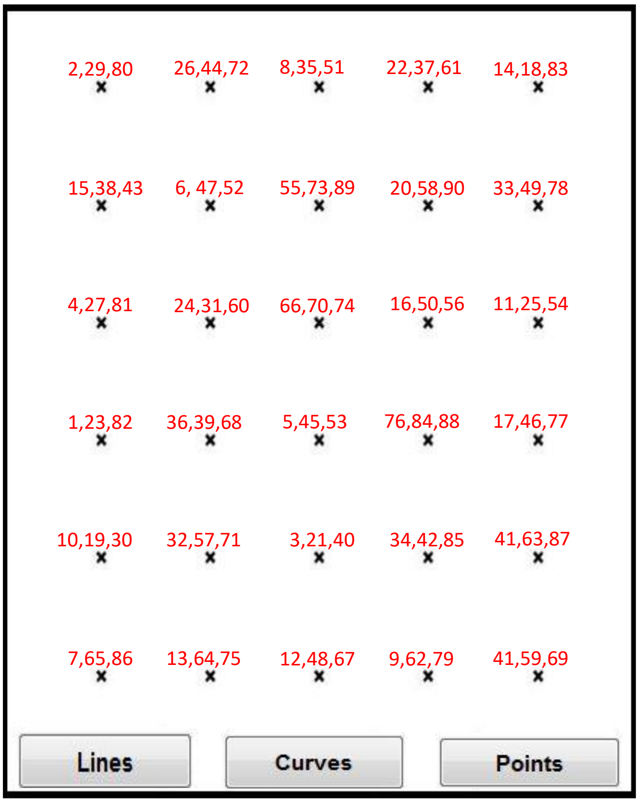


Figure 3.7: Curves challenge trajectory



2,29,80	26,44,72	8,35,51	22,37,61	14,18,83
15,38,43	6, 47,52	55,73,89	20,58,90	33,49,78
4,27,81	24,31,60	66,70,74	16,50,56	11,25,54
1,23,82	36,39,68	5,45,53	76,84,88	17,46,77
10,19,30	32,57,71	3,21,40	34,42,85	41,63,87
7,65,86	13,64,75	12,48,67	9,62,79	41,59,69

Figure 3.8: Points challenge grid

3.3.2 Generation of Attacks Scenarios

Three types of attack scenarios were investigated in this study. The scenarios considered here include that of an impostor attempting authentication by displaying to the camera an image on an iPad screen, a 2D photo mask or a 3D mask in front of the camera of the face recognition system. The attacks were carried out as follows.

1) **Displayed photo:** A photo of a genuine user was displayed on an iPad mini 2 (7.9" Retina display) which was held in front of the camera to generate the attack. To simulate the gaze tracking the stimulus, the attacker moved the device by hand in accordance to the challenge on the screen.

2) **2D photo mask:** A high-quality printed colour photo of a genuine user with holes made in the place of the pupils was held by the volunteer in front of the eyes as a mask and used to follow the graphical target. Three different sizes of 2D masks (small, medium and large) with different pupillary distance (PD) were produced. The reason for producing a set of artefacts with pupillary holes at different distances was to better fit the facial dimensions of the participants with different PDs. These 2D mask were printed on A4-size matte paper using a laser colour printer. Then 3 mm holes were made in the pupil centres of these 2D masks.

3) **3D mask:** For creating the 3D masks, a particular commercial service, ThatsMyFace.com (http://www.thatsmyface.com/), was used which offers a transformation of 2D portrait images into 3D sculptures. The service is called ThatsMyFace.com (http://www.thatsmyface.com/). Using this company, a 3D face model was constructed using frontal and profile images of a person. The 3D masks could be printed and delivered by post in several forms such as a wearable life-size mask in hard resin or a paper-cut file. For the KGDD, one frontal and two profile face images were taken from two different users (male and female) and uploaded to ThatsMyFace.com. For each user, a custom LifeMask was ordered. These masks were made out of a hard resin composite in full 24-bit colour. The 3D shapes of the masks were not precise since the

reconstructed models are only approximately computed from 2D images. The resulting mask quality depended on the input images as well as the performance of the reconstruction algorithm. Holes were then drilled at the pupil locations. The attackers held these masks in front of their face during the simulated attack.

3.3.3 Data Collection Equipment Setup

The hardware setup consisted of a PC, a display monitor for the PC and a webcam. The computer used has a processor with Intel i5-3470 quad-core 3.20 GHz CPU, 16GB RAM, and Windows 7 64bit system. It has a 23” LCD screen, a commonly used monitor type, with a resolution of 1920 × 1080 pixels.

A Logitech Quick Cam Pro 5000 camera was used centrally mounted on the top of the display monitor. The challenge (the visual stimulus) was displayed on the LCD screen for the user to follow by gaze. The camera captures an image at each location of the challenge. The average standoff distance between the face and the camera is approximately 30-50 cm. The data acquisition system setup was similar to that shown in Figure 3.9.

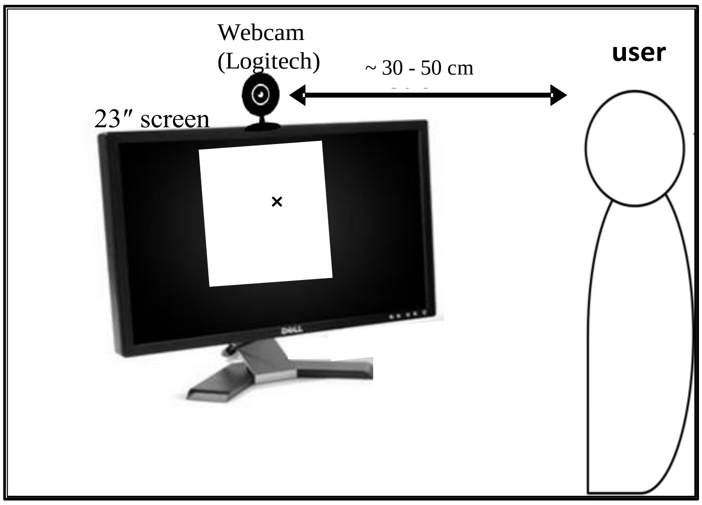


Figure 3.9: Hardware setup

Ethical approval was required as data collection contains material which are considered to be a sensitive and personal nature. An ethical approval application was filled and submitted to the Sciences Research Ethics Advisory Group (REAG) at the University of Kent, with a copy of the project protocol and other supporting documentation, including a participant information sheet (see Appendix A), a consent form (see Appendix B), participant registration form (see Appendix C), and an exit questionnaire sheet (see Appendix D). The application was approved with minor revisions.

3.3.4 Participants

The Kent Gaze Dynamics database (KGDD) consists of 2400 sets of genuine and attack attempts collected from 80 participants (33 females and 47 males). Metadata associated with each participant were also collected to facilitate experiments using specific subgroups. The available information includes age, gender, handedness and visual aids (glasses or contact lenses as opposed to none). Participants were all over 18 years old. There were slightly more males than females. The volunteers were of Asian, Middle Eastern, Eastern and Western European descent. The demographic distributions of the participants are presented in Table 3.1.

Table 3.1: Demographic distribution of the users

Demographic types		**Number of participants**
Gender	Male	47
	Female	33
Age Band	18–25	19
	26–35	43
	36–45	10
	46–55	4
	55+	4
Eyewear	Glasses	30
	Contact lenses	5
Handedness	Right-handed	71
	Left-handed	9

3.3.5 Data Collection

Participants performed genuine attempts with and without tinted glasses to investigate the impact of wearing these glasses on the attack detection accuracy. Figure 3.10 shows the tinted glasses that were used in data collection.

Data were collected from genuine attempts (Figure 3.11) where users were tracking a moving visual target (‘stimulus’ or ‘challenge’) with natural head/eye movements. Data were also collected from impostor attacks where users were holding a displayed photo (Figure 3.12), looking through a 2D mask (Figure 3.13) or wearing a 3D mask of a genuine user (Figure 3.14) and attempting to follow the stimulus during the simulated spoofing attacks. Each attempt caused the software to acquire images at every location of the challenge. The resolution of each image is 640 × 480 pixels. This resolution provided adequate picture quality for recognition of the facial landmarks for gaze analysis.

Each participant performed 18 attack attempts, using a displayed photo, a 2D mask or a 3D mask for three different challenges in both phone and tablet formats. Each participant also made 12 genuine attempts, with and without tinted glasses, for three different challenges in both phone and tablet formats. This created 1440 sets of attacks and 960 sets of genuine attempts in total. Figure 3.15 shows example images of genuine and presentation attack attempts in the KGDD database. For a subset of the KGDD database, only the data of 70 participants who provided their consent to have their face images are released.

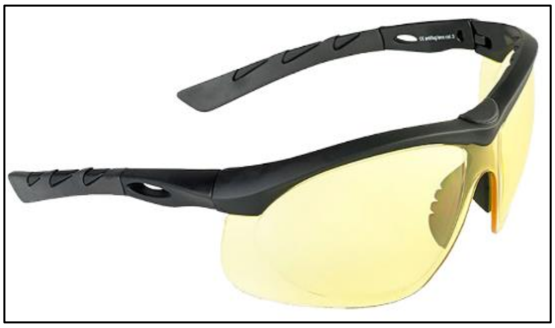


Figure 3.10: Tinted glasses used in data collection.

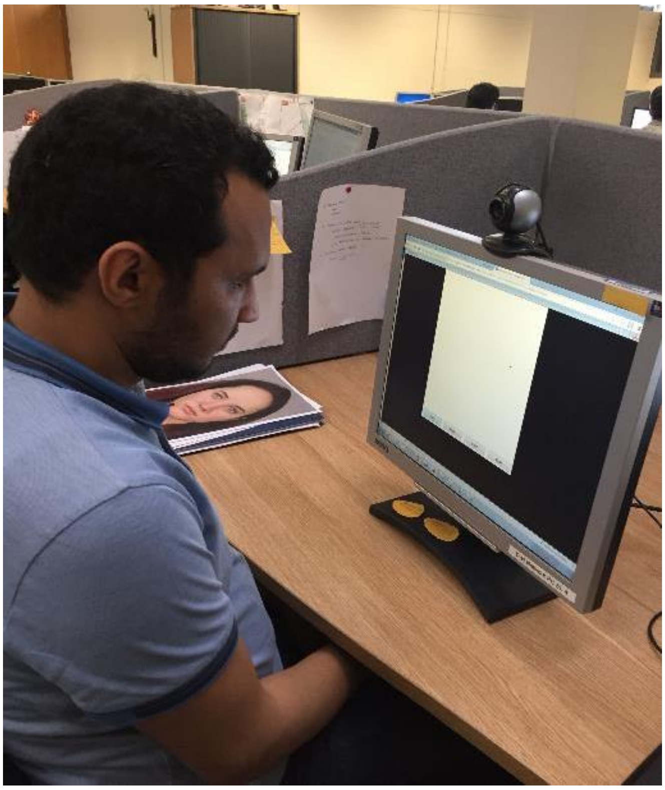


Figure 3.11: A genuine user tracking the moving challenge

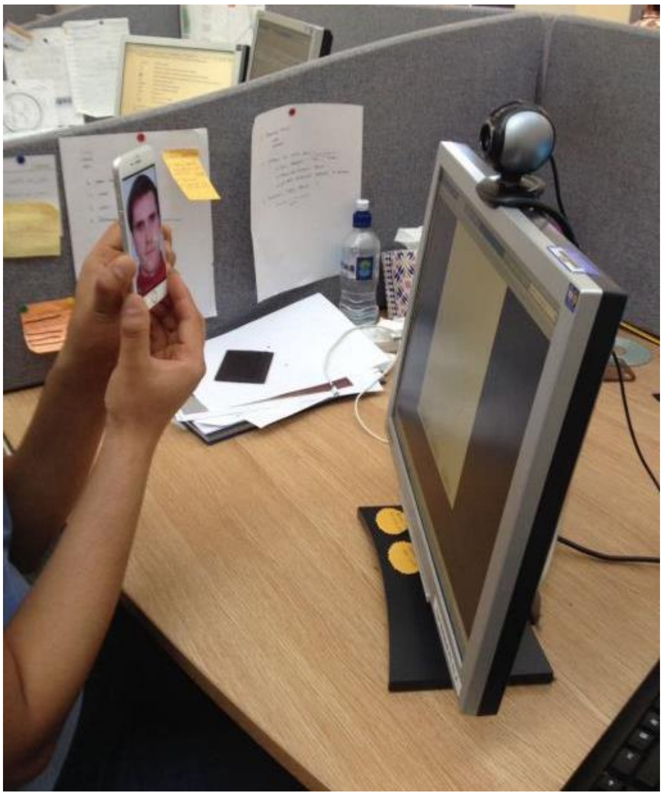


Figure 3.12: An impostor holding a displayed photo and attempting to follow the challenge during the simulated spoofing attack

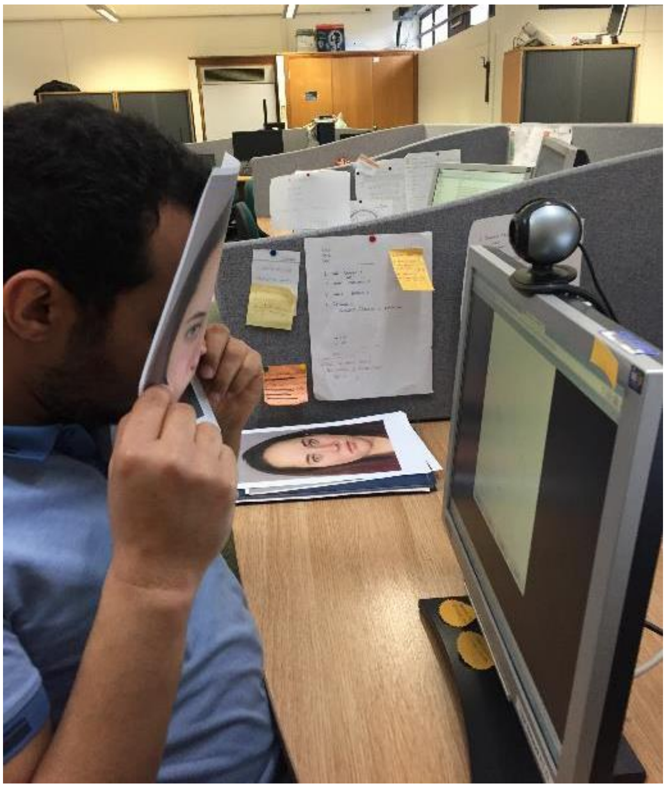


Figure 3.13: An impostor holding a 2D mask and attempting to follow the challenge during the simulated spoofing attack

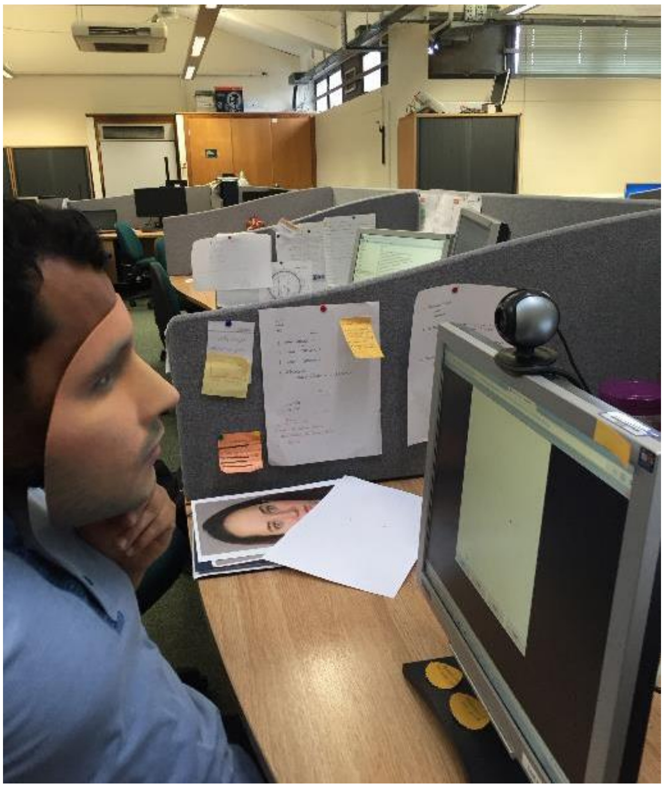


Figure 3.14: An impostor wearing a 3D mask and attempting to follow the challenge during the simulated spoofing attacks

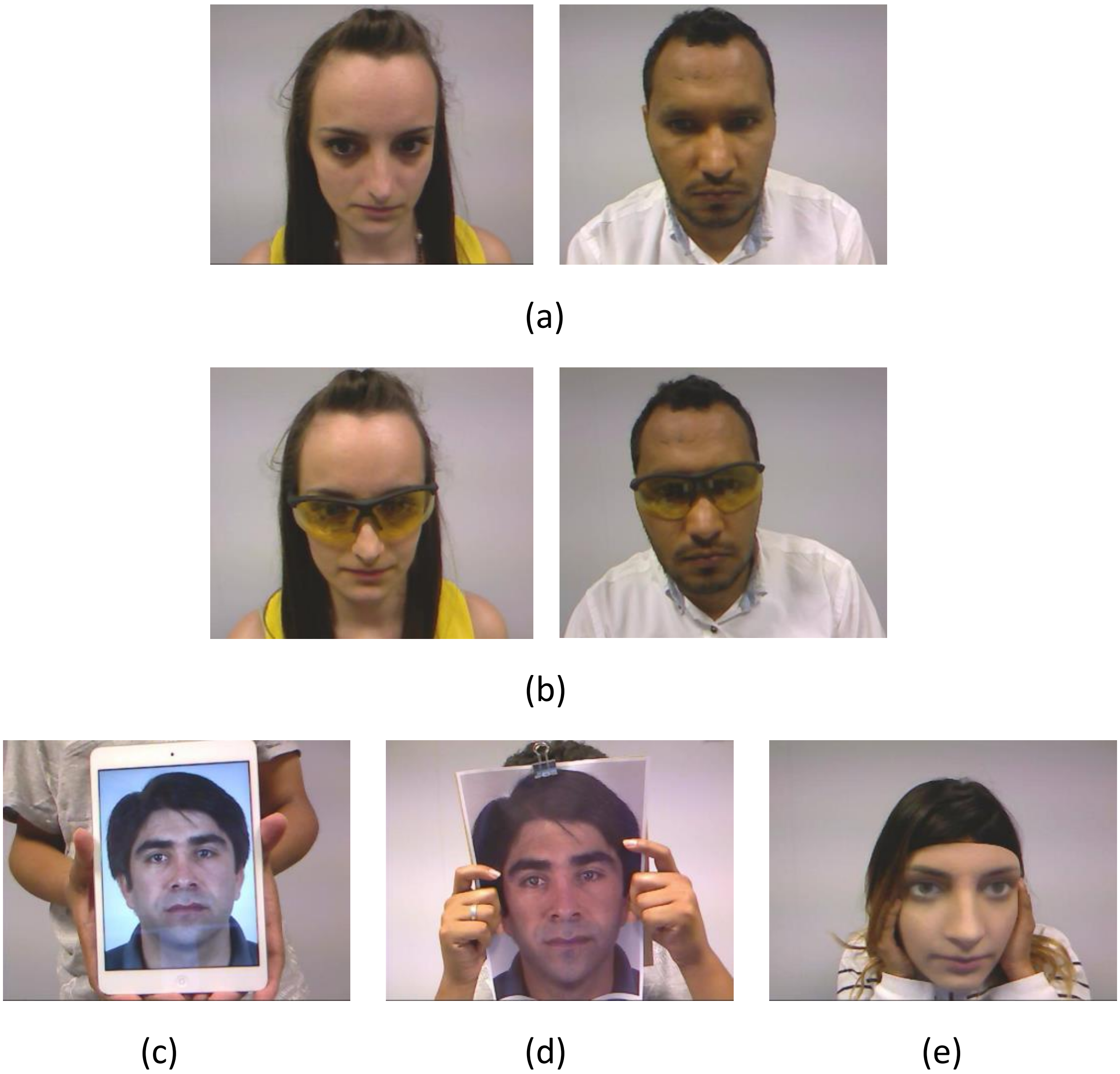


Figure 3.15: Example images of genuine and attack attempts in the collected database. (a) A genuine user; (b) a genuine user wearing tinted glasses; (c) a displayed photo attack using an iPad; (d) a 2D mask attack; (e) a 3D mask attack.

3.3.6 Usability Survey

At the end of each data collection session the participants were given an exit questionnaire (see appendix D) to find out their experience using the challenge-response system. The questionnaire explores the following issues:

- the system's ease of use;
- its applicability in real life;
- authentication delay acceptability.

Figure 3.16 shows a summary of these finds. The majority of participants agreed that following the stimulus using a 3D Mask was hard, whereas the stimulus was easy to track in genuine attempts.

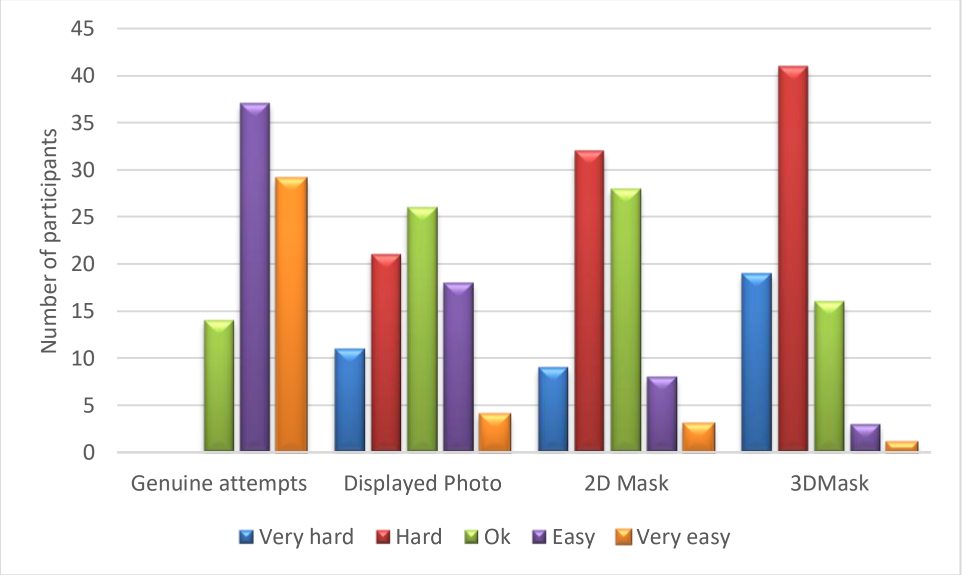


Figure 3.16: The ease of user interaction attempts

In terms of the usability of the challenge-response system in real life applications, results are shown in Table 3.2. The majority of the participants would have confidence in using the system to unlock their phones and at unattended passport controls. However, only a minority of the participants would be happy to use this system for online shopping and online banking. The last question in the survey was about the acceptable duration of the challenge for participants to use this system in real life applications. Figure 3.17 shows that 80% of participants would like to have this duration be less than 3 s for unlocking their phones, while 58.70% and 66.20% of users chose more than 5 s for online shopping and online banking, respectively.

Table 3.2 Usability of the system in real applications

Application	**No of participants**		
	Yes	**No**	**Not sure**
Unlocking your phone	64	7	9
Online shopping	24	17	39
Online banking	14	20	46
Unattended passport control	49	13	18

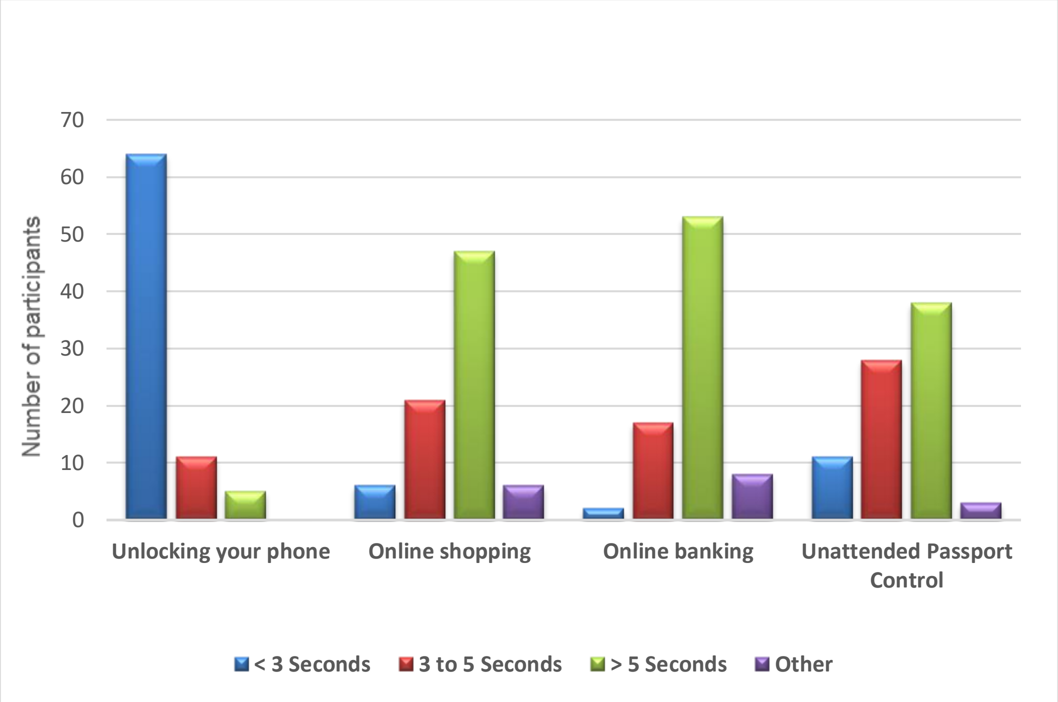


Figure 3.17: Authentication delay acceptability

3.3.7 Comparison with Previous Databases

Table 3.3 Table 3.3 gives a comparison between the existing face PAD databases described in Section 2.4 and the new KGDD in terms of sample size, acquisition device, presentation attack instruments and gender and race distributions of subjects. Existing public databases for the evaluation of PAD techniques cover a relatively narrow range of possible attack instruments and do not include the potentially more powerful challenge-response approaches for PAD.

Table 3.3: Comparison between the existing face PAD databases and the new KGDD

Database	No. of Subjects	No. of Attempts	Acquisition camera	Attack type	Gender
NUAA-PID [38]	15	24 genuine 33 spoof	Webcam (640 × 480)	▪ Printed photo	Male 80% Female 20%
ZJU Eyeblink [86]	20	80 genuine 100 spoof	Webcam (320 × 240)	▪ Printed photo	Male 65% Female 35%
REPLAY ATTACK [30]	50	200 genuine 1,000 spoof	MacBook 1300 camera (320 × 240)	▪ Printed photo ▪ Displayed photo (mobile/HD) ▪ Replayed video (mobile/HD)	Male 86% Female 14%
CASIA FASD [28]	50	150 genuine 450 spoof	Low-quality camera (640 × 480) Normal-quality camera (480 × 640) Sony NEX-5 camera (1280 × 720)	▪ Printed photo ▪ Cut photo ▪ Replayed video (HD)	Male 81% Female 19%
3DMAD [147]	17	170 genuine 85 spoof	Microsoft Kinect sensor	▪ 3D mask video	N/A
MSU MFSD [57]	35	110 genuine 330 spoof	MacBook Air 1300 camera (640 × 480) Google Nexus 5 camera (720 × 480)	▪ Printed photo ▪ Replayed video (mobile/HD)	Male 63% Female 37%
Replay-Mobile [143]	40	390 genuine 640 spoof	The front-camera of the mobile device	▪ Printed photo ▪ Replayed video	N/A
OULU-NPU [144]	55	1980 genuine 3960 spoof	The front cameras of six mobile devices	▪ Printed photo ▪ Replayed video (mobile/HD)	Male 73% Female 27%
IST LLFFSD [78]	50	100 genuine 600 spoof	Lytro ILLUM lenslet light field camera	▪ Printed photo ▪ Wrapped printed photo ▪ Displayed photo (laptop, tablet and 2 mobile phones)	Male 66% Female 34%
KGDD	80	960 genuine 1440 spoof	Web-cam (640 x 480)	▪ Displayed photo ▪ 2D photo mask ▪ 3D mask	Male 58 % Female 42%

3.4 Proposed Test Protocols

The database mainly considers the possible effect of various fake faces, challenge types and screen size. Therefore, it is necessary to provide a thorough analysis by verifying algorithms under different scenarios. Specifically, we designed a test protocol which could be grouped into two main categories, tablet format and phone format, based on the size of the screen that displayed the challenge. Each group contains three different challenge types; each challenge consisted of 3 scenarios. In total, we have had 18 scenarios.

Table **3.4** shows the summary of the database test protocol. The subset of KGDD database with 70 participants, who gave permission to make their data publicly available, can be split randomly into the training set (containing 42 subjects) and the testing set (containing 28 subjects). For each of the 18 scenarios, the corresponding data are to be selected from the training and testing set for model training and accuracy testing.

Table 3.4 The details of test protocols

Challenge type	**Scenario**	**Tablet**	**Scenario**	**Phone**
Lines	**1**	Photo attack test.	**10**	Photo attack test.
	2	2D mask attack test.	**11**	2D mask attack test.
	3	3D mask attack test.	**12**	3D mask attack test.
Curves	**4**	Photo attack test.	**13**	Photo attack test.
	5	2D mask attack test.	**14**	2D mask attack test.
	6	3D mask attack test.	**15**	3D mask attack test.
Points	**7**	Photo attack test.	**16**	Photo attack test.
	8	2D mask attack test.	**17**	2D mask attack test.
	9	3D mask attack test.	**18**	3D mask attack test.

3.5 Performance Metrics

Face presentation attack detection is a two-class classification problem. The performance of biometric-based authentication systems is generally represented in terms of these four possible classification results:

- **True positive (TP):** When a genuine attempt is correctly classified as genuine.
- **False positive (FP):** When a presentation attack attempt is incorrectly classified as genuine.
- **True negative (TN):** When a presentation attack attempt is correctly classified as a presentation attack.
- **False negative (FN):** When a genuine attempt is incorrectly classified as a presentation attack.

Based on these concepts, we can further define the following performance measurements: True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR) and False Negative Rate (FNR).

$$TPR = \frac{TP}{TP + FN} \tag{3.2}$$

$$TNR = \frac{TN}{FP + TN} \tag{3.3}$$

$$FPR = \frac{FP}{TN + FP} \tag{3.4}$$

$$FNR = \frac{FN}{TP + FN} \tag{3.5}$$

We also computed the recently standardised ISO/IEC 30107-3 metrics [1], namely the Attack Presentation Classification Error Rate (APCER) and the Bona Fide Presentation Classification Error Rate (BPCER), to evaluate the performance of the system.

- **Attack Presentation Classification Error Rate (APCER):** the proportion of attack presentations incorrectly classified as bona fide presentations in a specific scenario.

$$APCER = \frac{1}{N_{PA}} \sum_{i=1}^{N_{PA}} (1 - RES_i) \qquad (3.6)$$

where N_{PA} is the number of presentation attack, and RES_i takes 1 if the i^{th} presentation is classified as an attack presentation and value 0 if it is classified as a bona fide presentation.

- **Bona Fide Presentation Classification Error Rate (BPCER):** the proportion of bona fide presentations incorrectly classified as attack presentations in a specific scenario.

$$BPCER = \frac{\sum_{i=1}^{N_{BF}} RES_i}{N_{BF}} \qquad (3.7)$$

where N_{BF} is the number of bona fide presentation.

To summarise the performance of the overall system, the Average Classification Error Rate (ACER) is also computed. The ACER is defined as the average of APCER and BPCER and can be computed using (3.7).

$$ACER = \frac{APCER + BPCER}{2} \qquad (3.8)$$

To aid comparison with previously published works, we also report half total error rates (HTER) as defined in (3.8).

$$HTER(\tau, d) = \frac{FPR(\tau, d) + FNR(\tau, d)}{2} \qquad (3.9)$$

where τ is the threshold and d is the dataset. In this work, we report the minimum HTER achievable as the threshold is changed [32].

Receiver operating characteristic (ROC) plots are also used to present the trade-off between FPR and FNR for different values of the threshold.

3.6 Summary

This chapter covers the experimental system, challenge design, hardware setup and landmarks extraction and classification protocol. The challenge is a small shape following a randomised trajectory and presented on two screen sizes: tablet and phone formats. Three different challenge trajectories were designed: Lines, Curves and Points.

This chapter also presents the database (KGDD) that was collected for evaluation of face PAD algorithms. The KGDD consists of genuine access and attack attempts performed by 80 participants. In its presentation attacks, the KGDD includes three presentation attack instruments: displayed photo attacks, 2D mask attacks and 3D mask attacks. Each participant performed 18 attack attempts (using displayed photos, 2D masks and 3D masks in Phone and Tablet formats for three different challenges) and 12 genuine attempts (with or without tinted glasses in the phone and tablet formats for three different challenges) in total, creating 1440 sets of attack attempts and 960 sets of genuine attempts. A subset of this database, consisting of 70 participants, will be made public at http://kgddatabase.eda.kent.ac.uk. The test protocols are designed to evaluate the robustness of the developed face PAD methods. These protocols study the effect of the attack type, screen size and challenge type

The main contributions of this chapter are as follows.

- Challenge-response software design.
- An evaluation framework.
- The KGDD database.

In the next chapter, three gaze-based features that have been evaluated using the KGDD database for face PAD will be presented.

Chapter 4

Gaze-based Features for Detection of Presentation Attacks

4.1 Introduction

The work reported in this chapter explores a novel feature set for face presentation attack detection. The kinds of challenge described in Section 3.3.1 are employed here to direct user gaze. Here, the challenge moves randomly in different directions on the screen; the user's gaze is directed to these locations. The system records the gaze of the user by means of an ordinary webcam. Features based on measurement of the observed gaze are then extracted from facial images captured at each challenge location and used to discriminate between genuine attempts responding to the challenge and those of impostors. The proposed features are based on the assumption that spatial and temporal coordination of the eye movements involved in following the challenge are significantly different when a genuine attempt is made as compared to attack attempts of certain types.

To measure the similarity of eye movement and challenge movement trajectories, three gaze-based features are introduced: gaze correlation analysis, gaze alignment consistency, and gaze time dynamics features.

Performance of the system was tested using the KGDD dataset. Three presentation attack instruments were explored: displayed photo attack, 2D mask attack, and 3D mask attack. The effect of using tinted glasses for genuine user attempts was also investigated. Additionally, three different challenge types were explored: Lines, Curves, and Points. The effect of challenge duration on system performance was also considered, as challenge duration should ideally be kept to a minimum.

The rest of this chapter is organized as follows. The principle underlying the gaze correlation analysis and the experimental results of this feature are presented in Section 4.2. Section 4.3 describes the gaze alignment consistency, along with detailed experimental results. The gaze time dynamics features methodology and evaluations of this feature are reported in Section 4.4. Section 4.5 gives further analysis of proposed system, details computational time requirements for the gaze features and compares these approaches with other existing works. Section 4.6 presents observations and discussion. Finally, Section 4.7 includes a chapter summary and concluding remarks.

4.2 Gaze Correlation Analysis

The eye movements of a genuine user follow the same trajectory as the stimulus. By correlating a moving stimulus's trajectory with the eyes' trajectory, it is therefore possible to detect whether the stimulus is being looked at. The use of correlation implies that eye movements can be used with any size of screen, as the correlation coefficient inherently normalises the data; the gaze coordinates are not necessarily in the same range as the stimulus coordinates [153]. Here, the correlation feature was used to measure the similarity between pupil movement trajectory and object movement trajectory. The underlying hypothesis is that the correlation between eye movements and challenge movements should be greater for genuine user attempts and smaller for attack attempts. This phenomenon is then exploited to distinguish between genuine and impostor attempts.

4.2.1 Methodology

Figure 4.1 shows a block diagram of the proposed system. The visual stimulus (challenge) appears on the display screen and moves in different directions. The user is required to track the moving stimulus across the screen with their gaze, using natural head and eye movements. The camera captures the facial images at each location of the challenge. Then the facial landmarks in the captured frames are extracted. From these landmarks, various features are computed and then used to classify the presentation attempt as a genuine attempt (i.e. coming from a live sample) or a presentation attack attempt (i.e. coming from an impostor presenting a displayed photo, 2D mask or 3D mask attack). After the pre-processor (Chehra software) extracts eye points, the crucial task is to effectively measure the similarity between eye movement and stimulus movement trajectories. The essential problem can be defined as how to measure the similarity between two lines. It is anticipated that the genuine user's eye movement trajectory will be similar to that of the moving stimulus. As each eye point contains two variables (*x* and *y* coordinates), the correlation between eye movement and stimulus movement has to be calculated separately for *x* coordinates and *y* coordinates.

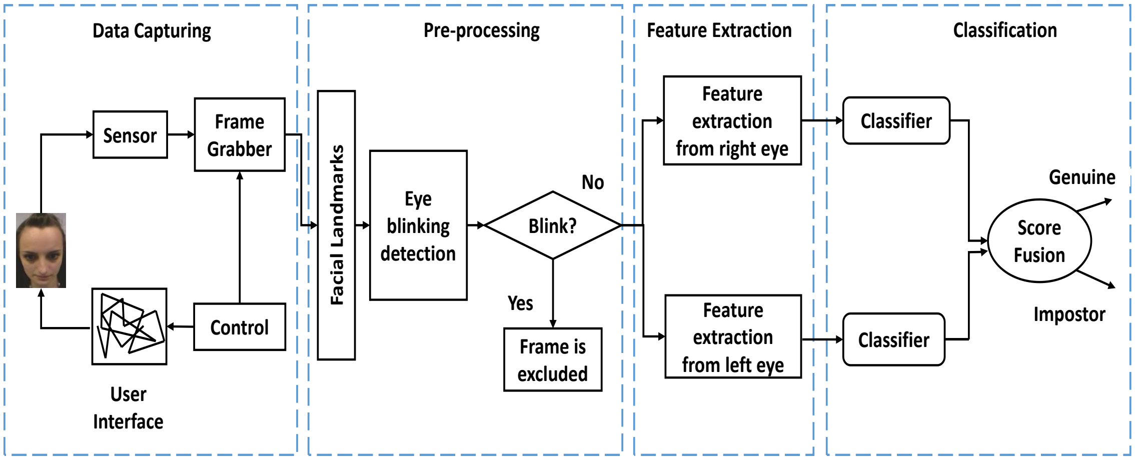
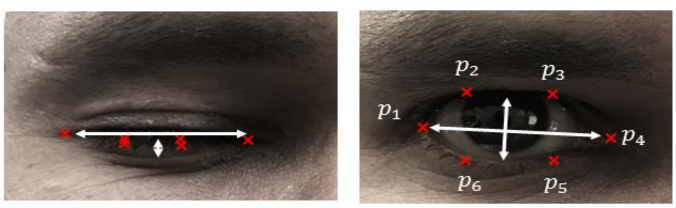


Figure 4.1: Diagram of the proposed system

As gaze correlation analysis rely on the eye being open in order to locate the centre of the eye, it is important to detect blinks; frames that include the blinking eye can then be excluded before calculating the feature [154]. The upper and lower eyelid positions were used to detect blinking. The eye aspect ratio (EAR) (between the height and width of the eye) was calculated for each frame as follows [155]:

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \qquad (4.1)$$



(a) (b)

Figure 4.2: Examples of (a) closing eye and (b) opening eye

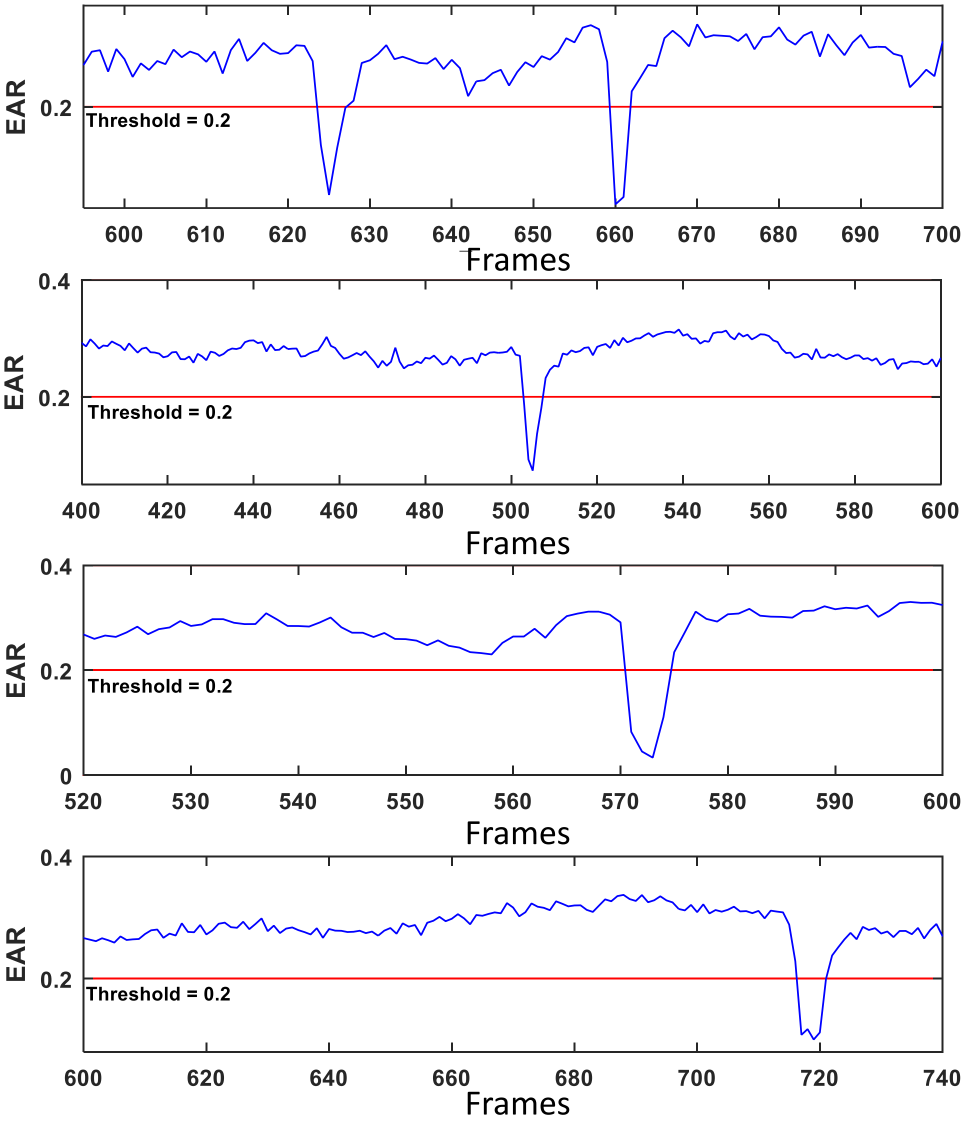


Figure 4.3: Eye aspect ratio (EAR) plotted for several frames from four videos

where $p_1, . . . , p_6$ are the 2D landmark locations as can be seen in Figure 4.2 and $||$ is the norm.

$$Decision \ = \begin{cases} Eye\ closed, & if\ EAR \leq threshold; \\ \\ Eye\ open, & otherwise. \end{cases} \tag{4.2}$$

The EAR values are mostly constant when the eye is open and approaches zero at the time of blink. The threshold (0.2) was empirically determined. Some examples of EAR values in the video sequences are shown in Figure 4.3.

After detecting blinks, corresponding frames with closed eyes were excluded, and the remaining frames had been used in the correlation feature-extraction step.

The correlation coefficient calculation for horizontal dimension x are described in equation (4.3 - 4.6); the same computation applies for the vertical dimension y as well. Similarity of gaze and challenge trajectories calculated using Pearson's product-moment correlation coefficient [156] for the horizontal coordinates r_x between the challenge PAD_x and the eye Eye_x, is defined as

$$r_x = \frac{cov(Eye_x, PAD_x)}{\sigma_{Eye_x}\sigma_{PAD_x}} \tag{4.3}$$

$$r_x = \frac{\sum_{i=1}^{N}\left(Eye_{x_i} - \overline{Eye_x}\right)\left(PAD_{x_i} - \overline{PAD_x}\right)}{\sigma_{Eye_x}\sigma_{PAD_x}} \tag{4.4}$$

where $\overline{Eye_x}$ and σ_{Eye_x} are the mean and standard deviation of the horizontal gaze positions.

$$\overline{Eye_x} = \frac{1}{N}\sum_{i=1}^{N} Eye_{x_i} \tag{4.5}$$

$$\sigma_{Eye_x} = \sqrt{\frac{\sum_{i=1}^{N}(Eye_{x_i} - \overline{Eye_x})^2}{N}} \tag{4.6}$$

These measurements are also computed for coordinates of the object, where $\overline{PAD_x}$ and σ_{PAD_x} are the mean and standard deviation of the horizontal challenge location.

$$\overline{PAD_x} = \frac{1}{N}\sum_{i=1}^{N} PAD_{x_i} \tag{4.7}$$

$$\sigma_{PAD_x} = \sqrt{\frac{\sum_{i=1}^{N}(PAD_{x_i} - \overline{PAD_x})^2}{N}} \tag{4.8}$$

The coefficient is calculated for both horizontal coordinates, r_{x}, and vertical coordinates, r_{y}. The closer the coefficient's value is to 1, the more correlated are the two coordinates and the more similar the eyes' trajectory to the object's trajectory.

These two correlation coefficients are concatenated to form the feature vector F_{corr}.

$$F_{corr} = [\, r_{\mathrm{x}} \,, r_{\mathrm{y}}] \tag{4.9}$$

A *k*-NN based multi-classifier system is then used for presentation attack detection as shown in Figure 4.1.

4.2.2 Experimental Evaluation of Gaze Correlation Analysis

The experimental analysis was performed using the KGDD database (as described in Chapter 3). In this section, performance was reported for a subset from the KGDD database with 70 participants, who gave us permission to make their data publicly available. Performance of the proposed features on the entire dataset of 80 participants is provided in Appendix E.

Several sets of experiments were performed to investigate the performance of the proposed features, which were extracted from the data captured during **Lines**, **Curves,** and **Points** challenges in **Tablet** and **Phone** formats. The first set of experiments involved genuine presentation (without tinted glasses) against all attack scenarios for Tablet format. A further set of experiments followed to investigate whether this feature is effective if tinted glasses are worn by genuine users while accessing the device. The subsequent set of experiments was performed to explore the same feature for a much smaller screen size (Phone format). The purpose of these experiments was to investigate whether this feature would be adequate for presentation attack detection in the smaller mobile device formats.

The results presented in this section are organized as follows:

- Performance of gaze correlation analysis using **Tablet** format (analysed in Section 4.2.2.1)
- Performance of gaze correlation analysis using **Phone** format (presented in Section 4.2.2.2)

4.2.2.1 Tablet Format

Following the feature extraction step, a *k*-nearest neighbour (*k*-NN) classifier was applied (*k* = 1) to discriminate presentation attack attempts from genuine attempts. The classification protocol was as follows. Features are extracted from each eye and then passed to separate classifiers (1-NN) to obtain individual classification scores for each eye. These scores are then combined using score fusion technique to obtain the final PAD outcome as mentioned in Section 3.2.3. Different fusion schemes were explored. Compared to other fusion schemes (see Table 4.1), the product rule [152] achieved the lowest ACER (in (3.8)) and was therefore used in all of the subsequent experiments reported in this thesis.

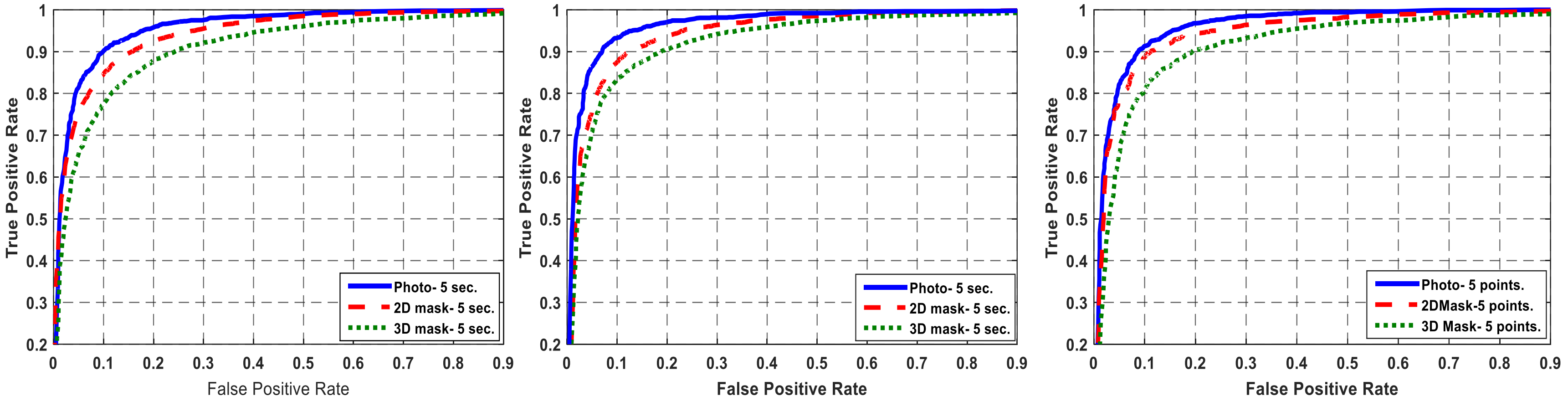
Table 4.1: Comparison of various score fusion schemes' performance using the entire database (80 participants), captured during Lines challenge of five-second duration

	Rule-based fusion	**APCER = 0.01**		**APCER = 0.05**		**APCER = 0.10**	
		BPCER	ACER	BPCER	ACER	BPCER	ACER
Lines	Product	0.43	**0.22**	0.19	**0.12**	0.04	**0.07**
	Sum	0.49	0.25	0.21	0.13	0.08	0.09
	Majority vote	0.61	0.31	0.33	0.19	0.18	0.14
Curves	Product	0.33	**0.17**	0.11	**0.08**	0.02	**0.06**
	Sum	0.39	0.20	0.27	0.16	0.12	0.11
	Majority vote	0.47	0.24	0.33	0.19	0.16	0.13
Points	Product	0.43	**0.22**	0.15	**0.10**	0.04	**0.07**
	Sum	0.51	0.26	0.25	0.15	0.10	0.10
	Majority vote	0.53	0.27	0.27	0.16	0.14	0.12

Table 4.1 specifies the performances (in the ISO metrics) for the entire dataset, with 80 participants using various score fusion schemes.

Figure 4.4 shows the performance for three attack scenarios—displayed photos, 2D masks, and 3D masks—for the **Tablet** format, using gaze correlation features extracted from the dataset (70 participants) during **Lines**, **Curves,** and **Points** challenges of five-second duration. At FPR = 10%, TPRs were 90%, 94%, and 91% for displayed photo attack using Lines, Curves, and Points challenges, respectively. However, these dropped to 85%, 88%, and 89% when using the 2D mask. For 3D mask attack detection, TPRs were 79%, 83%, and 80% using Lines, Curves, and Points challenges, respectively. These results confirm that superior performance can be achieved by using this feature to detect photo attack for all three challenge types, possibly because the impostor can follow the challenge target more accurately when looking through the small holes in the masks as compared to a displayed photo attack, in which the participant needs to move the photo with their hands. Similar trends were observed at other FPR settings as well.

A similar set of tests was conducted in which the dataset was captured from genuine users wearing **tinted glasses**. The purpose of this set of experiments was to investigate the system's performance when genuine users attempted to access the system while wearing tinted glasses. This scenario was devised to simulate the real-world use of tinted glasses during outdoor operation. Figure 4.5 shows the ROC curves for displayed photo, 2D mask, and 3D mask attacks, using the proposed feature extracted from the dataset (70 participants) and captured during **Lines**, **Curves,** and **Points** challenges of five-second duration. At 10% FPR, TPRs were about 41%, 20%, and 32% for displayed photo, 2D mask and 3D mask, respectively, during a Lines challenge. For Curves challenges, displayed photo attack detection ranked top while 3D mask attack detection ranked second, and 2D mask ranked third, with TPRs of 41%, 31%, and 40%, respectively. At 10% FPR, TPRs of 37%, 39%, and 36% were achieved for displayed photo, 2D mask, and 3D mask attacks, respectively, when using a Points challenge. It is clear that the proposed method does not perform very well when users are wearing tinted glasses.



(a)

(b)

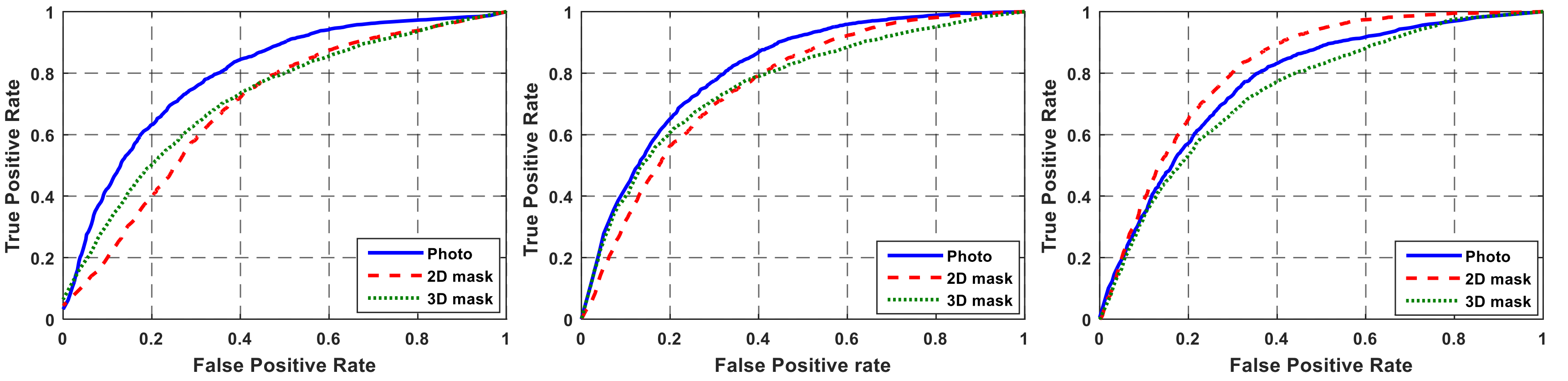
(c)

Figure 4.4: ROC curves for photo, 2D mask, and 3D mask in Tablet format, using data captured during (a) Lines, (b) Curves, and (c) Points challenges for gaze correlation analysis

Table 4.2: Performance of the system at APCER = 0.05 for various challenge types using gaze correlation analysis

Attack type	**Lines**		**Curves**		**Points**	
	APCER = 0.05**		**APCER = 0.05****		**APCER = 0.05****	
	BPCER	**ACER**	**BPCER**	**ACER**	**BPCER**	**ACER**
Photo	0.19	0.12	0.12	0.08	0.17	0.11
2D mask	0.25	0.15	0.21	0.13	0.20	0.12
3D mask	0.32	0.18	0.30	0.17	0.30	0.17
Overall	0.32	0.18	**0.30**	**0.17**	**0.30**	**0.17**

** Using the 70 publicly available participants in the KGDD (42 for training and 28 for testing)



(a) (b) (c)

Figure 4.5: ROC curves for photo, 2D mask, and 3D mask in Tablet format, using data captured during (a) Lines, (b) Curves, and (c) Points challenges, using tinted glasses and gaze correlation analysis

Table 4.3: Performance of the system at APCER = 0.05 for various challenge types using tinted glasses and gaze correlation analysis

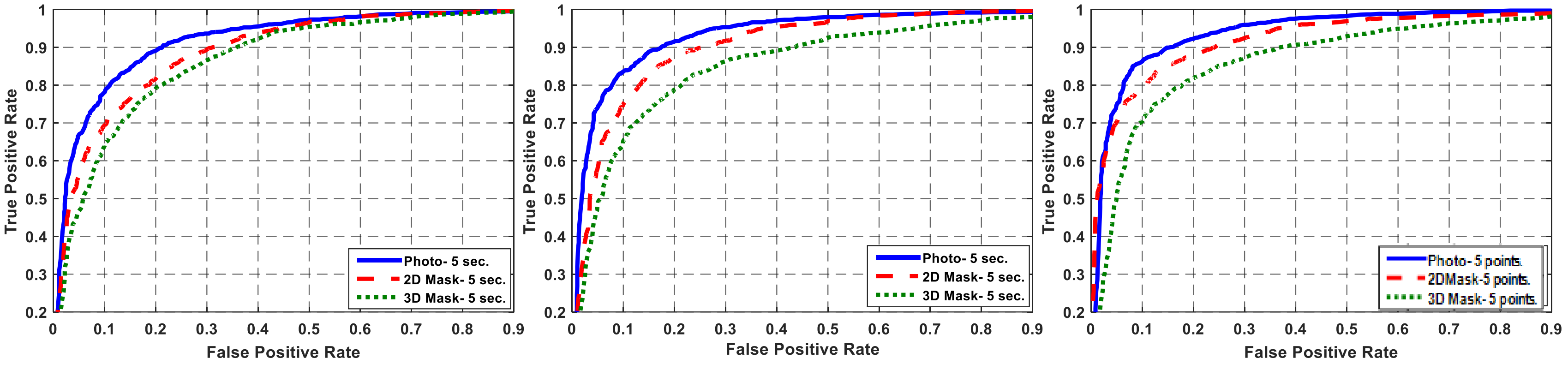
Attack type	**Lines**		**Curves**		**Points**	
	APCER = 0.05**		**APCER = 0.05****		**APCER = 0.05****	
	BPCER	**ACER**	**BPCER**	**ACER**	**BPCER**	**ACER**
Photo	0.75	0.40	0.69	0.37	0.80	0.42
2D mask	0.87	0.46	0.82	0.43	0.81	0.43
3D mask	0.80	0.42	0.70	0.38	0.85	0.42
Overall	0.87	0.46	0.82	0.43	**0.85**	**0.42**

Tables 4.2 and 4.3 summarise overall performances (using the newly standardized APCER and BPCER metrics as defined in the ISO/IEC CD 30107-3 standard) when the proposed features were tested on the KGDD dataset of 70 subjects, which will be publicly available.

4.2.2.2 Phone Format

Experiments similar to those with Tablet format devices were performed to explore the same feature in a much smaller screen size (the Phone platform). The purpose of these experiments was to investigate whether this feature would be adequate for PAD in the smaller mobile handsets. The ROC curves for displayed photo, 2D mask, and 3D mask attack detection in **Phone** format using gaze correlation analysis are presented in Figure 4.6; data were again captured during **Lines, Curves,** and **Points** challenges of five-second duration. As in the case of the Tablet format, a displayed photo attack is easier to detect in all three challenge types when using the proposed feature. The 2D and 3D mask attacks were challenging and difficult to discriminate from genuine attempts. At FPR = 10%, displayed photo detection TPRs were about 79%, 82%, and 86% for **Lines, Curves,** and **Points** challenges, respectively. TPRs for 2D mask attack detection were 70%, 73%, and 89%, dropping to 62%, 65%, and 70% when the proposed features were used to detect 3D masks during **Lines, Curves,** and **Points** challenges, respectively. Generally, it can be found that the performance of displayed photo detection is better compared to other PAIs. This is similar to the analysis in section 4.2.2.1.

The system was also tested using data captured from users wearing tinted glasses when performing genuine presentations in Phone format. As shown in Figure 4.7, at 10% FPR, displayed photo detection TPR was about 48%; 2D mask attack detection TPR was 38%, and 3D mask attack detection was about 21% for data captured during Lines challenges. For Curves challenges, the TPRs were 20%, 29%, and 18% at 10% FPR for displayed photo attack, 2D mask attack, and 3D mask attack, respectively. At 10% FPR, TPRs were 60%, 48%, and 30% for displayed photo, 2D mask, and 3D mask for five-second Points challenges. Tables 4.4 and 4.5 summarise overall performance (using the newly standardized APCER and BPCER metrics, defined in the ISO/IEC CD 30107-3 standard).



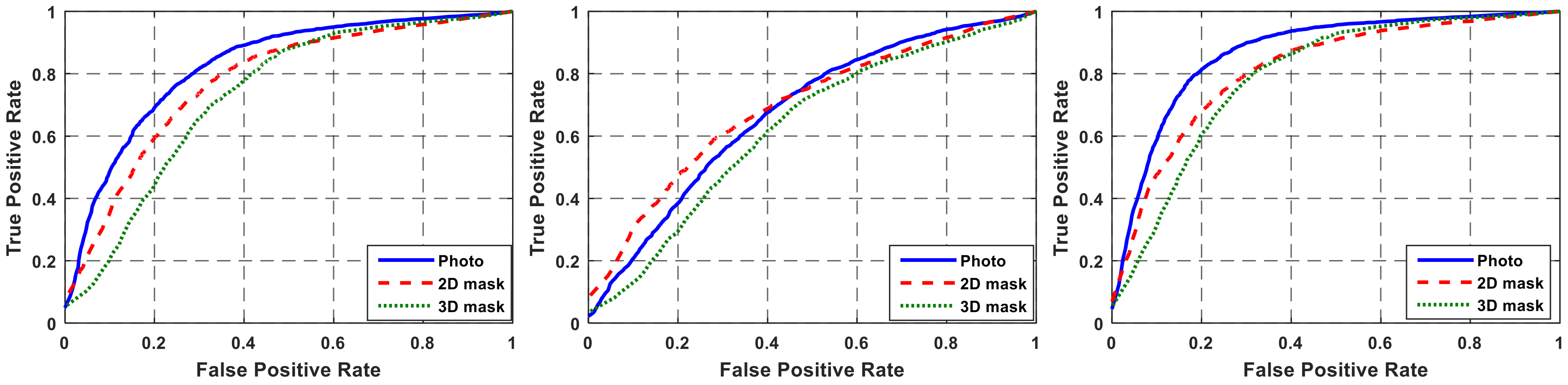
(a) (b) (c)

Figure 4.6: ROC curves for photo, 2D mask, and 3D mask in Phone format using data captured during (a) Lines, (b) Curves, and (c) Points challenges for gaze correlation analysis

Table 4.4: Performance of the system at APCER = 0.05 for various challenge types using gaze correlation analysis

Attack type	**Lines**		**Curves**		**Points**	
	APCER = 0.05**		**APCER = 0.05****		**APCER = 0.05****	
	BPCER	**ACER**	**BPCER**	**ACER**	**BPCER**	**ACER**
Photo	0.33	0.18	0.24	0.14	0.25	0.15
2D mask	0.40	0.22	0.35	0.20	0.30	0.17
3D mask	0.50	0.28	0.50	0.27	0.48	0.26
Overall	0.51	0.28	0.50	0.27	**0.48**	**0.26**

** Using the 70 publicly available participants in the KGDD, 42 for training and 28 for testing



(a) **(b)** **(c)**

Figure 4.7: ROC curves for photo, 2D mask and 3D mask in Phone format using data captured during (a) Lines, (b) Curves, and (c) Points challenge using tinted glasses for gaze correlation analysis

Table 4.5: Performance of the system at APCER = 0.05 for various challenge types using tinted glasses and gaze correlation analysis

Attack type	Lines		Curves		Points	
	APCER = 0.05**		APCER = 0.05**		APCER = 0.05**	
	BPCER	ACER	BPCER	ACER	BPCER	ACER
Photo	0.65	0.35	0.87	0.46	**0.58**	**0.31**
2D mask	0.77	0.41	0.80	0.42	0.60	0.32
3D mask	0.89	0.47	0.90	0.47	0.78	0.41
Overall	0.89	0.47	0.90	0.47	0.78	0.41

** Using the 70 publicly available participants in the KGDD (42 for training and 28 for testing)

4.3 Gaze Alignment Consistency

This section introduces a new feature based on gaze alignment consistency for detecting presentation attacks. Alignment consistency features are extracted from sets of images captured when the stimulus is on a given line. This novel feature described here is designed to capture the angular similarity, for each line segment, between the stimulus trajectory and the trajectory of the participant’s pupil movement, which are then used for PAD classification.

4.3.1 Methodology

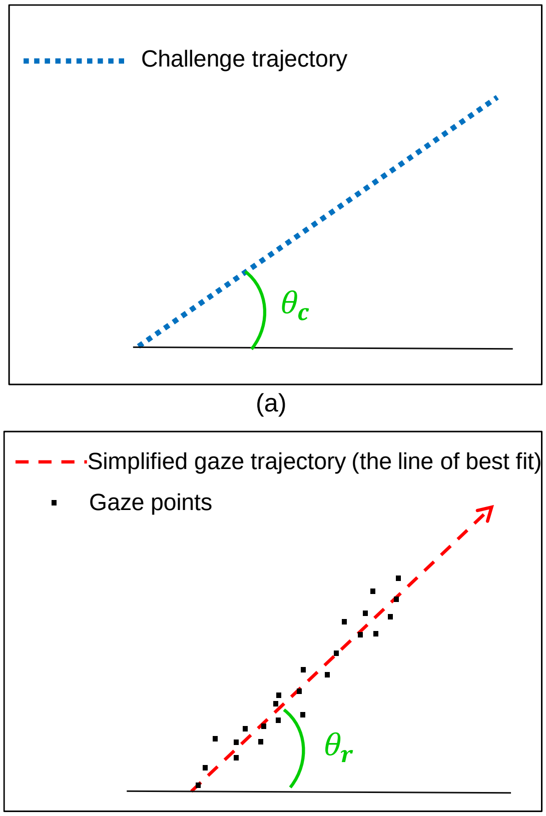
The key idea of the gaze-based feature is that the eye movements of a genuine user are expected to be similar to that of the stimulus being tracked. The alignment of eye movement trajectory and challenge trajectory is measured by computing the angular differences between the two.

Let (Obj_x, Obj_y) be a point on the challenge trajectory where the stimulus moves along a straight-line segment, with (Eye_x, Eye_y) is the corresponding pupil centre. Let θ_c be the angle of the challenge, using the start and end points i and j along the trajectory of the line. Then,

$$\theta_c = tan^{-1}\left(\frac{Obj_{y_j} - Obj_{y_i}}{Obj_{x_j} - Obj_{x_i}}\right) \tag{4.10}$$

For the response trajectory, let θ_r be the angle of the line of best fit calculated using the Least Squares regression method; then

$$\theta_r = tan^{-1}\left(\frac{\sum_{i=1}^{N}(Eye_{x_i} - \overline{Eye_x})(Eye_{y_i} - \overline{Eye_y})}{\sum_{i=1}^{N}(Eye_{x_i} - \overline{Eye_x})^2}\right) \tag{4.11}$$



(a)

(b)

Figure 4.8: How the system compares (a) challenge and (b) gaze trajectories

where $\overline{Eye_x}$ and $\overline{Eye_y}$ are the means of all Eye_{x_i} and Eye_{y_i} values, respectively, and N is the number of frames acquired for each line segment. Figure 4.8 illustrates how the system compares gaze and challenge trajectory.

The feature vector F then consists of the differences between these two angles for each of the line segments included in the challenge:

$$\Delta\theta_v = |\theta_{cv} - \theta_{rv}| \qquad (4.12)$$

$$F = [\Delta\theta_1, \Delta\theta_2, \dots, \Delta\theta_v] \qquad (4.13)$$

where v is the number of line segments used in each attempt.

4.3.2 Experimental Evaluation of Gaze Alignment Consistency

To evaluate the proposed method, gaze-based alignment consistency were extracted for both eyes from the data collected during presentation of the stimulus in the **Lines** challenge. The evaluation framework shown in Figure 4.9 were used to conduct experiment to explore the performance of the gaze alignment consistency. Gaze alignment features were extracted from both eyes and were then passed to separate classifiers for PAD testing using this framework. The final outcome is through the fusion of these two outputs. Performance is reported for the subset of KGDD database with 70 participants who gave permission to make their data publicly available. Performance of the proposed features on the entire dataset of 80 participants is provided in Appendix E.

The performance of the proposed alignment features was investigated for genuine presentation (without tinted glasses) for all attack scenarios. A subsequent set of experiments investigated whether this feature is effective for PAD if tinted glasses are used by genuine users while accessing the device.

The results in this section are organized as follows.

- Performance of the gaze correlation analysis using **Tablet** format is analysed in Section 4.3.2.1.
- Performance of the gaze correlation analysis using **Phone** format is presented in Section 4.3.2.2.

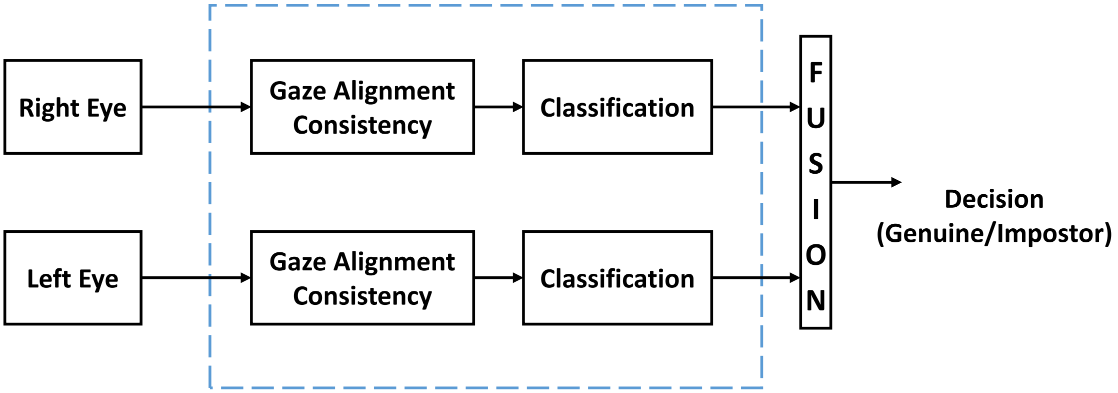

Right Eye
Left Eye
Gaze Alignment Consistency
Gaze Alignment Consistency
Classification
Classification
FUSION
Decision (Genuine/Impostor)

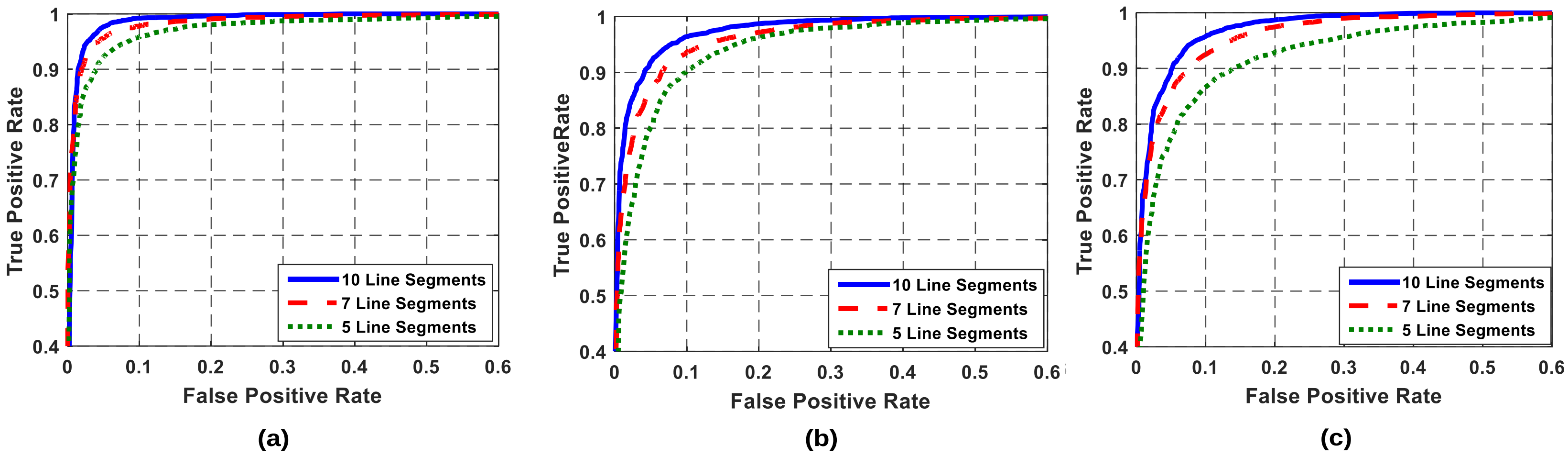
Figure 4.9: Score fusion using features extracted from left and right eye

4.3.2.1 Tablet format

Figure 4.10 shows performance of the system for the three attack scenarios—displayed photo (Figure 4.10(a)), 2D mask (Figure 4.10(b)) and 3D mask (Figure 4.10(c))—for **Tablet** format, using gaze alignment consistency tested on the KGDD subset of 70 participants. System performance was found to be lower in the case of 3D mask detection than for the other attack scenarios; this may be because it is easier for the attacker to adjust the mask to fit their eye positions and so track the stimulus more accurately. The system TPRs were 99%, 97%, and 96% at 10% FPR for photo, 2D mask, and 3D mask, respectively when using 10 line segments. Performance dropped to 96%, 90%, and 86% when using 5 line segments for the proposed feature. In general, system performance increased with number of line segments. This observation suggests that the proposed features that is based on a longer challenge duration may provide more information to achieve attack detection.

A further set of experiments was conducted in which genuine users wore tinted glasses. Figure 4.11 shows the ROC curves for displayed photo, 2D mask, and 3D mask attacks using the proposed scheme with different line segments. At 10% FPR, TPRs were about 48%, 61%, and 40% for displayed photo, 2D mask, and 3D mask, respectively, when using 5 line segments. In general, these results suggest that system performance is poor when genuine users wear tinted glasses. This may be because (1) the tinted glasses slightly cover the eye centres and (2) the reflection of the white screen on the glasses may affect the facial landmark detection software.

Tables 4.6 and 4.7 summarise overall performance (using the newly standardized APCER and BPCER metrics, defined in the ISO/IEC CD 30107-3 standard) when the proposed features were tested on the KGDD datasets with 70 participants.



(a) (b) (c)

Figure 4.10: ROC curves for (a) displayed photos, (b) 2D masks, and (c) 3D masks in Tablet format based on data captured for the Lines challenge using gaze alignment consistency

Table 4.6: Performance of the system at APCER = 0.05 for the Lines challenge using gaze alignment consistency

Attack type	5 Line Segments		7 Line Segments		10 Line Segments	
	APCER = 0.05**		APCER = 0.05**		APCER = 0.05**	
	BPCER	ACER	BPCER	ACER	BPCER	ACER
Displayed photo	0.09	0.07	0.05	0.05	**0.03**	**0.04**
2D mask	0.18	0.11	0.12	0.08	0.08	0.06
3D mask	0.23	0.14	0.16	0.10	0.11	0.08
Overall	0.23	0.14	0.16	0.10	**0.11**	**0.08**

** Using the 70 publicly available participants in the KGDD (42 for training and 28 for testing)

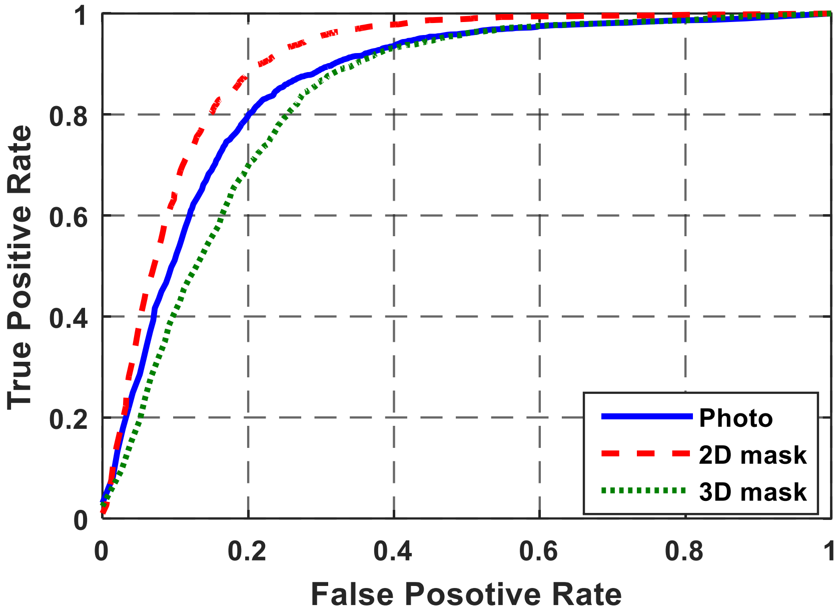


Figure 4.11: ROC curves for displayed photos, 2D masks, and 3D masks in Tablet format based on data captured for the Lines challenge using gaze alignment consistency and tinted glasses

Table 4.7: Performance of the system for the Lines challenge using gaze alignment consistency and tinted glasses

Attack type	APCER = 0.05**		APCER = 0.10**	
	BPCER	ACER	BPCER	ACER
Photo	0.71	0.36	0.50	0.30
2D mask	0.60	0.32	0.39	0.24
3D mask	0.80	0.42	0.60	0.35
Overall	0.80	0.42	**0.60**	**0.35**

** Using the 70 publicly available participants in the KGDD (42 for training and 28 for testing)

4.3.2.2 Phone format

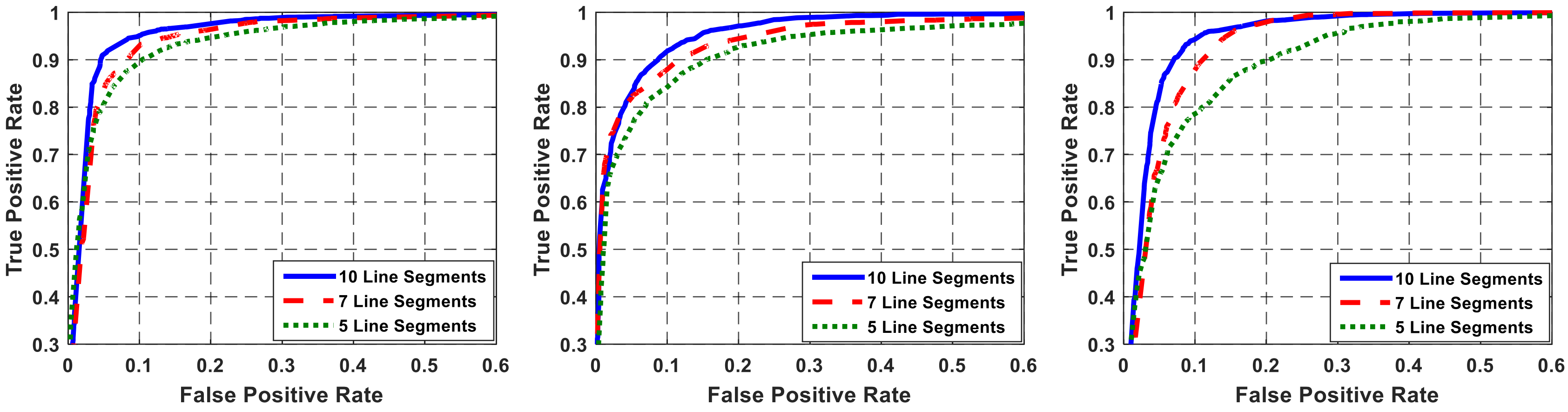
This subsection reports experiments performed to estimate the proposed gaze alignment feature's ability to distinguish between genuine and presentation attack attempts in a much smaller screen size format (**Phone** format).

Figure 4.12 presents the ROC curves showing feature performance when tested on the KGDD subset of 70 participants.

Performance improved as the number of line segments used to extract the feature vector increased, up to 10 line segments. As they were randomly generated, each line varied in length, but the average time was about 1 second per line. For 5 and 10 line segments, system performance for 3D mask attack detection was lower than for displayed photo and 2D mask attack detection. System TPRs were 96%, 91%, and 95% at 10% FPR for displayed photo, 2D mask, and 3D mask, respectively, when using sets of 10 line segments. Performance fell to 91%, 84%, and 80% when only 5 line segments were used for the proposed feature. The reason is similar to the one stated in the above analysis in section 4.3.2.1, where the proposed features based on a longer challenge duration may provide more information to discriminate between genuine and impostor attempts. From Figure 4.12, it is observed that the performance of displayed photo detection is better compared to other PAIs. A possible reason is that the impostor can follow the challenge easier when looking through the small holes in the masks.

The effect of wearing tinted glasses on the proposed feature's accuracy was also evaluated. Figure 4.13 shows ROC performance of the alignment consistency when tested on the KGDD subset of 70 participants. From the figure it can be seen that the proposed method does not perform very well when users are wearing tinted glasses (TPR = 40%, TPR = 24%, and TPR = 37% @ FAR = 0.10 for displayed photo, 2D mask, and 3D mask, respectively, when using 5 line segments).

Tables 4.8 and 4.9 summarise overall performance (using the newly standardized APCER and BPCER metrics, defined in the ISO/IEC CD 30107-3 standard) when the proposed features were tested on the KGDD database with 70 participants.



(a) (b) (c)

Figure 4.12: ROC curves for (a) photos, (b) 2D masks, and (c) 3D masks in Phone format using data captured for Lines challenge gaze alignment consistency

Table 4.8: Performance of the system at APCER = 0.05 for various challenge types using gaze alignment consistency

Attack type	**5 Line Segments**		**7 Line Segments**		**10 Line Segments**	
	APCER = 0.05**		**APCER = 0.05****		**APCER = 0.05****	
	BPCER	**ACER**	**BPCER**	**ACER**	**BPCER**	**ACER**
Photo	0.20	0.12	0.16	0.10	**0.09**	**0.07**
2D mask	0.34	0.19	0.30	0.17	0.15	0.10
3D mask	0.32	0.18	0.27	0.16	0.16	0.11
Overall	0.34	0.19	0.30	0.17	**0.16**	**0.11**

** Using the 70 publicly available participants in the KGDD (42 for training and 28 for testing)

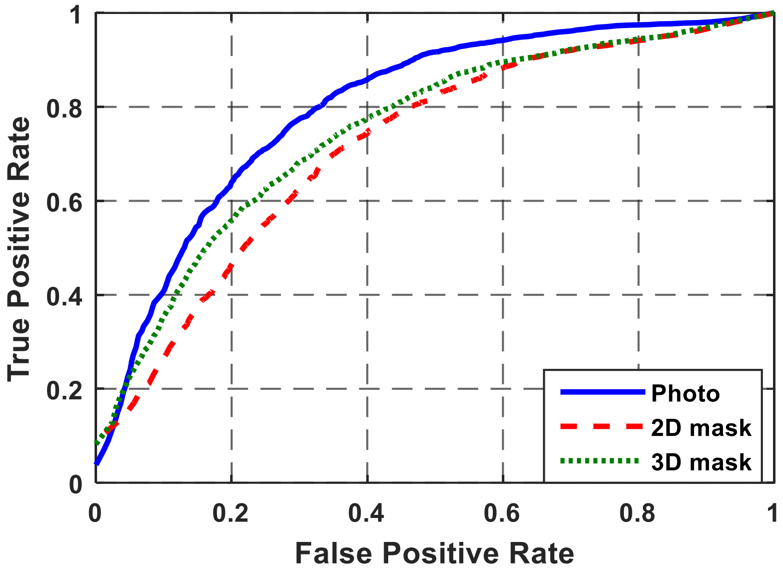


Figure 4.13: ROC curves for photos, 2D masks, and 3D masks in Phone format based on data captured for Lines challenge using gaze alignment consistency and tinted glasses

Table 4.9: Performance of the system for various challenge types using gaze alignment consistency and tinted glasses

Attack type	APCER = 0.05**		APCER = 0.10**	
	BPCER	ACER	BPCER	ACER
Photo	0.75	0.40	0.60	0.35
2D mask	0.81	0.43	0.70	0.37
3D mask	0.77	0.41	0.62	0.35
Overall	0.81	0.43	**0.70**	**0.37**

** Using the 70 publicly available participants in the KGDD (42 for training and 28 for testing)

4.4 Gaze Time Dynamics Feature

This section describes a third gazed-based feature: gaze time dynamics feature. The previous feature (Section 4.3) was extracted from the coordinates of eye centre. In gaze time dynamics, time is considered in feature extraction step. This feature is extracted using dynamic time warping (DTW). DTW is an algorithm that searches for optimal alignment of two temporal sequences, returning a distance measure of their similarities [157]. Here, the DTW measure is used to assess the similarity between the gaze trajectory and the challenge trajectory following optimal DTW alignment. The idea behind the proposed feature is that an impostor cannot align the PAIs for the same location as accurately as a genuine user. The underlying hypothesis is that the similarity between gaze trajectory and challenge trajectory should be greater in genuine user attempts and small for presentation attack attempts. This phenomenon is exploited to distinguish between genuine users' and impostors' attempts.

4.4.1 Methodology

As before, a small shape (stimulus) was presented to the user while seated in front of the computer screen, and they were instructed to follow its movement across the screen using natural head and eye movements. Facial images were captured at each challenge location, and pupil coordinates were extracted from each image. These coordinates were then used to extract gaze time dynamics features. The proposed gaze time dynamics features have been extracted when the stimulus was moving in **Lines** and **Curves** trajectories. The feature vector was then passed to the classifier to discriminate between genuine and impostor attempts.

As the spatial coordinates of landmarks for each session and challenge trajectory are in different range, they were normalized using the Min-Max technique [139] prior to feature extraction. Let x_i and $\widehat{x_i}$ denote the original and normalized x coordinates, respectively, and let y_i and $\widehat{y_i}$ denote the original and normalized y coordinates, respectively. The normalized values are:

$$\widehat{x_i} = \frac{x_i - \min x}{\max x - \min x} \tag{4.14}$$

$$\hat{y_\imath} = \frac{y_i - \min \mathrm{y}}{\max \mathrm{y} - \min \mathrm{y}} \tag{4.15}$$

To measure similarity, DTW was used to compute the warping distance between the two trajectories.

Suppose that the input response (gaze) sequence of *x* coordinates is denoted as *R$_x$(i)*, *i*=1,….,*n*, and the challenge sequence of x coordinates is denoted as *C$_x$(j)*, *j*=1,….,*m*, where n and m represent the number of frames and challenge locations, respectively and *n* = *m*; DTW distance is then defined as the minimum distance from the beginning of the DTW table to the current position (*i; j*) [158]. The DTW table $D(i,j)$ can be calculated as

$$DTW(R,C) = D(i,j) = d(i,j) + min \begin{cases} D(i-1,j) \\ D(i,j-1) \\ D(i-1,j-1) \end{cases} \tag{4.16}$$

where $D(i,j)$ is the node cost associated with $R(i)$ and $C(j)$ as defined in

$$d(i,j) = (R(i) - C(j))^2 \tag{4.17}$$

As challenge and gaze are specified by two coordinates (*x* and *y*), we performed DTW separately for each coordinate, and the summed distances of x and y coordinates were used to measure similarity of a sequence pair [158, 159].

DTW_L denotes the cumulative distances of both coordinates measured independently under DTW [142]. If $DTW(R_x, C_x)$ is the DTW distance of the *x* coordinates of R and C, and $DTW\left(R_y, C_y\right)$ is the DTW distance of the *y* coordinates of *R* and *C*, DTW_L can be expressed as

$$DTW_L(R,C) = \; DTW(R_x, C_x) + \; DTW\left(R_y, C_y\right) \tag{4.18}$$

In the above equation, each dimension is considered independent, and DTW is allowed the freedom to warp each dimension independently of the others.

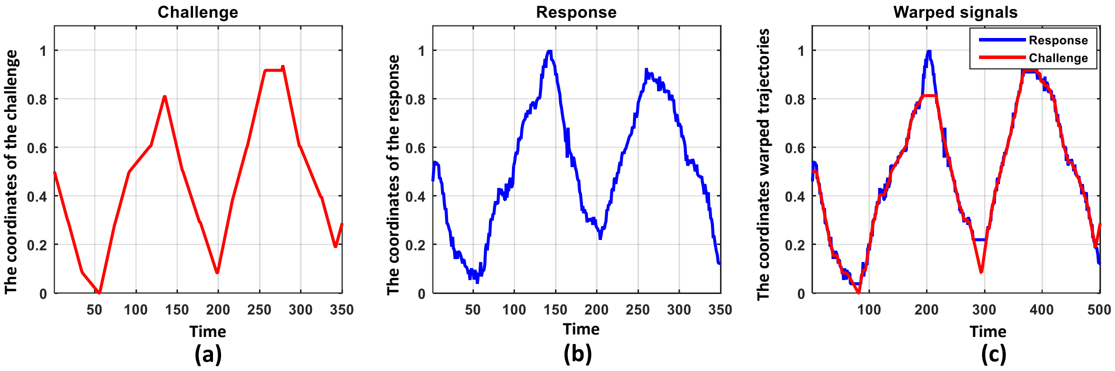


Figure 4.14: (a) x coordinates of challenge trajectory; (b) x coordinates of response by a genuine user; (c) both trajectories aligned with DTW

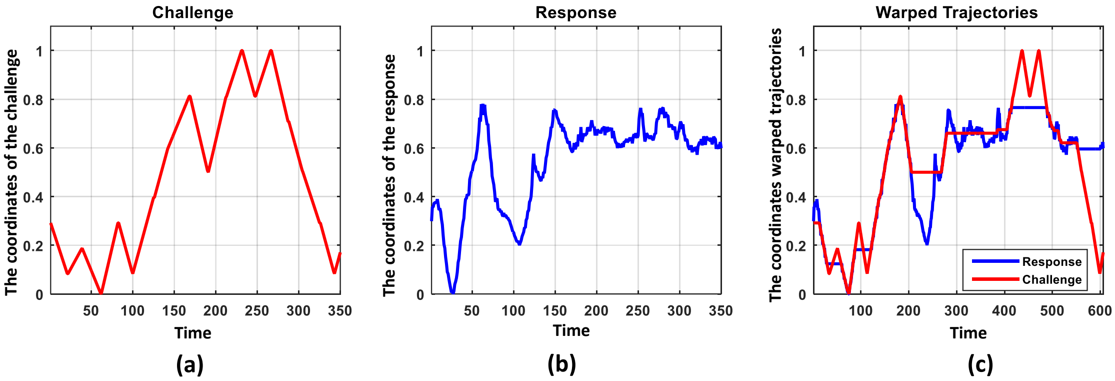


Figure 4.15: (a) x coordinates of challenge trajectory; (b) x coordinates of response by an impostor using displayed photo; (c) both trajectories aligned with DTW

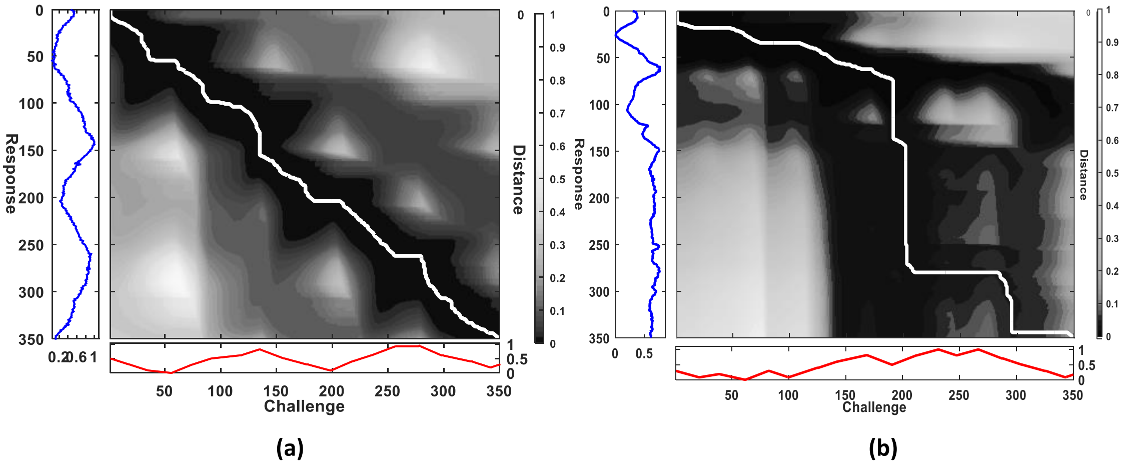


Figure 4.16: Accumulated distances (cost) matrix of challenge and response and optimal warping paths (white line): (a) genuine attempt; (b) attack attempt

Figure 4.14 illustrates an example of two trajectories C (challenge trajectory) and R (response or gaze trajectory of a genuine user). Figure 4.15 shows the two trajectories for an impostor response using displayed photo attack. In Figure 4.16, optimal paths are indicated by the solid white line.

In this way, a feature vector is formed, based on the distance calculated using DTW. As gaze DTW distances were extracted from the left eye (DTW_l) and right eye (DTW_r), feature vectors can be constructed for the right eye F_{DTW_r} and left eye F_{DTW_l} for use in presentation attack detection:

$$F_{DTW_r} = [DTW_r] \tag{4.19}$$

$$F_{DTW_l} = [\ DTW_l] \tag{4.20}$$

The two feature vectors are fed to the *k*-NN classifiers and their outputs are fused to get the final result of presentation attack detection as shown in Figure 4.17.

4.4.2 Experimental Evaluation of Gaze Time Dynamics Features

Several sets of experiments were conducted to explore the performance of gaze time dynamics features. The typical evaluation framework is shown in Figure 4.17. Gaze time dynamics features were extracted using both eyes and were then passed to separate classifiers for training and testing using this framework. In this section, performance was reported for a subset from the KGDD database with 70 participants, who gave us permission to make their data publicly available. Performance of the proposed features on the entire dataset of 80 participants is provided in Appendix E.

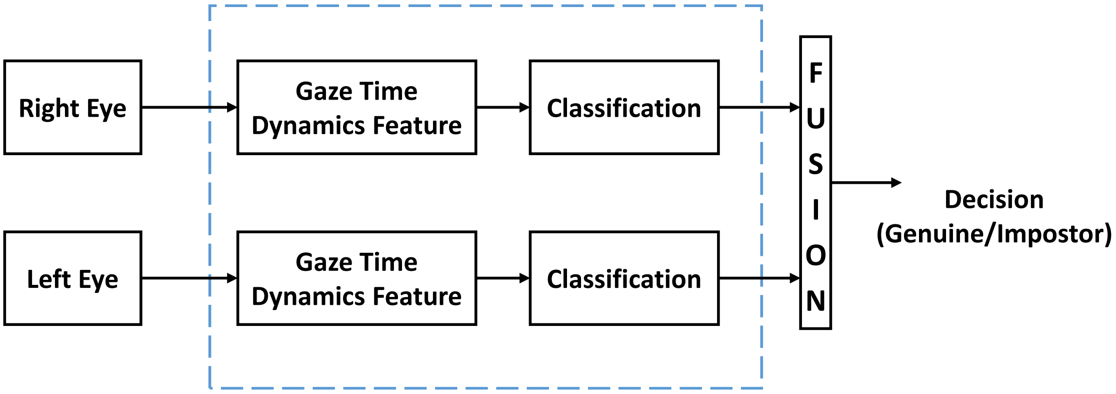

Right Eye
Gaze Time Dynamics Feature
Classification
Left Eye
Gaze Time Dynamics Feature
Classification
FUSION
Decision (Genuine/Impostor)

Figure 4.17: Score fusion using features extracted from left and right eyes

The results in this section are organized as follows.

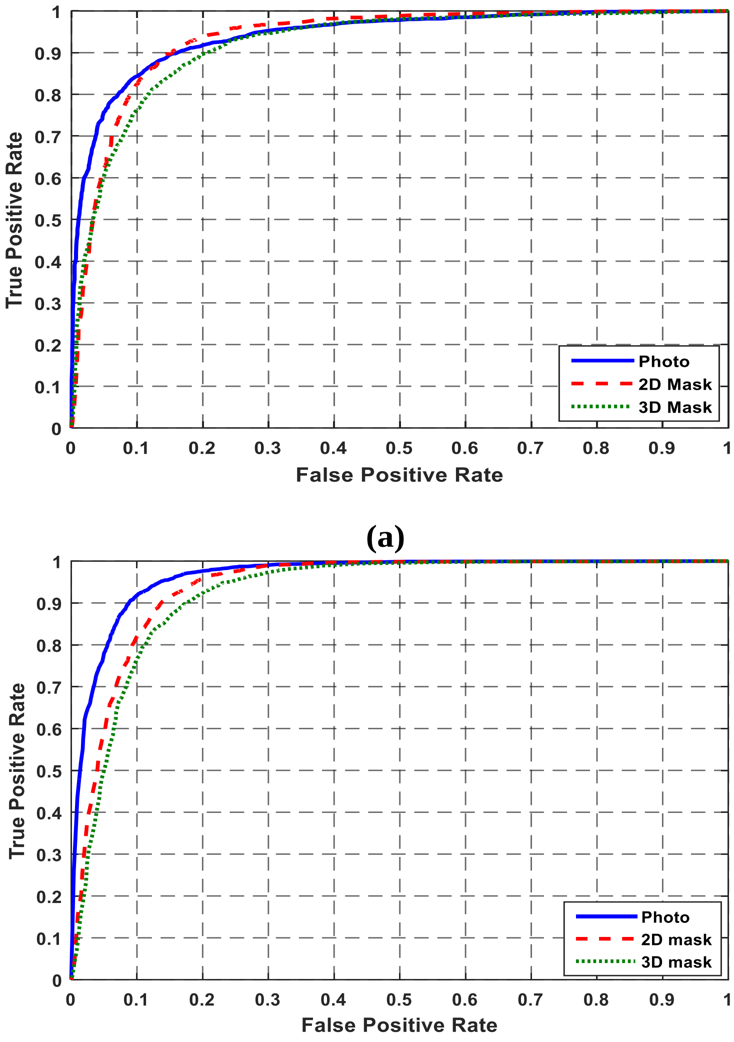
- Performance of the gaze time dynamics features using the KGDD database and **Tablet** format is analysed in subsection 4.4.2.1.
- Performance of the gaze time dynamics features using the KGDD database and **Phone** format is presented in subsection 4.4.2.2.

4.4.2.1 Tablet format

Figure 4.18 shows the performance of the system for three attack scenarios—displayed photos, 2D masks, and 3D masks—for Tablet format, using gaze time dynamics feature extracted from the subset of 70 participants, captured during **Lines** and **Curves** challenges of five-second duration. At FPR = 10%, TPRs are 85%, and 91% for displayed photo attack using Lines and Curves challenges, respectively. However, these fell to 82% and 81% when using 2D mask. For 3D mask, TPRs were 77% and 79% for Lines and Curves challenges, respectively. In general, performance was lower for mask attack detection than for displayed photo attack detection. An explanation of this observation is that the attacker can adjust the mask to fit their eye positions and so they may track the stimulus more accurately as mentioned in Section 4.2.2.

In a further set of similar tests, the dataset was instead captured from genuine users wearing **tinted glasses**. Figure 4.19 shows the ROC curves for displayed photo, 2D mask, and 3D mask attacks using the proposed feature extracted from the subset of 70 participants, captured during **Lines** and **Curves** challenges of five-second duration. At 10% FPR, TPRs were about 52%, 52%, and 40% for displayed photo, 2D mask, and 3D mask, respectively, using data captured during a Lines challenge. For Curves challenges and at 10% FPR, displayed photo attack detection TPR was 58%; 2D mask attack detection TPR was about 53%, and 3D mask attack detection TPR was about 39%. It is clear that the proposed features does not perform very well when users are wearing tinted glasses. This may be because the tinted glasses slightly obscure the eye centres.

Tables 4.10 and 4.11 summarise overall performance (using the newly standardized APCER and BPCER metrics, defined in the ISO/IEC CD 30107-3 standard) when the proposed features were tested on the KGDD database with 70 participants.



(a)

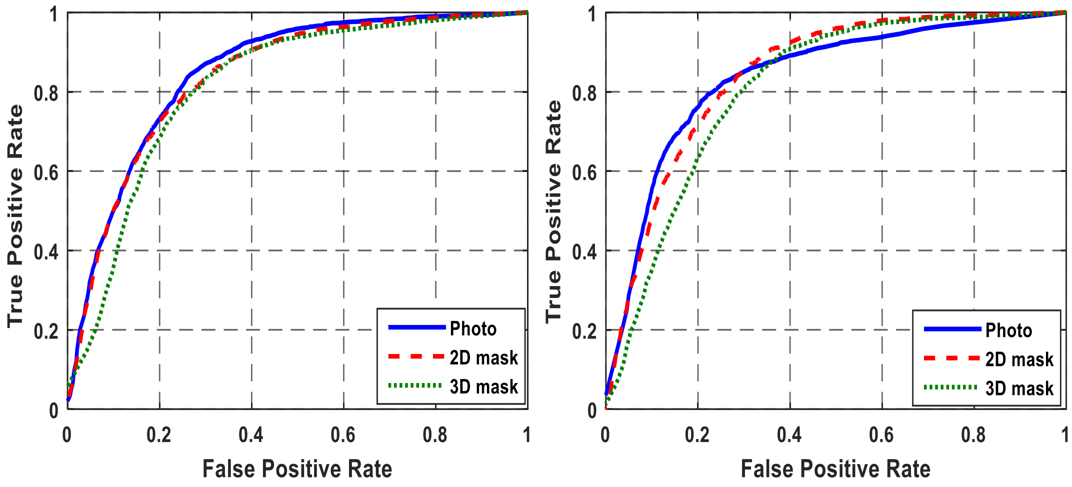
(b)

Figure 4.18: ROC curves for the three presentation attacks in **Tablet** format based on data captured for (a) Lines and (b) Curves challenges using gaze time dynamics features

Table 4.10: Performance of the system at APCER = 0.05 for Lines and Curves challenges using gaze time dynamics features

Attack type	**Lines challenge**		**Curves challenge**	
	APCER = 0.05**		**APCER = 0.05****	
	BPCER	**ACER**	**BPCER**	**ACER**
Photo	0.25	0.15	0.20	0.12
2D mask	0.37	0.21	0.40	0.22
3D mask	0.40	0.22	0.50	0.27
Overall	**0.40**	**0.22**	0.50	0.27

** Using the 70 publicly available participants in the KGDD (42 for training and 28 for testing)



(a) **(b)**

Figure 4.19: ROC curves for photos, 2D masks, and 3D masks in Tablet format based on data captured for (a) Lines and (b) Curves challenges using gaze time dynamics features and tinted glasses

Table 4.11: Performance of the system at APCER = 0.05 for Lines and Curves challenges using gaze time dynamics features and tinted glasses

Attack type	**Lines challenge**		**Curves challenge**	
	APCER = 0.05**		**APCER = 0.05****	
	BPCER	**ACER**	**BPCER**	**ACER**
Photo	0.62	0.33	0.61	0.32
2D mask	0.62	0.33	0.61	0.32
3D mask	0.80	0.42	0.81	0.43
Overall	**0.80**	**0.42**	0.81	0.43

** Using the 70 publicly available participants in the KGDD (42 for training and 28 for testing)

4.4.2.2 Phone Format

Experiments similar to those with Tablet format devices were performed to explore the same feature in the Phone platform. The performance of gaze time dynamics features for displayed photo, 2D mask, and 3D mask attacks as tested on the KGDD subset of 70 participants are presented in Figure 4.20.

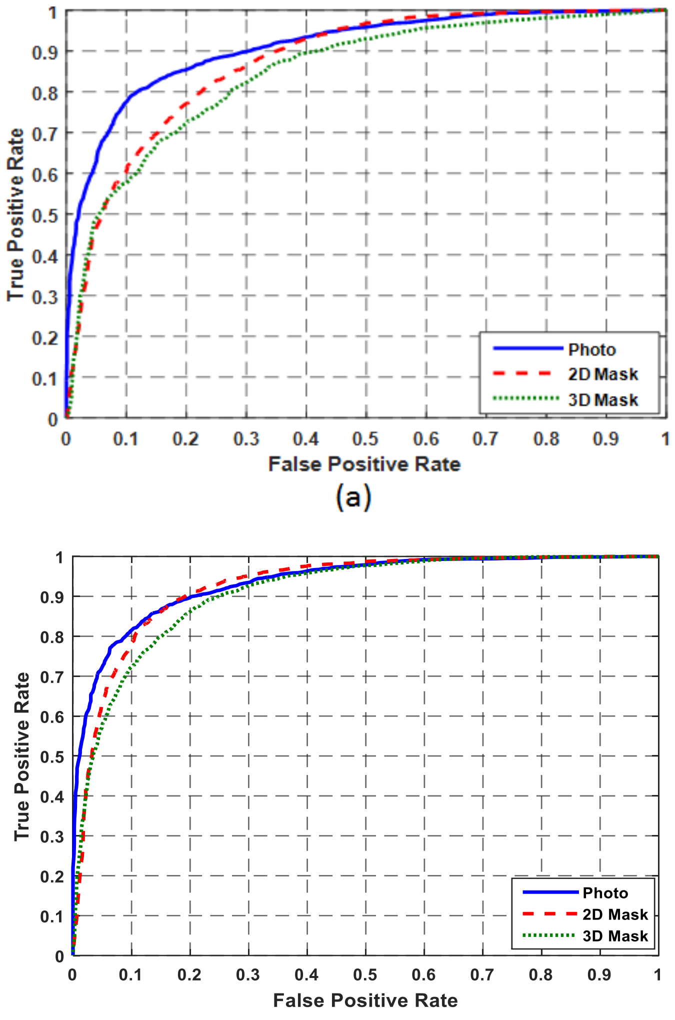
As can be seen from the figure, displayed photo attack detection returned the best performance while 2D mask attack detection ranked second, followed by 3D mask attack detection. At 10% FPR, about 79%, 60%, and 58% TPRs were achieved for displayed photo, 2D mask, and 3D mask detection, respectively, using data captured for Lines challenges.

For Curves challenges, at 10% FPR, TPRs of about 81%, 80%, and 71% were achieved for displayed photo, 2D mask, and 3D mask detection, respectively. The reason is similar to the one stated in the previous analysis in Section 4.2.2.1 and 4.4.2.2

The system was also tested in Phone format, based on data captured from users performing genuine presentations while wearing tinted glasses. In Figure 4.21, at 10% FPR, displayed photo detection TPR was about 71%; 2D mask attack detection TPR was 60%, and 3D mask attack detection was about 58% for Lines challenges.

For Curves challenges, TPRs were 61%, 57%, and 40% at 10% FPR for displayed photo attack, 2D mask attack, and 3D mask attack, respectively. System performance was poor, but some presentation attack attempts were still classified correctly.

Tables 4.12 and 4.13 summarise overall performance (using the newly standardized APCER and BPCER metrics, defined in the ISO/IEC CD 30107-3 standard) when the proposed features were tested on the KGDD database with 70 participants.



(a)

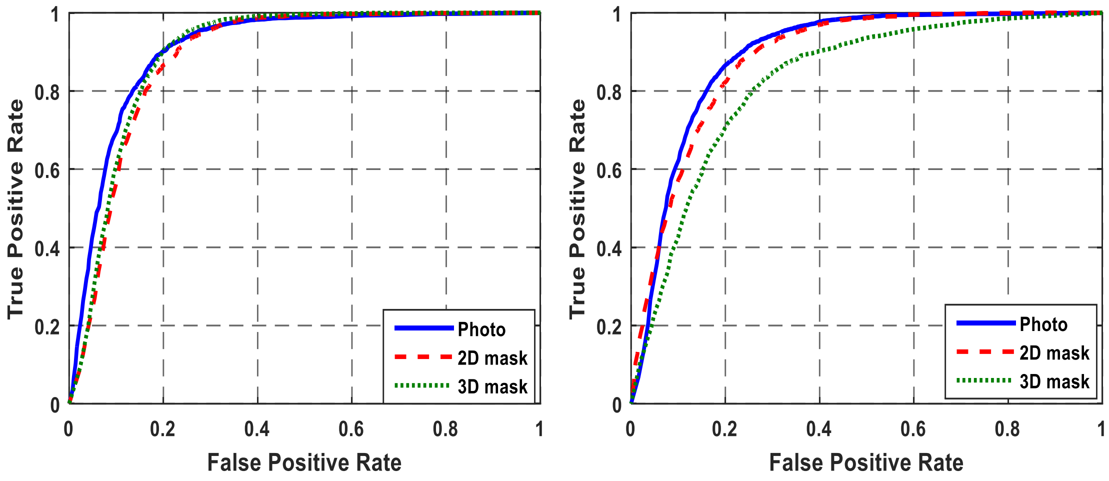
(b)

Figure 4.20: ROC curves for the three presentation attacks in Phone format based on data captured for (a) Lines and (b) Curves challenges using gaze time dynamics features

Table 4.12: Performance of the system at APCER = 0.05 for Lines and Curves challenges using gaze time dynamics features

Attack type	**Lines challenge**		**Curves challenge**	
	APCER = 0.05**		**APCER = 0.05****	
	BPCER	**ACER**	**BPCER**	**ACER**
Photo	0.35	0.20	0.29	0.17
2D mask	0.50	0.27	0.40	0.22
3D mask	0.50	0.27	0.43	0.24
Overall	0.50	0.27	**0.43**	**0.24**

** Using the 70 publicly available participants in the KGDD (42 for training and 28 for testing)



(a) **(b)**

Figure 4.21: ROC curves for photos, 2D masks, and 3D masks in Phone format based on data captured for (a) Lines and (b) Curves challenges using gaze time dynamics features and tinted glasses

Table 4.13: Performance of the system at APCER = 0.05 for Lines and Curves challenges using gaze time dynamics features and tinted glasses

Attack type	**Lines challenge**		**Curves challenge**	
	APCER = 0.05**		**APCER = 0.05****	
	BPCER	**ACER**	**BPCER**	**ACER**
Photo	0.50	0.27	0.67	0.36
2D mask	0.65	0.35	0.68	0.36
3D mask	0.65	0.35	0.80	0.42
Overall	**0.65**	**0.35**	0.80	0.42

** Using the 70 publicly available participants in the KGDD (42 for training and 28 for testing)

4.5 Further Analysis of Proposed PAD Scheme

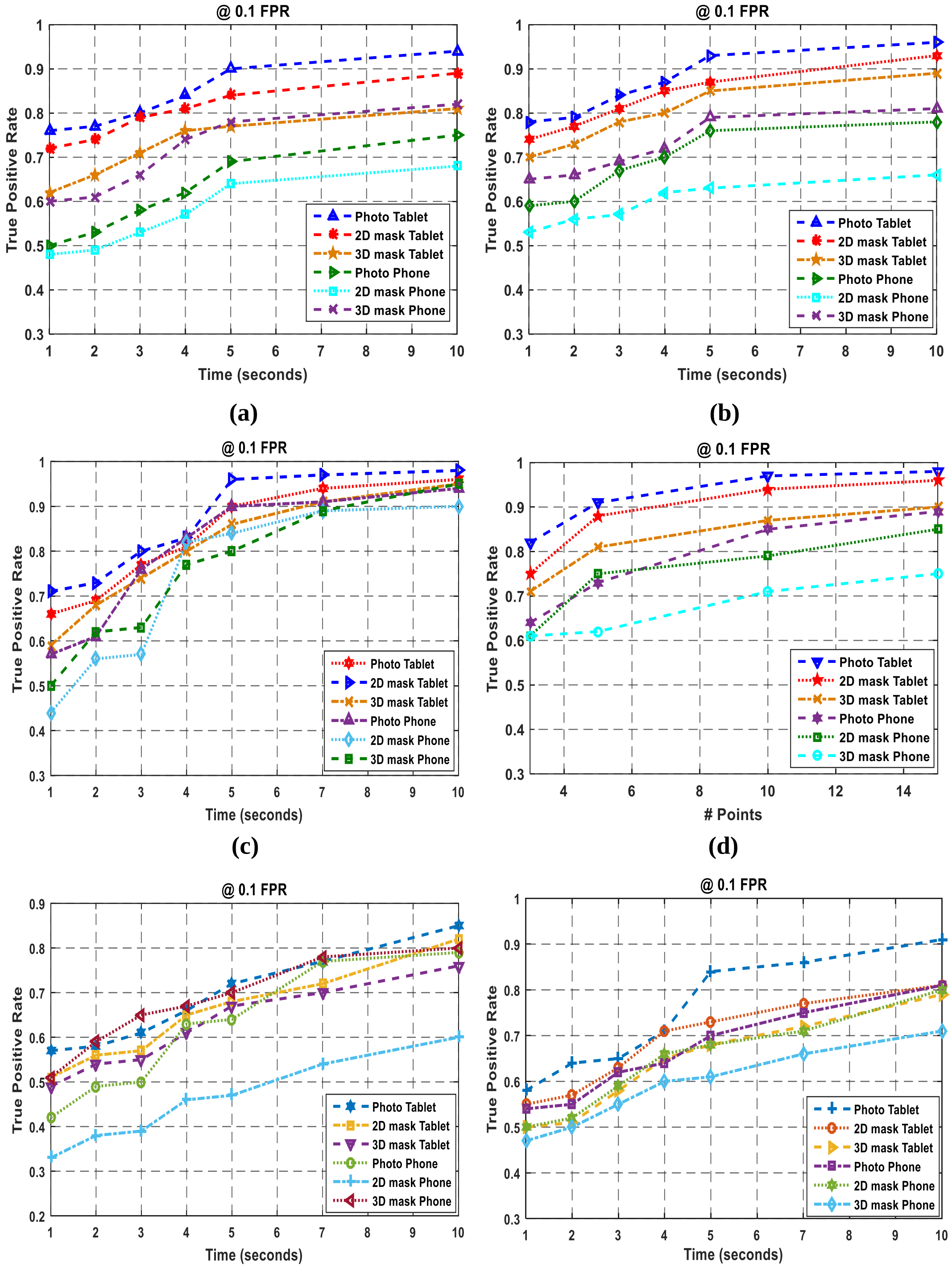
This section performs further analysis for the proposed gaze-based features. Firstly, the experimental setting is introduced, including dataset and then the impact of challenge duration on the performance of the proposed PAD is analysed (Section 4.5.1). Next, the computational load of the proposed method is studied (Section 4.5.2). Finally, the overall performances of proposed gaze-based features are compared with state-of-the-art algorithms (Section 4.5.3).

4.5.1 Effect of Challenge Duration

Database. To examine the effect of challenge duration, additional experiments were performed on the KGDD dataset with various challenge duration.

Figure 4.22 shows the effects of data capture duration at FPR = 10% for different challenge types using the three proposed gaze-based features (gaze correlation analysis, gaze alignment consistency and gaze time dynamics features) for Tablet and Phone formats. For the gaze alignment consistency, the stimulus moves along each line for an average of 1 s per line. Considering the result shown in Figures 4.22 (a, b, c, d, e, f), it is found that the challenge length has an impact on the overall performance of the system. When the challenge duration increases, the system performance generally improves. For example, in Figure 4.22 (a), the TPR of the proposed system using gaze correlation analysis for Tablet format decreases from 95% to 90% when the challenge duration drops from 10 to 5 seconds.

In all, the experimental results in this subsection show that the performance of the system using displayed photo, 2D mask, and 3D mask attack scenarios using Tablet and Phone formats generally improves with an increase in the challenge duration. The possible reason is that data based on a longer challenge duration contains more information for PAD.



(a) **(b)**

(c) **(d)**

(e) **(f)**

Figure 4.22: Effects of data capture duration at FPR = 0.1 for (a) Lines challenge using gaze correlation analysis; (b) Curves challenge using gaze correlation analysis; (c) Points challenge using gaze correlation analysis; (d) Lines challenge using gaze alignment consistence; (e) Lines challenge using gaze time dynamics; (f) Curves challenge using gaze time dynamics features.

4.5.2 Computational Time Requirement

As computational cost is known to be a critical factor in real applications, this section addresses the additional computational cost of the proposed methods. PAD systems based on the challenge-response method would be expected to incur higher computational costs than other methods because they need more time to run the challenge and record the response.

Computational cost is estimated by measuring the time required for feature extraction and classification process using the function 'tic' 'toc'. Measurement is based on a Matlab implementation of algorithms, running on a desktop Intel i5-3470 quad-core 3.20 GHz CPU with 16 GB RAM, using Windows 7 (64-bit) and Matlab 2015a (64-bit).

Table 4.14 lists the computational requirements of the three face-PAD approaches for 5-second challenge duration. For gaze alignment consistency, as an average of 1 s per line, the computational requirements in Table 4.14 is for 5 lines. The acquisition process and facial land mark detection perform in Off-line. When using gaze correlation analysis, the total time needed for feature extraction, and classification is about 0.02 seconds; additional time (1.16 seconds) is needed when the system is based on gaze alignment. For the gaze time dynamics features, total time from feature extraction to classification is approximately 0.04 seconds. As the system using the gaze alignment consistency features performs better, there is a trade-off between time consumption and performance.

Table 4.14: System computational time requirement for 5-second challenge duration (seconds)

Operation	Feature Extraction	Classification	Total
Face-PAD-based gaze correlation analysis	**0.004**	**0.02**	**0.02**
Face-PAD-based gaze alignment consistency	1.123	0.04	~1.16
Face-PAD-based gaze time dynamics feature	0.022	**0.02**	0.04

4.5.3 Comparison with Other State-of-the-art Methods

The published face PAD methods are commonly evaluated using public domain face spoofing databases (e.g. IDIAP and CASIA). For comparison with previously published work, we reproduced the results of two widely used methods: LBP [32] and LBP-TOP [160]. These experiments were performed using Bob [161], a freely downloaded signal and image processing toolbox, which includes a library with implementations of several presentation attack detection algorithms. These methods were tested on the KGDD database captured when displaying the challenge on a Tablet format for 5-second challenge duration.

Table 4.15 reports overall performance data for the different methods. For the system using gaze alignment consistency, ACER value decreases from 0.23 to 0.14 when compared to the LBP+SVM scheme; ACER is lower by about 4% for the same BPCER performance when compared to the LBP-TOP+SVM scheme. When using gaze correlation analysis, the system achieves better performance when compared to the LBP+SVM and LBP-TOP+SVM schemes. In all cases, the proposed gaze-based methods returned the lowest ACER results apart from the gaze time dynamics feature.

Table 4.15: Comparison of performance reports for the same collected database

Method	APCER	BPCER	ACER
$\mathrm{LBP}\,{}^{\mathrm{u2}}_{\mathrm{3x3}} + \mathrm{SVM}$ [30]	0.20	0.26	0.23
$\mathrm{LBP} - \mathrm{TOP}_{8,8,8,1,1,[1-2]} + \mathrm{SVM}$ [157]	0.13	0.23	0.18
Gaze correlation analysis	0.05	0.29	0.17
Gaze alignment consistency	**0.05**	**0.23**	**0.14**
Gaze time dynamics feature	0.05	0.42	0.23

Table 4.16 compares our experimental results to recently published performances based on different databases. As the work presented here is based on the challenge-response system, it was difficult to draw a fair and direct comparison with approaches that use different methods for liveness detection. Nevertheless, the results indicate the potentially superior performance of the proposed approach.

Table 4.16: Comparison of performance reports for different databases

Method	Database used	# of users	Type of attack	HTER
IQA [162]	Replay-Attack	50	Printed photo Display photo Video replay	0.15
IDA [57]	Replay-Attack	50	Printed photo Display photo Video replay	**0.07**
Gaze stability [128]	Self-collected	30	Photo 2D Mask Video replay	0.10
Gaze correlation analysis	KGDD	70	Photo 2D Mask 3D Mask	0.11
Gaze alignment consistency	KGDD	70	Photo 2D Mask 3D Mask	**0.07**
Gaze time dynamic feature	KGDD	70	Photo 2D Mask 3D Mask	0.16

4.6 Observations and Discussion

Several conclusions can be drawn from the results reported in this chapter which are summarised as follows.

First, it is clear that the proposed gaze-based features for detecting displayed photo attack generally perform better compared to the other PAIs (2D mask and 3D mask attacks) for all three challenge types and for both screen sizes (Tablet and Phone formats). There are two possible reasons for this; (1) it is easier for the attacker to adjust the mask to fit their eye positions and so they can track the stimulus more accurately and (2) in the display photo attack, visually guided hand movements are required to orient the artefact to the challenge direction, reducing the similarity of challenge and response trajectories as hand-eye coordinate is difficult.

Second, system performance was clearly better when challenge duration increased (see Figure 4.22). This suggests that data based on a longer challenge duration contains more information to discriminate between genuine and attack attempts as mentioned in Section 4.5.1.

Third, based on the results for the proposed gaze based features when using different screen sizes, system performance was generally better when the challenge was displayed in the Tablet format as compared to the Phone format. This suggests that the larger screen (Tablet format) may prompt large eye movements, so providing more information for attack detection for most of the assessed features and challenge types.

Fourth, the effects of different challenge trajectories were also evaluated. In general, the proposed features performed similarly on Lines and Curves challenges, suggesting that challenge trajectory has little influence on the performance of the proposed PAD algorithm. The only exception is Points challenges, where the proposed gaze correlation analysis performed better for Points challenge compared to other challenges types. As the stimulus jumps to random locations, this may indicate that saccadic movements are preferable for gaze data of this type.

Finally, the results suggest that performance is poor when genuine users attempt to access the system while wearing tinted glasses. As mentioned in Section 4.3.2.1, this may be because (1) the tinted glasses slightly obscure the eye centres and (2) the reflection of the white screen on the glasses affects the facial landmark detection software.

4.7 Summary

This chapter describes novel features for face PAD applications. The proposed system employs a challenge-response approach, using a visual stimulus to measure the user's gaze in order to establish the presence of displayed photo, 2D mask, and 3D mask attacks.

Three gaze-based features extracted from the pupil centres have been used to distinguish between genuine and presentation attack attempts: gaze correlation analysis, gaze alignment consistency, and gaze time dynamics features. Several sets of experiments were performed to explore different features, PAIs, and screen sizes for challenge display. The experiments reported here also investigated the effect of challenge duration on the PAD accuracy, as well as the effect of genuine users wearing tinted glasses.

The experiments used the KGDD database under different test protocols. Based on the experimental observations in section 4.6, the main conclusions can be summarised as follows.

- The gaze-based features described in this chapter generally perform better in detecting displayed photo attacks than for other PAIs (2D mask and 3D mask attacks).
- These gaze-based features generally perform better when challenges are displayed on a larger screen (Tablet format).
- As the system performs better for a long challenge duration, there is a trade-off between challenge duration and performance.
- Performance is similar for Lines and Curves challenges. For Points challenges, the system performs only slightly better in most cases.
- The gaze alignment consistency delivers significant advantage when compared to the other two gaze-based features.
- Performance is worse when genuine users attempt to access the system while wearing tinted glasses, indicating that tinted glasses impact significantly on outcomes when using gaze-based features.

- In comparison to state-of-the-art methods, the gaze-based method achieves comparable performance on the KGDD.

The chapter's main contributions are as follows.

- Testing of three novel gaze-based features
- Exploring three types of PAI (displayed photo, 2D mask, and 3D mask attacks)
- Investigating the effect of tinted glasses on PAD accuracy
- Assessing the impact of challenge duration on PAD performance
- Investigating different challenge trajectories (Lines, Curves, and Points)
- Exploring the effect of different screen sizes

The next chapter explores the effectiveness of the gaze-based features described in this chapter when using a real handheld device.

Chapter 5

Presentation Attack Detection using a Mobile Handheld Device

5.1 Introduction

This chapter investigates the feasibility of the proposed scheme on a true handheld device. To test the features described in Chapter 4, another database was collected using a handheld mobile device. The new database was collected to explore whether the proposed gaze-based features are affected by the hand movements or not. The same kind of visual stimuli presented in Section 3.3.1 are used here to direct the gaze of the user to random trajectories on the screen. Two challenge trajectory types were explored: Lines, and Points. The Curves challenge was excluded in this collected data, as not all gaze features were extracted using this challenge and also the performance of the system using this particular challenge and testing on the KGDD database (Chapter 4) was similar to the performance of Lines challenge. The system records the user’s gaze via the frontal camera of the mobile device. Features based on the measured similarity between the observed gaze and the presented challenge are then used to discriminate between genuine attempts and those by impostors. Experiments reported later in this chapter analyse the effectiveness of the proposed method in detecting presentation attacks. The scenario in the experiments reported in this chapter involves a biometric authentication system for mobile devices; the presentation attack is assumed to be an impostor using a displayed photo or a 2D or 3D mask.

The chapter is organized as follows. The new Mobile Kent Gaze Dynamic Database (MKGDD) is described in section 5.2, and section 5.3 explores the effectiveness of gaze-based features using MKGDD. Section 5.4 summarises the experimental results for gaze-based features. The influence of mobile device on system performance is given in Section 5.5, followed by some concluding remarks in Section 5.6.

5.2 The Mobile Kent Gaze Dynamic Database

The MKGDD comprises short video recordings of 30 participants, collected from users' genuine attempts to track a moving visual target (the 'stimulus' or 'challenge'), with natural head/eye movements, and impostor attacks in which users were holding a displayed photo, looking through a 2D mask or holding a 3D mask of a genuine user while holding the mobile device and attempting to follow the stimulus. Figure 5.1 shows examples of genuine and presentation attack attempts from the MKGDD database. For each attempt, we also collected additional sensor data (e.g. accelerometer and gyroscope

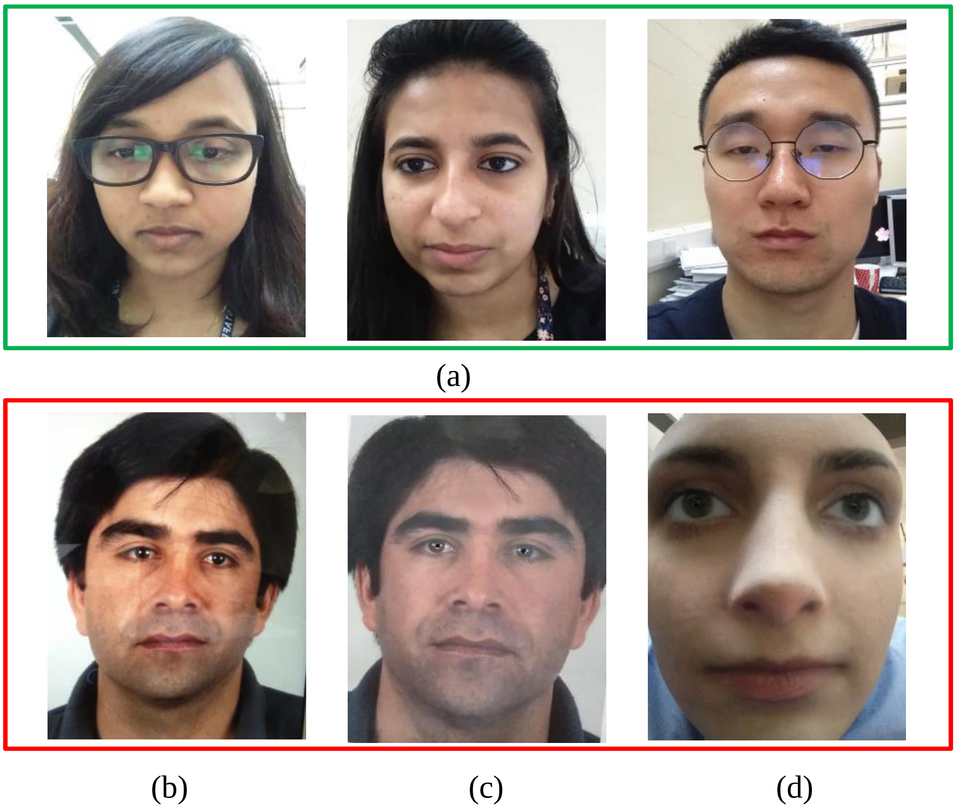
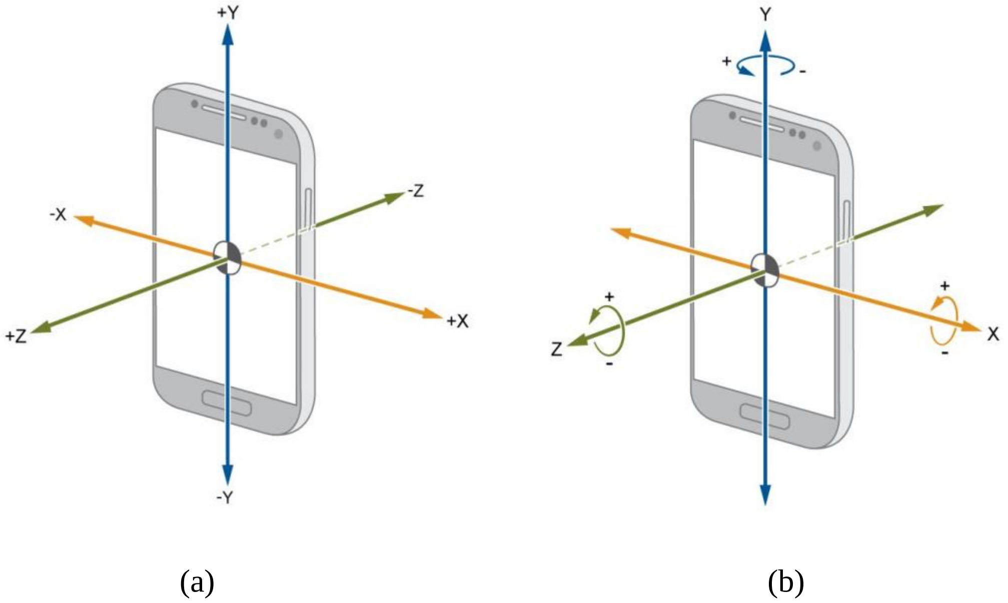


Figure 5.1: Examples images of genuine and attack attempts in the collected MKGDD database: (a) genuine users; (b) displayed photo attack using an iPad; (c) 2D mask attack; (d) 3D mask attack

parameters) from the mobile device. The accelerometer measures the acceleration applied to the mobile device on all three axes and has three corresponding outputs (x, y and z) (see Figure 5.2(a)). The gyroscope is used to measure rotation around the x, y and z axes [163] (see Figure 5.2(b)). An accelerometer measures tilt motion and orientation in a mobile phone. Metadata associated with each participant were also collected to facilitate experiments using specific subgroups. The available information includes gender, handedness and visual aids (glasses or contact lenses as opposed to none). The demographic distributions of the participants are presented in Table 5.1. Two different types of challenge trajectory (Lines and Points) were presented to each participant. This section details the data collection process and explains the evaluation protocols.



(a) (b)

Figure 5.2: a) the accelerometer and b) the gyroscope parameters on the Android device [160]

Table 5.1: Demographic distribution of the users

Demographic types		Number of participants
Gender	Male	13
	Female	17
Eyewear	Glasses	6
	Contact lenses	2
Handedness	Right-handed	28
	Left-handed	2

5.2.1 Mobile Application Design

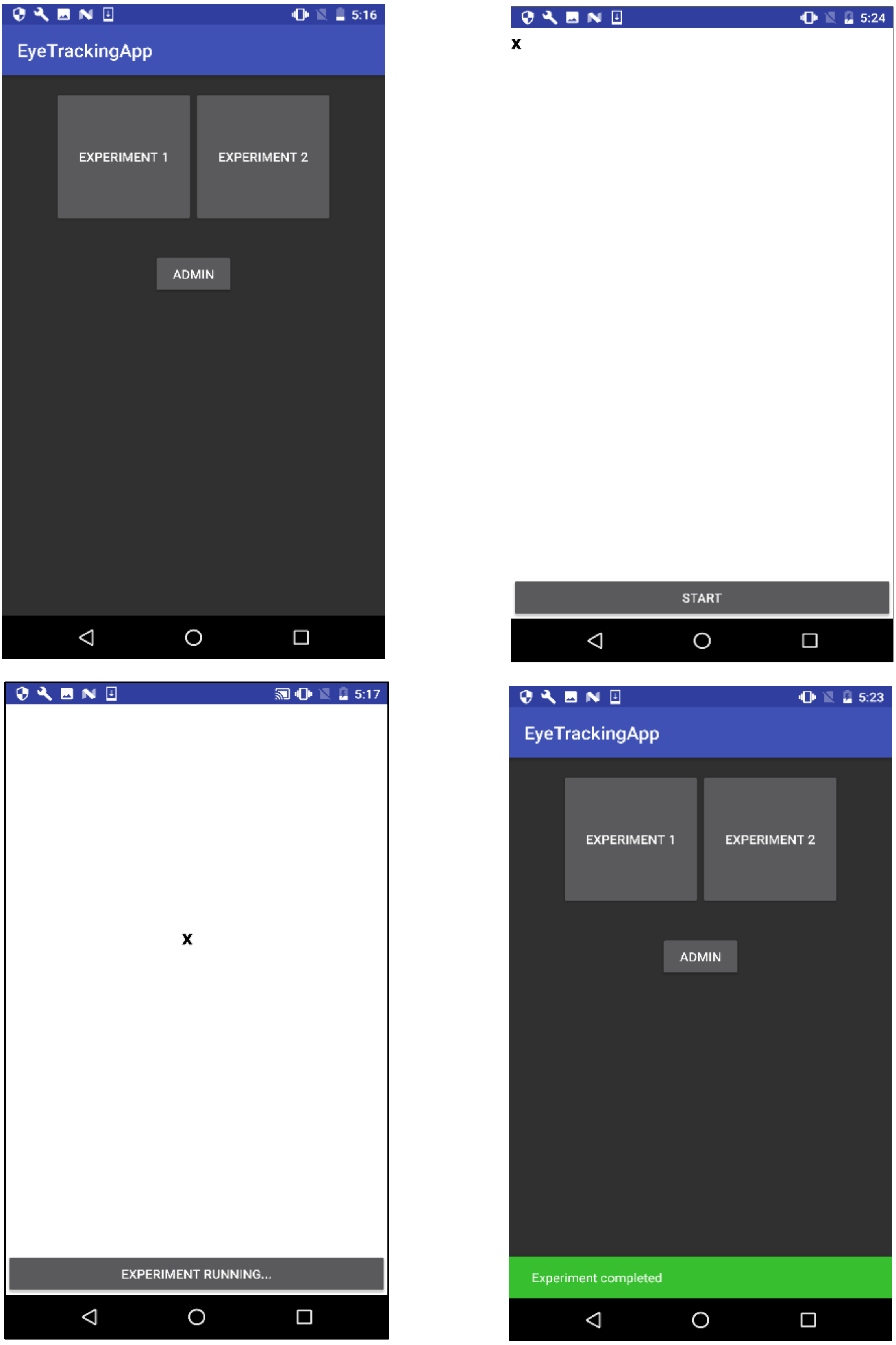
An Android application was developed to collect data from individual users; this was loaded onto a Google Nexus 6 phone. A graphical user interface (GUI) was designed to display a moving challenge and to video record the user's attempts to follow the challenge during simulated spoofing attacks.

5.2.2 Mobile Platform Selection

The proposed PAD system was implemented for true mobile handheld devices. Of the three main operating systems for mobile devices (IOS, Android OS, and Windows Phone OS) [164], the Android OS was chosen for several reasons. First, Android is an open source OS and has free tools for developing applications. Second, Android applications can be developed for any operating system (e.g. Mac OS, Linux, Windows) while iOS and Windows Phone applications are developed in Mac OS and Windows OS, respectively. Finally, because of the companies' policies, it is easier to publish Android applications to the market as compared to other operating systems [164]. For these reasons, Android OS seemed to be the optimal choice for this project.

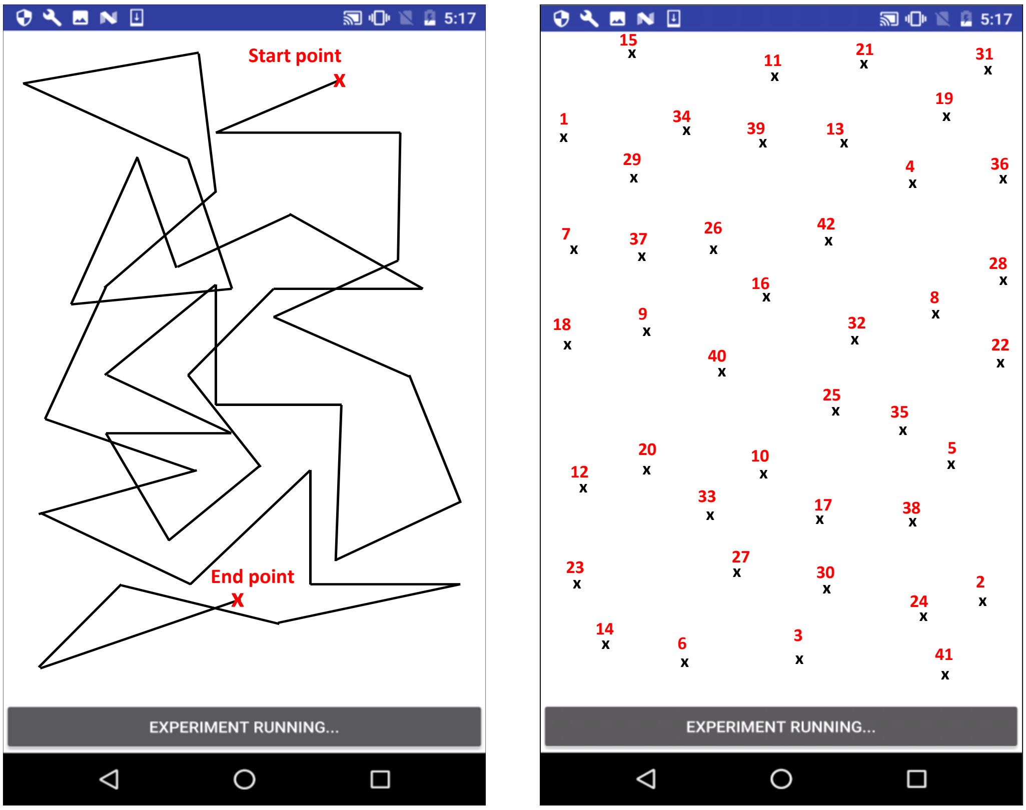
5.2.3 Challenge Design and Response Acquisition

When participants open the application, the GUI appears, displaying two buttons: EXPERIMENT 1 and EXPERIMENT 2 (as shown in Figure 5.3). Clicking either of these buttons displays the challenge interface, which contains one button (START). On pressing START, the challenge remains static for one second to allow the user's eyes to fixate on it before it begins to move and the camera begins to record a video of the attempt. Two challenge trajectories (Lines and Points) were shown to the user for each attempt. EXPERIMENT 1 is a Lines challenge as can be seen in Figure 5.4(a), in which the stimulus moves along a set of connected straight lines in various random directions. Each session includes several lines of differing lengths. EXPERIMENT 2 includes a Points challenge (see Figure 5.4(b)), in which the stimulus appears at random locations on the screen.



Experiment completed

Figure 5.3: The application interface



(a) (b)

Figure 5.4: Challenge trajectories. a) Lines challenge; b) Points challenge

The participant is holding the mobile device in front of his/her face and instructed to follow the shape (stimulus) that appears on the mobile screen using natural head/eye movements, and the front camera (sensor) records a video of the attempt as the stimulus moves on the screen (see Figure 5.5). The challenge is designed so that the moving stimulus is synchronized with the video recording. In each case, data were recorded for the entire duration of the attempt (approximately 30 seconds). It is anticipated that in a practical application, only a short presentation of the stimulus of the order of a few seconds may be sufficient to detect presentation attempts. However, the data collected were of much longer duration, which were then suitably partitioned to establish the trade-off between the duration and accuracy of performance of the system.

5.2.4 Generation of Attacks

The same types of PAI used in KGDD (cf. Section 3.3.2) were used in these experiments. The PAIs included an impostor attempting authentication by displaying an image on an iPad screen as can be seen in Figure 5.6, a 2D photo mask (Figure 5.7) or a 3D mask (Figure 5.8) to the face recognition system's camera.

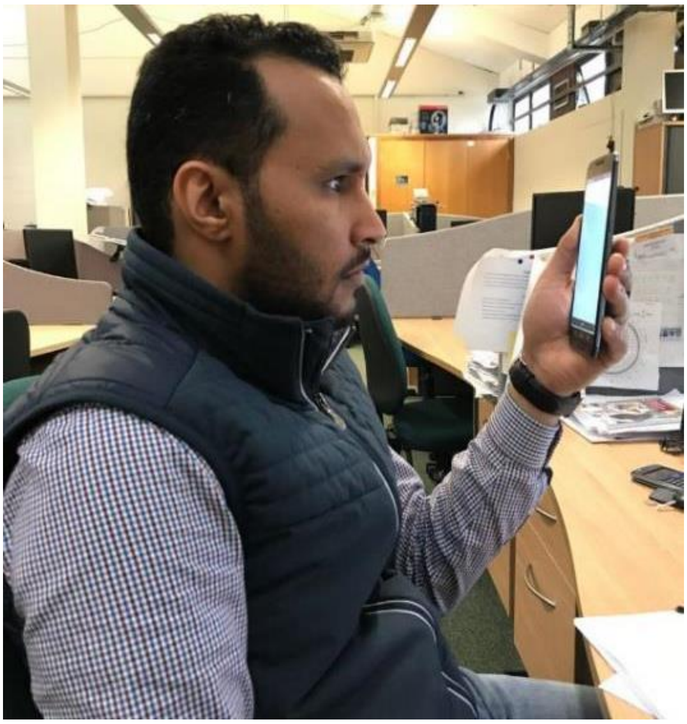
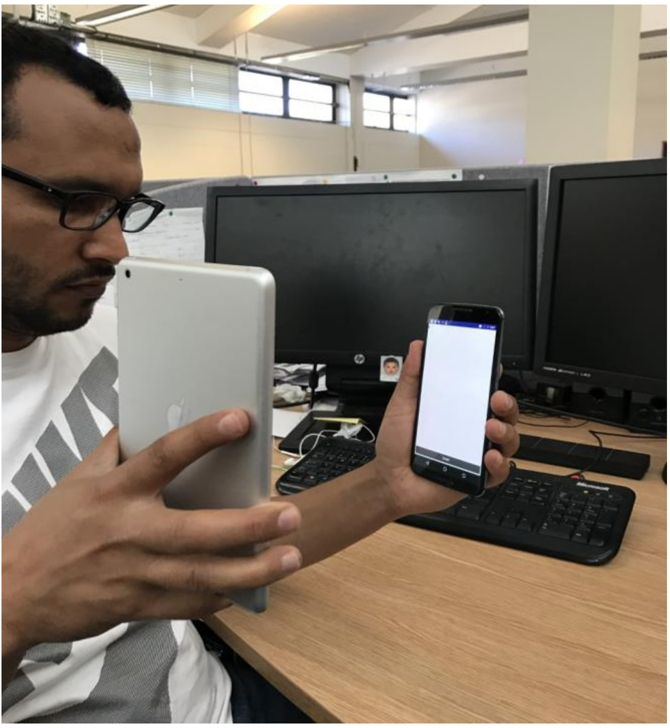


Figure 5.5: A genuine user using the mobile device to track the moving challenge



:

Figure 5.6: An impostor holding a displayed photo and attempting to follow the challenge during the simulated spoofing attack

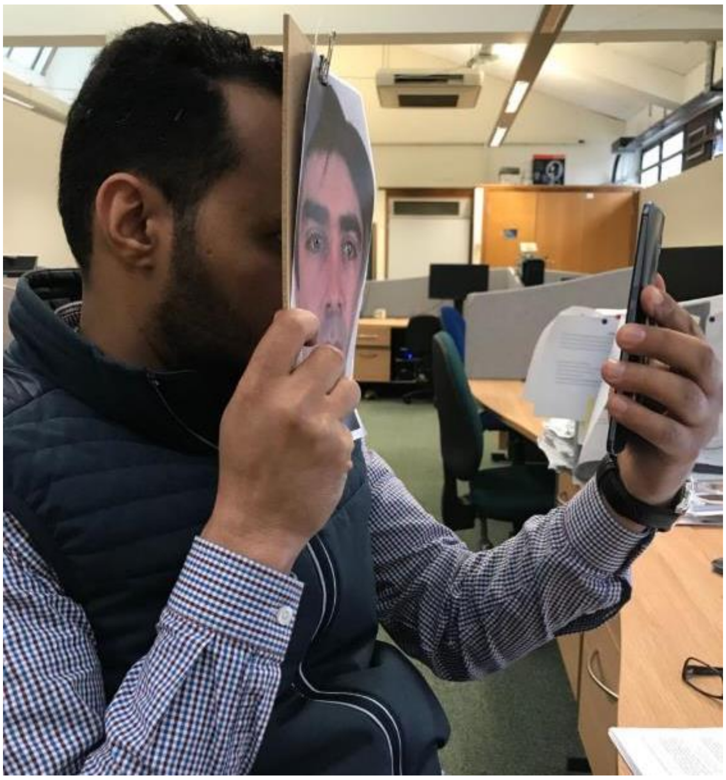


Figure 5.7: An impostor holding a 2D mask and attempting to follow the challenge during a simulated spoofing attack

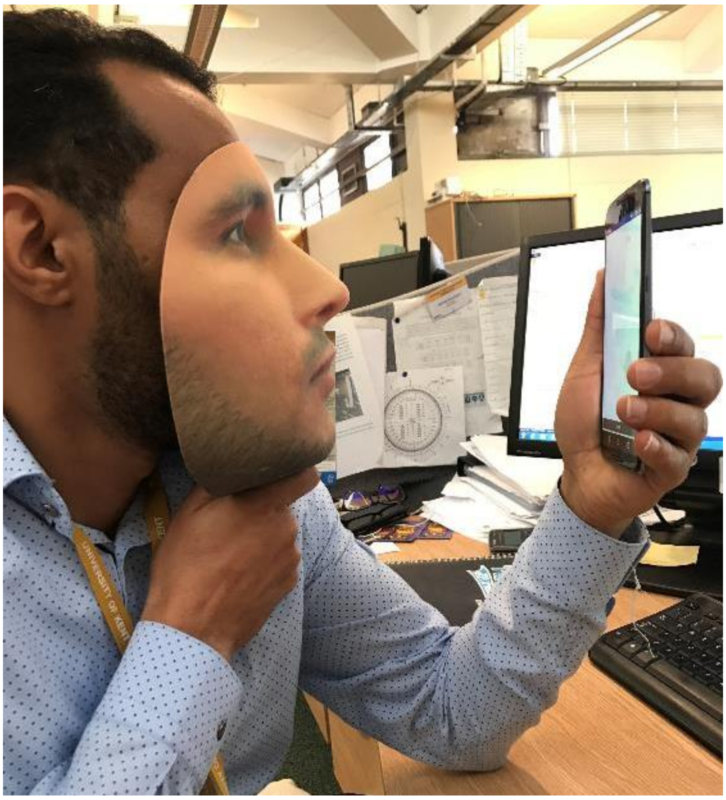


Figure 5.8: An impostor wearing a 3D mask and attempting to follow the challenge during a simulated spoofing attack

5.2.5 Hardware Selection

The Android application was installed on a Google Nexus mobile, which was used to collect genuine and impostor attempts. This phone was chosen for data collection as it has a most common screen size. Figure 5.9 shows the Google Nexus mobile phone used for data collection; the hardware specifications are listed in Table 5.2.

Figure 5.9: The Google Nexus 6 phone used in data collection [165]

Table 5.2: Google Nexus 6 specifications

Operating System	Android 5.0 (Lollipop), upgradeable to 7.1.1 (Nougat)
CPU	Quad-core 2.7 GHz Krait 450
Memory	32/64 GB
Display size	5.96 inches (~74.1% screen-to-body ratio)
Display Resolution	1440 x 2560 pixels, 16:9 ratio (~493 ppi)
Primary camera	13 MP
Secondary camera	2 MP
Video	2160pixel@30fps

5.3 The Proposed System

To explore the gaze-based features introduced in Chapter 4, the proposed system was tested on the MKGDD database. As shown in Figure 5.10, the challenge was presented to the user as a visual stimulus on the device's screen, and the participant was asked to hold the device in front of his face and follow the challenge using natural head/eye movements while the front camera (sensor) recorded a video of their attempts. As mentioned in Section 3.2.1, the system used Chehra software to extract facial landmarks (eye centres) from the video frames, computing a number of features from these landmarks. As described in detail in Chapter 4, these include gaze correlation analysis, gaze alignment consistency and gaze time dynamics. Following feature extraction, a k-nearest neighbour (*k*-NN) classifier was applied with *k* = 1 to discriminate attack attempts from genuine attempts. The classification protocol was as follows: features are extracted from each eye and then passed to separate classifiers (1-NN) to obtain individual classification scores for each eye. These scores are then fused, using score fusion to obtain the final PAD result based on the product. The MKGDD database can be split randomly into the training set (containing 18 subjects) and the testing set (containing 12 subjects). Each experiment was run 100 times with random data sets for training and testing.

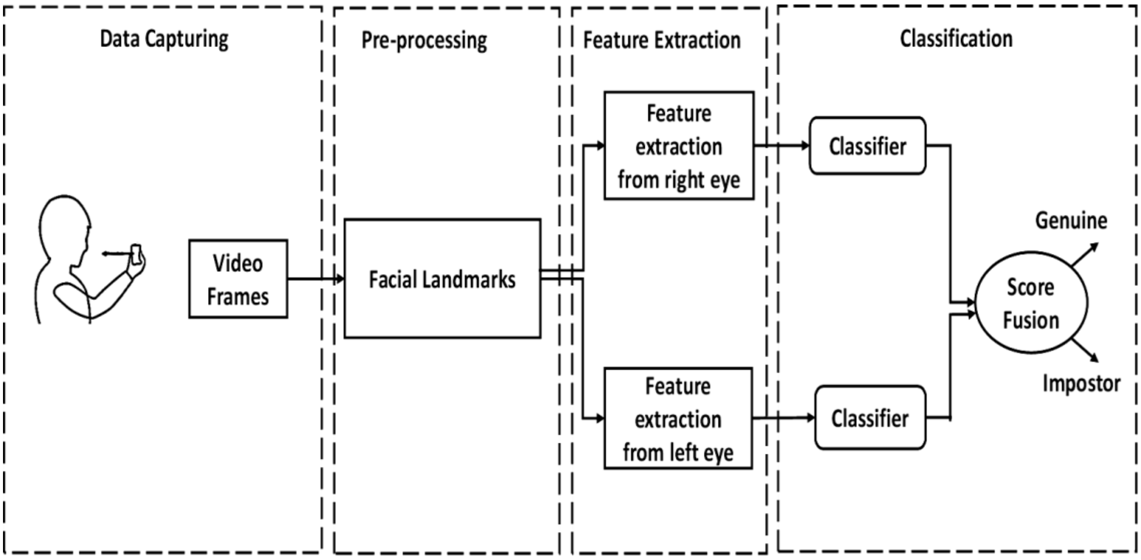


Figure 5.10: Diagram of the proposed system

5.4 Experimental Results for Gaze-based Features

In this part, the impact of using a mobile device on overall system performance was investigated. This section presents the results obtained after extensive testing of the different gaze-based features (gaze correlation analysis, gaze alignment consistency and gaze time dynamics). Details of these features and their calculation can be found in Chapter 4. The MKGDD database used for these experiments includes different challenges (Lines and Points) and PAIs (displayed photo, 2D and 3D masks).

5.4.1 Gaze Correlation Analysis

Correlation analysis was used to measure the similarity between gaze and challenge trajectories. Further detail on feature calculation can be found in section 4.2.1.

Figures 5.11 and 5.12 show ROC curves for the displayed photo, 2D mask and 3D mask attacks, using gaze correlations extracted from the dataset captured during **Lines** and **Points** challenges of five seconds duration. At FPR = 10%, the TPRs are 81%, and 86% for displayed photo attack using Lines and Points challenges, respectively. However, these fall to 75%, and 80% when using the 2D mask. At 10% FPR, 3D mask attack detection TPRs are about 68% for the Lines challenge and 78% for the Points challenge. As mentioned in section 4.6, the performance of gaze-based features in detecting displayed photo attacks generally achieves the best performance; one possible explanation is that it is easier for the attacker to look through and accurately track the stimulus because the masks have holes in the pupil centres.

Table 5.3 summarises overall performance for **Lines** challenges of different durations when tested on the MKGDD database, using the standardized APCER and BPCER metrics as defined in the ISO/IEC CD 30107-3 standard. The system ACERs were 14%, 17% and 21% at APCER = 10% for displayed photo, 2D mask and 3D mask, respectively, for a challenge duration of 5 seconds. These error rates fell to 10%, 14% and 15% for a 10-second challenge. Table 5.4 summarises overall performance for gaze correlation feature on the MKGDD database using **Points** challenges of differing durations. At APCER = 10%, ACERs were 12%, 11% and 9% for displayed photo challenge durations of 5, 7 and 10 seconds, respectively. For 2D mask attack detection, overall performance for a 10-second challenge duration was 10% ACER. However, this increased to 0.15 for a challenge duration of 5 seconds. ACERs for 3D mask attack detection based on 5-, 7- and 10-second challenges were 16%, 15% and 14% respectively. Tables 5.3 and Table 5.4 show that system performance was clearly better when challenge duration increased. This suggests that data based on a longer challenge duration are more useful in discriminating between genuine and presentation attack attempts (as noted in section 4.5.1).

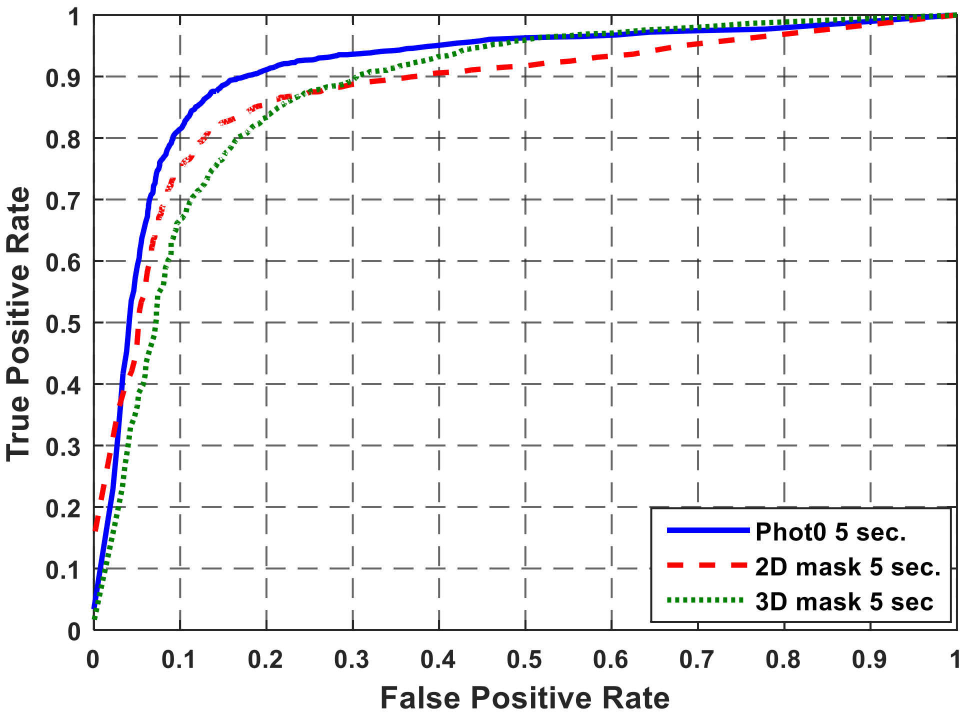


Figure 5.11: ROC curves for photo, 2D mask and 3D mask, using data captured during Lines challenge for correlation analysis

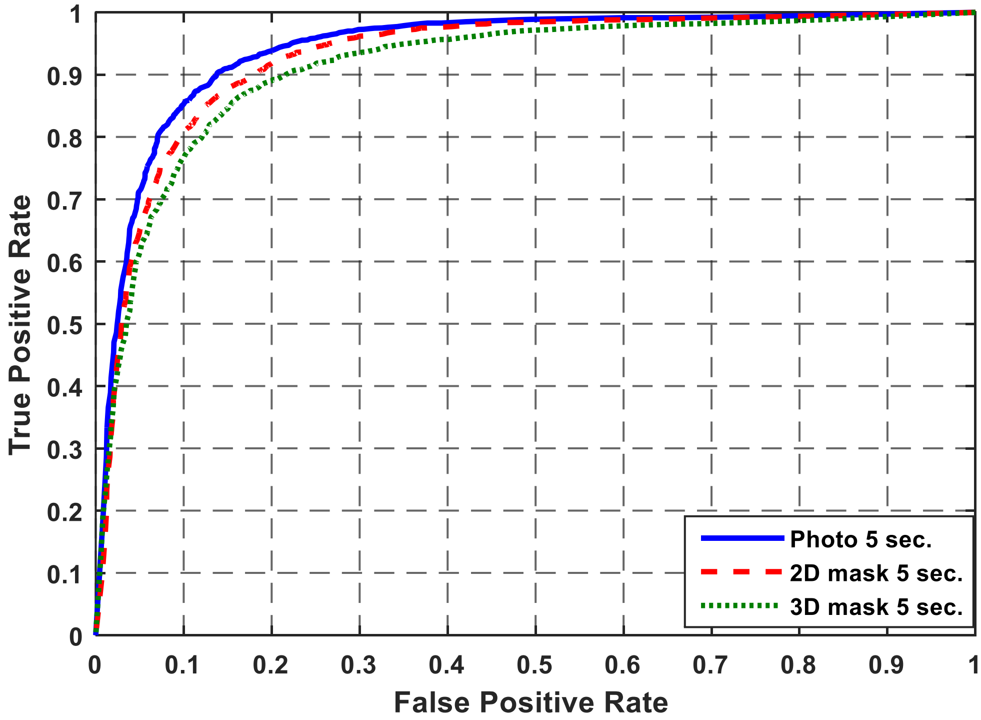


Figure 5.12: ROC curves for photo, 2D mask and 3D mask, using data captured during Points challenge for gaze correlation analysis.

Table 5.3: System performance at APCER = 0.10, based on data captured during Lines challenge for correlation analysis at various challenge durations.

Attack type	5 seconds		7 seconds		10 seconds	
	BPCER	ACER	BPCER	ACER	BPCER	ACER
Photo	0.19	0.14	0.17	0.13	0.11	0.10
2D mask	0.24	0.17	0.22	0.16	0.18	0.14
3D mask	0.32	0.21	0.29	0.19	0.21	0.15
Overall	0.32	0.21	0.29	0.19	**0.21**	**0.15**

Table 5.4: System performance at APCER = 0.10, based on data captured during Points challenge for correlation analysis at various challenge durations.

Attack type	5 seconds		7 seconds		10 seconds	
	BPCER	ACER	BPCER	ACER	BPCER	ACER
Photo	0.15	0.12	0.12	0.11	0.08	0.09
2D mask	0.20	0.15	0.19	0.14	0.10	0.10
3D mask	0.22	0.16	0.21	0.15	0.18	0.14
Overall	0.22	0.16	0.21	0.15	**0.18**	**0.14**

5.4.2 Gaze Alignment Consistency

To test gaze alignment consistency on the MKGDD, the system recorded the user's gaze using the front camera of the mobile device. Features based on the measured alignment of the observed gaze were then used to discriminate between genuine attempts and impostors. Gaze alignment consistency was extracted from sets of images captured when the stimulus was on a given line. These features were designed to capture angular similarities between the stimulus trajectory and the trajectory of the participant's pupil movement for each line segment (see equation (4.12)). The feature vector was the difference between the two angles for each of the line segments included in the challenge (see equation (4.13)).

Further details of feature calculation are provided in Section 4.3.1.

Gaze alignment consistency performance for displayed photo, 2D mask and 3D mask attacks using this feature set and tested on the MKGDD is shown in Figure 5.13. Displayed photo attack detection returned the best performance, followed by the 2D mask attack detection and the 3D mask attack detection. At FPR = 10%, displayed photo, 2D mask and 3D mask detection achieved approximately 91%, 81% and 76% TPR, respectively, for a challenge duration of 5 seconds.

Table 5.5 summarises overall performance gaze alignment consistency tested on the MKGDD database using the **Lines** challenge at durations of 5, 7 and 10 seconds, respectively, using the newly standardized APCER and BPCER metrics as defined in the ISO/IEC CD 30107-3 standard. The results demonstrate that overall system performance generally improves with increased challenge duration, for reasons similar to Section 5.4.1.

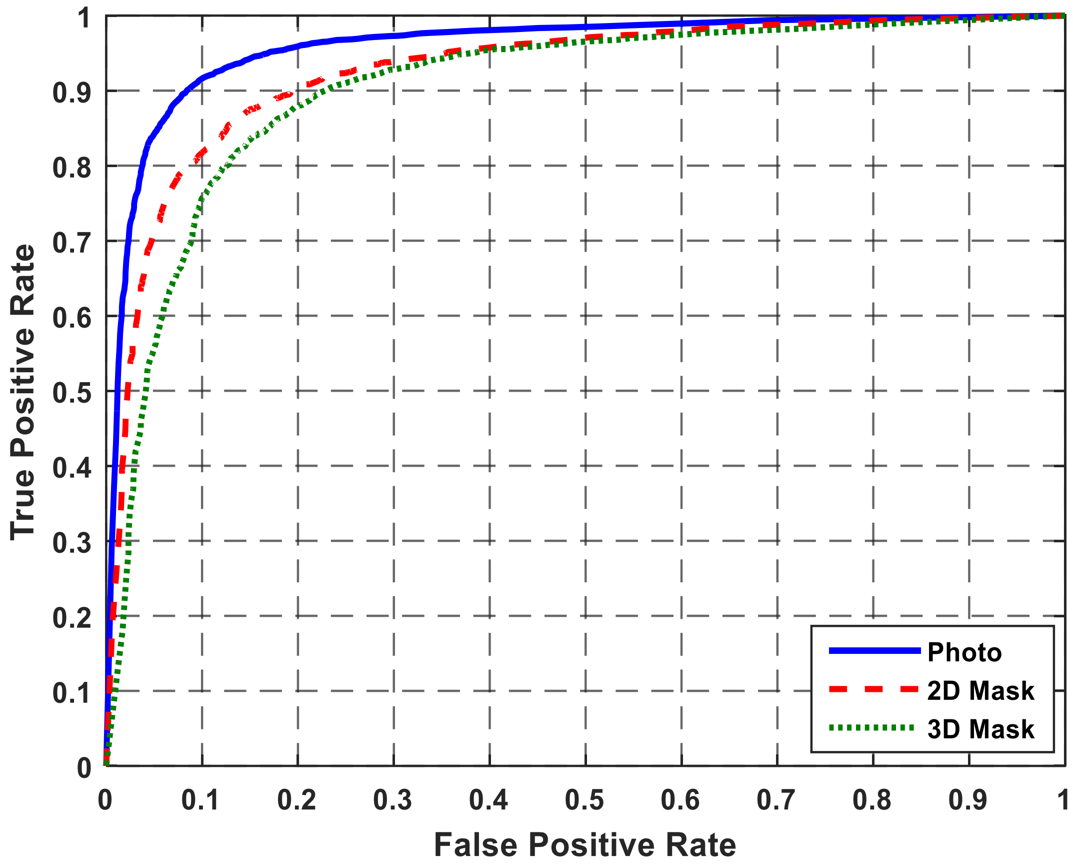


Figure 5.13: ROC curves for photo, 2D mask and 3D mask, using data captured during Lines challenge for gaze alignment consistency.

Table 5.5: Performance of the system at APCER = 0.10, using data captured during Lines challenge for gaze alignment consistency at various challenge durations.

Attack type	**5 seconds**		**7 seconds**		**10 seconds**	
	BPCER	ACER	BPCER	ACER	BPCER	ACER
Photo	0.09	0.10	0.08	0.09	0.05	0.07
2D mask	0.19	0.15	0.15	0.13	0.14	0.12
3D mask	0.24	0.17	0.21	0.16	0.20	0.15
Overall	0.24	0.17	0.21	0.16	**0.20**	**0.15**

5.4.3 Gaze Time Dynamics Features

This section presents results for gaze time dynamics features when discriminating between genuine and impostor attempts using the MKGDD. As before, the user's gaze was directed to random positions on the mobile screen, and the user's pupil coordinates were extracted and used to calculate the proposed features.

Here, DTW was used to measure the dissimilarity between gaze and challenge trajectories after these were optimally aligned. A feature vector F_{DTW} was based on the distance calculated using DTW. As gaze DTW distances were extracted from the left eye (DTW_l) and right eye (DTW_r), feature vectors could be constructed for the right eye F_{DTW_r} and for the left eye F_{DTW_l} for use in presentation attack detection:

$$F_{DTW_r} = [DTW_r] \tag{5.1}$$

$$F_{DTW_l} = [\ DTW_l] \tag{5.2}$$

Further details regarding feature calculation can be found in section 4.4.1.

Figure 5.14 shows ROC curves for displayed photo, 2D mask and 3D mask attacks, using the proposed feature extracted from the dataset as captured during **Lines** challenges of 5 seconds duration. At FPR = 10%, TPRs were approximately 85%, 80% and 70% for displayed photo, 2D mask and 3D mask, respectively. Detection of the mask attack proved difficult; as noted earlier, it may be easier for the attacker to adjust the mask to their eye positions and so to more accurately follow the stimulus.

Overall performance for the proposed features tested on the MKGDD database for different challenge durations using the newly standardized metrics are summarised in Table 5.6. The reported results show that challenge length affects overall system performance, which generally improves slightly as challenge duration increases. At APCER = 10%, ACER increased from 16% to 20% when challenge duration decreased from 10 to 5 seconds. When challenge duration was 7 seconds, overall system performance was 19% and 28% for ACER and BPCER, respectively, at APCER = 10%.

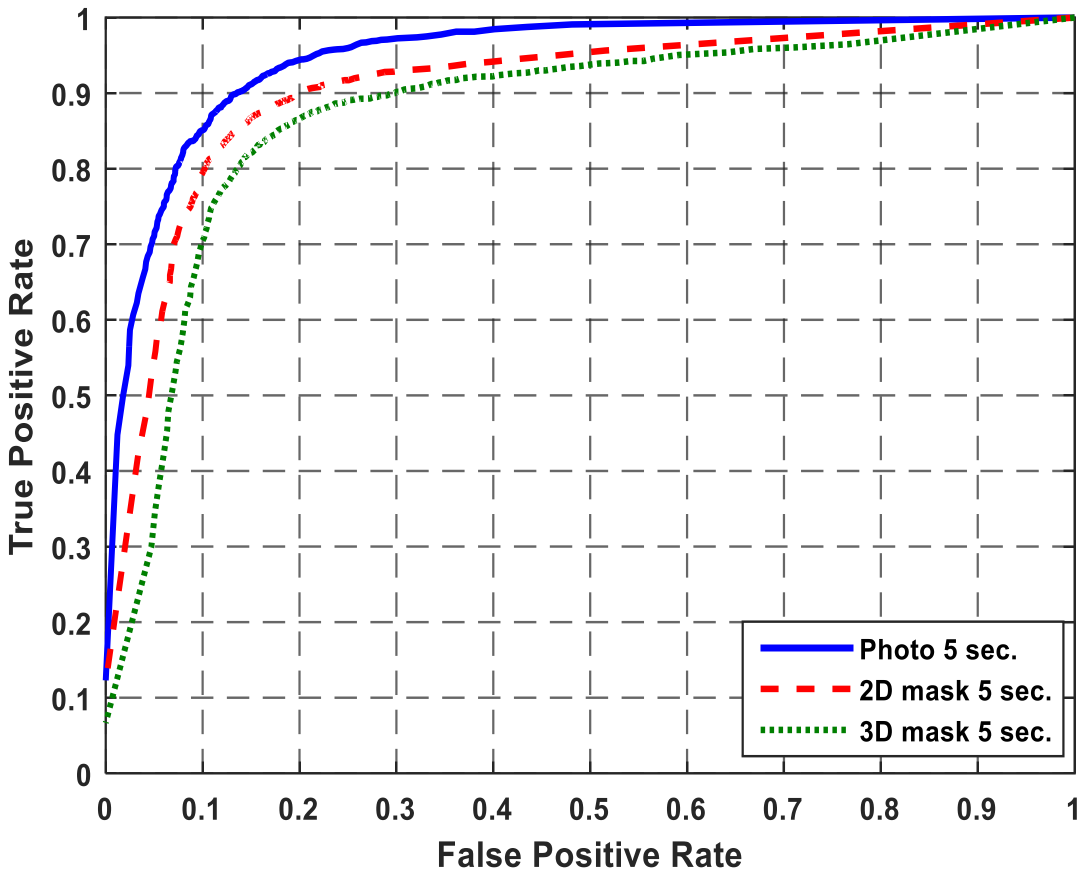


Figure 5.14: ROC curves for photo, 2D mask and 3D mask, using data captured during Lines challenge for gaze DTW feature.

Table 5.6: Performance of the system at APCER = 0.10, using data captured during Lines challenge for gaze DTW feature at various challenge durations.

Attack type	5 seconds		7 seconds		10 seconds	
	BPCER	**ACER**	**BPCER**	**ACER**	**BPCER**	**ACER**
Photo	0.15	0.13	0.12	0.11	0.09	0.10
2D mask	0.20	0.15	0.18	0.14	0.16	0.13
3D mask	0.30	0.20	0.28	0.19	0.23	0.16
Overall	0.30	0.20	0.28	0.19	**0.23**	**0.16**

5.5 Influence of Mobile Device on System Performance

This section investigates the effect of using mobile device on the proposed system by comparing the performance of gaze features collected using a desktop computer and those collected using a mobile device.

Table 5.7 presents TPRs at FPRs 10% for gaze correlation features extracted from the two databases capturing during **Lines** and **Points** challenges of 5 seconds duration. Tables 5.8 and 5.9 show the results for gaze alignment and gaze DTW features, respectively, which were extracted from the two databases captured during a **Lines** challenge. The results in Table 5.6 show that use of a mobile device causes degradation in performance. For gaze correlation features extracted during the Lines challenge and tested on the MKGDD, performance decreased from 90% when tested on the KGDD to 81% TPR for displayed photo detection. For gaze alignment and gaze DTW features, the TPRs dropped from 95% to 91% and from 85% to 84%, respectively.

Based on the above observations, it can be concluded that performance of the proposed features was affected by use of the mobile device and requires further attention. One possible explanation is that it is much more challenging to track the user's gaze on a mobile device than on a desktop computer because both user and device are moving. To explore the motion of user and device, Figure 5.15 and Figure 5.16 show examples of acceleration and gyroscope data of one attempt respectively. From the figures, it can be seen that there are fluctuations in axes which indicate the device movements. To eliminate this problem, in future research, movements of the device and of the user's hands should be separated, using data from the mobile's various sensors, including the gyroscope and accelerometer, to filter out unwanted information.

Table 5.7: System performance at FPR = 0.10 on the KGDD and MKGDD databases for Lines and Points challenges using gaze correlation analysis.

Database	PAI	Lines challenge	Points challenge
		TPR	TPR
KGDD	Displayed photo	0.90	0.91
	2D mask	0.85	0.90
	3D mask	0.79	0.80
MKGDD	Displayed photo	0.81	0.85
	2D mask	0.75	0.80
	3D mask	0.68	0.77

Table 5.8: System performance at FPR = 0.10 on the KGDD and MKGDD databases for Lines challenge using gaze alignment.

Database	PAI	TPR
KGDD	Displayed photo	0.95
	2D mask	0.93
	3D mask	0.90
MKGDD	Displayed photo	0.91
	2D mask	0.81
	3D mask	0.77

Table 5.9: System performance at FPR = 0.10 on the KGDD and MKGDD databases for Lines challenge using gaze time dynamic features.

Database	PAI	TPR
KGDD	Displayed photo	0.85
	2D mask	0.82
	3D mask	0.78
MKGDD	Displayed photo	0.84
	2D mask	0.80
	3D mask	0.71

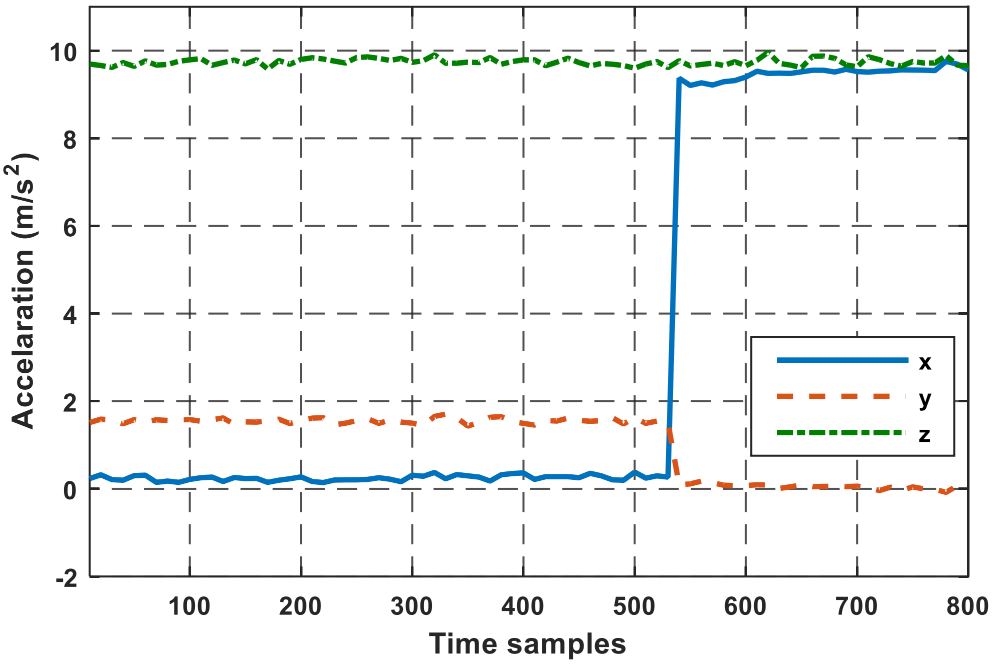


Figure 5.15: The acceleration data extracted from the mobile's sensor when a genuine user holding the device and following the challenge by his gaze

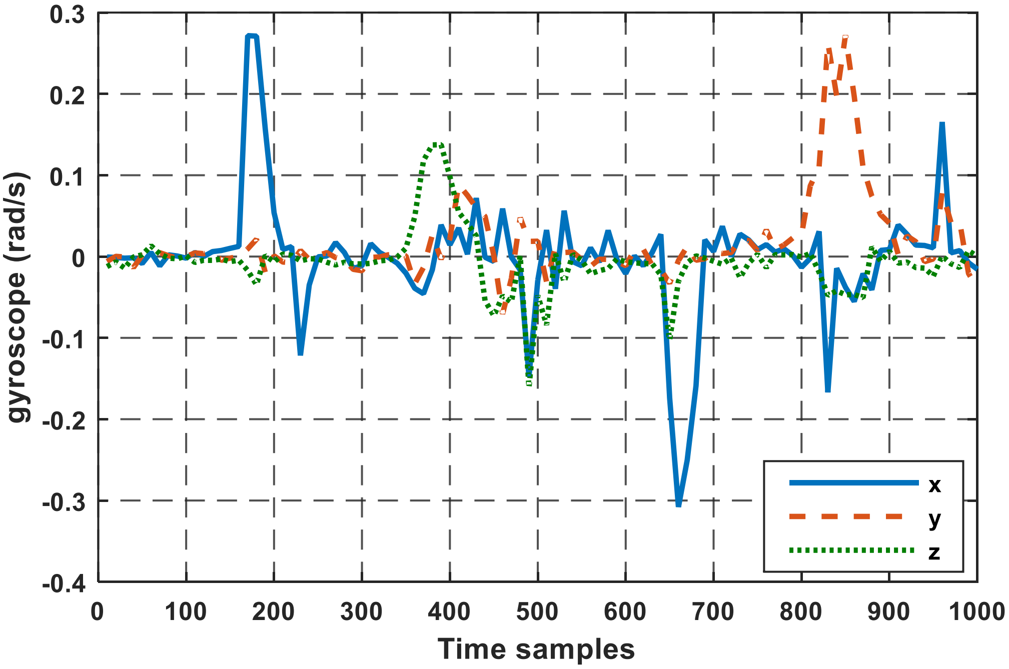


Figure 5.16: The gyroscope data extracted from the mobile's sensor when a genuine user holding the device and following the challenge by his gaze

5.6 Summary

This chapter introduces the new Mobile Kent Gaze Dynamic Database (MKGDD), which was collected from 30 participants using a mobile device. The data collection application developed to display the challenge and to collect participant gaze information was installed on a Google Nexus 6 phone. Data were collected from genuine attempts to track a moving visual target ('challenge') using natural head/eye movements and from impostor attacks, where users held a displayed photo, looked through a 2D mask and held a 3D mask of a genuine user while attempting to follow the stimulus. Experiments to investigate the feasibility of gaze-based features on a real handheld device were conducted using the MKGDD.

As the experimental observations indicate that use of a handheld device impacts on the accuracy of gaze-based features, future research should consider using data from the mobile's various sensors, including the gyroscope and accelerometer, to detect and eliminate the motion of the device and the user's hand.

The chapter's main contributions are as follows:

- Collection of the new database (MKGDD) using a mobile handheld device.
- Testing three novel gaze-based features on the MKGDD database.
- Investigating the effect of using a handheld device on PAD accuracy.

.

Chapter 6

Conclusions and Recommendations for Future Work

This chapter includes a summary of the work completed for this thesis, followed by a discussion of the principal research findings and recommendations for future work.

6.1 Summary of the Research

The work presented in this thesis relates to the security of biometric systems and, in particular, of face recognition systems. The proposed challenge-response presentation attack detection system was used to assess the user’s gaze in response to a randomly moving stimulus on the screen. The user was required to visually track the moving stimulus using natural head/eye movements, and the camera captured facial images at each challenge location. From these images, facial landmarks were extracted and used for feature extraction. Three gaze-based features were proposed for the purpose of face presentation attack detection, and a comprehensive experimental analysis was performed for each feature. The results confirm that the proposed gaze-based features were effective in discriminating between genuine and presentation attack attempts. Details of the completed work are as follows.

- Chapter 2 presented a comprehensive overview of face presentation attacks, including an introduction to relevant instruments. A number of algorithms have been reported in the literature addressing the problem of face presentation attack. Methods of detecting face presentation attacks were assigned to two main categories and further subdivided into groups, based on the cues used to distinguish between real access and spoofing attacks. The chapter also detailed publicly available databases used in the literature.

- Chapter 3 described the design and implementation of the proposed face presentation attack detection system, including further details of the hardware and software used to conduct the experiments. Challenge-response software was designed for the proposed system. The challenge was presented to each user on a display screen as a visual stimulus following three different randomized trajectories: Lines, Curves, and Points. As the stimulus changed location, the user was instructed to follow it with their gaze, using natural head/eye movements, and the camera captured facial images at the various locations of the stimulus on the screen. A new database, the Kent Gaze Dynamic Database (KGDD), was assembled from a sample of 80 participants. This database comprises 2,400 sets of genuine and attack attempts, where participants were assumed to be using a mobile device with two possible screen sizes for biometric authentication. The KGDD contains gaze information and associated evaluation protocols for three types of attack instrument: displayed photo, 2D mask, and 3D mask. The evaluation strategies used for this purpose were also specified.

- Chapter 4 investigated three gaze-based features for face presentation attack detection: gaze correlation analysis, gaze alignment consistency, and gaze time dynamics. The proposed features were based on the assumption that eye/head movements when a genuine attempt is made will differ significantly from certain types of presentation attack attempt. Features were extracted from facial images captured at each challenge location. The experimental results confirmed that the proposed features were effective in detecting presentation attacks. A comparison with state-of-the-art algorithms demonstrated that the proposed algorithm achieves generally improved performance.

- Chapter 5 described how the new Mobile Kent Gaze Dynamic Database (MKGDD) was collected from 30 participants using a handheld mobile device. An Android application was developed to display the challenge and to collect gaze information from the participants. The data collection application was uploaded to a Google Nexus 6. Data were collected from genuine attempts, where users employed natural head/eye movements to track a moving visual target or 'challenge', and from impostor attacks, where users held a displayed photo, looked through a 2D mask, or held a 3D mask of a genuine user while attempting to follow the stimulus. Several sets of experiments were performed using the MKGDD to investigate the viability of gaze-based features on a real handheld device.

6.2 Key Findings

Based on the experimental observations completed for this thesis, the main conclusions can be summarised as follows.

- *Effect of PAIs on performance*. For all challenge types, the gaze-based features generally performed best in detecting displayed photo attacks on both databases (KGDD and MKGDD) as compared to the other PAIs (2D mask and 3D mask attacks). There are two possible explanations. (1) It is easier for the attacker to adjust the mask to fit their eye positions and so track the stimulus more accurately. (2) In the case of a displayed photo attack, the fact that visually guided hand movements are required to follow the challenge means there is less similarity between the trajectories of challenge and hand movement.

- *Effect of screen size*. The gaze-based features tested on the KGDD generally performed better when the challenge was displayed on a large screen (Tablet format) as compared to small screen (Phone format).

- *Impact of challenge duration on performance*. The system performed better when gaze features were based on a longer challenge duration, implying that the data contained a better excerpt of what the trajectories look like. This observation confirmed that there is a trade-off between challenge duration and performance.

- *Effect of challenge trajectory types*. Performance was similar for Lines and Curves challenges. In most cases, the system performed slightly better on Points challenges.

- *Impact of tinted glasses on system performance*. The system's performance was found to be worse when genuine users attempted to access the system while wearing tinted glasses. This indicates that tinted glasses had a significant influence on the results for gaze-based features.

- In comparison to state-of-the-art methods, gaze-based methods were able to achieve comparable performance on the KGDD database.

- A handheld device affect the accuracy of the proposed gaze-based features.

6.3 Main Contributions

The main contributions of this thesis can be summarized as follows.

1. A comprehensive review of recent advances in face-PAD algorithms and existing face-PAD databases.

2. Development of challenge-response software with three different challenge trajectories: Lines, Curves, and Points.

3. A newly compiled database (KGDD) containing gaze information and associated evaluation protocols, with three types of attack artefact (displayed photo, 2D

mask, and 3D mask) and two different screen sizes displaying three types of challenge.

4. Three gaze-based features for face-PAD, including gaze correlation analysis, gaze alignment consistency, and gaze time dynamics.

5. A comparative analysis encompassing state-of-the-art face-PAD techniques.

6. A second database (MKGDD) using a handheld device.

6.4 Recommendations for Future Work

Future research can build on the work presented in this thesis to further improve the performance of face-PAD.

First, better performance may be achieved by fusing the gaze-based features investigated here. The effectiveness of the three proposed features (gaze correlation analysis, gaze alignment consistency, and gaze time dynamics) should be explored in combination for detection of presentation attacks. Different fusion schemes (feature-based and score-based fusion) may be considered to achieve improvements.

Second, more improvements can be made to the feature extraction stage of the proposed system. As shown in Chapter 4, the three proposed features depend on facial landmark detection. It follows that more robust landmark detection algorithms can be used to extract gaze information more accurately. Another fruitful area of research would be to investigate novel features such as duration of fixation and user saccades in following the challenge. Other features worth investigating include acceleration between fixation and saccade.

Third, future work might also consider designing the challenge system so that challenge duration is minimised while maintaining high accuracy of presentation attack detection.

Fourth, future research should use data from the mobile’s various sensors, including the gyroscope and accelerometer, to detect and eliminate the motion of the device and of the user’s hand.

Fifth, data from mobile sensors that record device motion, such as the gyroscope and accelerometer, could be used as features to discriminate between genuine and impostor attempts.

Sixth, only one mobile device was used to collect the data from 30 participants. In the future, it would be useful to collect more data using different mobile devices that can record videos with different qualities. Additionally, data should be collected under different illumination conditions and with different background scenes.

References

[1] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: a tool for information security," *IEEE Transactions on Information Forensics and Security,* vol. 1, pp. 125-143, 2006.

[2] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of biometrics*: Springer, 2008.

[3] D. Gafurov, "A survey of biometric gait recognition: Approaches, security and challenges," in *Annual Norwegian computer science conference,* 2007, pp. 19-21.

[4] D. S. Matovski, M. S. Nixon, S. Mahmoodi, and J. N. Carter, "The effect of time on gait recognition performance," *IEEE transactions on information forensics and security,* vol. 7, pp. 543-552, 2012.

[5] J. E. Boyd and J. J. Little, "Biometric gait recognition," in *Advanced Studies in Biometrics*, Springer, 2005, pp. 19-42.

[6] J. A. Markowitz, "Voice biometrics," *Communications of the ACM,* vol. 43, pp. 66-73, 2000.

[7] D. A. Reynolds, "An overview of automatic speaker recognition technology," in *ICASSP*, 2002, pp. 4072-4075.

[8] J. Mantyjarvi, J. Koivumaki, and P. Vuori, "Keystroke recognition for virtual keyboard," in *proceedings of IEEE International Conference onMultimedia and Expo (ICME'02. P),* 2002, pp. 429-432..

[9] A. K. Jain, F. D. Griess, and S. D. Connell, "On-line signature verification," *Pattern recognition,* vol. 35, pp. 2963-2972, 2002.

[10] D.-Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll, "SVC2004: First international signature verification competition," in *Biometric Authentication*, Springer, 2004, pp. 16-22.

[11] F. Alonso-Fernandez, M. C. Fairhurst, J. Fierrez, and J. Ortega-Garcia, "Automatic measures for predicting performance in off-line signature," in *IEEE International Conference on Image Processing (ICIP)*, 2007, pp. I-369-I-372.

[12] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on circuits and systems for video technology,* vol. 14, pp. 4-20, 2004.

[13] K. Delac and M. Grgic, "A survey of biometric recognition methods," in *Proceedings of 46th International Symposium Electronics in Marine*, 2004, pp. 184-193.

[14] J. Galbally, J. Fierrez, and J. Ortega-García, "Vulnerabilities in biometric systems: Attacks and recent advances in liveness detection," *Database,* vol. 1, p. 4, 2007.

[15] K. A. Nixon, V. Aimale, and R. K. Rowe, "Spoof detection schemes," in *Handbook of biometrics*, ed: Springer, 2008, pp. 403-423.

[16] N. K. Ratha, J. H. Connell, and R. M. Bolle, "An analysis of minutiae matching strength," in *International Conference on Audio-and Video-Based Biometric Person Authentication*, 2001, pp. 223-228.

[17] ISO/IEC JTC 1/SC 37 Biometrics. Information technology – Biometric presentation attack detection – Part 1: Framework. International Organization for Standardization. Available: https://www.iso.org/obp/ui/#iso:std:iso-iec:30107:-1:ed-1:v1:en , accessed Jan. 2017.

[18] S. Marcel and N. Erdogmus, "Introduction," in *Handbook of Biometric Anti-Spoofing: Trusted Biometrics under Spoofing Attacks*, S. Marcel and N. Erdogmus, Eds., ed London: Springer London, 2014, pp. 1-11.

[19] J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," *IEEE Access,* vol. 2, pp. 1530-1552, 2014.

[20] Y. Li, Y. Li, K. Xu, Q. Yan, and R. Deng, "Empirical study of face authentication systems under OSNFD attacks," *IEEE Transactions on Dependable and Secure Computing*, 2016, pp. 223-228.

[21] I. B. Group, "Biometrics Market and Industry Report 2009-2014," Available: http://www.biometricgroup.com/reports/public/market_report. php [consultado el 20-07-09], accessed May 2018.

[22] B. Gipp, J. Beel, and I. Rössling, "ePassport: The World's New Electronic Passport," *A Report about the ePassport's Benefits, Risks and it's Security. CreateSpace*, 2007.

[23] S. Chakraborty and D. Das, "An Overview of Face Liveness Detection," *International Journal on Information Theory (IJIT)*, , vol. 3, no. 2, pp. 11–25, April 2014.

[24] J. Yang, Z. Lei, D. Yi, and S. Z. Li, "Person-specific face antispoofing with subject domain adaptation," *IEEE Transactions on Information Forensics and Security,* vol. 10, pp. 797-809, 2015.

[25] A. Ali, N. Alsufyani, S. Hoque, and F. Deravi, "Biometric Counter-Spoofing for Mobile Devices Using Gaze Information," in *International Conference on Pattern Recognition and Machine Intelligence*, 2017, pp. 11-18.

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

[27] J. Komulainen, "Software-Based Countermeasures to 2d Facial Spoofing Attacks," *PhD. thesis,* University of Oulu, Finland, 2015.

[28] P. Tome, M. Vanoni, and S. Marcel, "On the vulnerability of finger vein recognition to spoofing," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2014, pp. 1-10.

[29] N. Duc and B. Minh, "Your face is not your password," in *Black Hat Conference*, 2009.

[30] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *5th IAPR international conference on Biometrics (ICB)*, 2012, pp. 26-31.

[31] N. Erdogmus and S. Marcel, "Spoofing Face Recognition With 3D Masks," *IEEE Transactions on Information Forensics and Security*, vol. 9, pp. 1084-1097, 2014.

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

[33] O. Kähm and N. Damer, "2D face liveness detection: An overview," in *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2012, pp. 1-12.

[34] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of fourier spectra," in *Biometric Technology for Human Identification*, 2004, pp. 296-304.

[35] N. Kose and J. Dugelay, "Classification of captured and recaptured images to detect photograph spoofing," in *International Conference on Informatics, Electronics & Vision (ICIEV)*, 2012, pp. 1027-1032.

[36] W. R. Schwartz, A. Rocha, and H. Pedrini, "Face spoofing detection through partial least squares and low-level descriptors," in *International Joint Conference on Biometrics (IJCB)*, 2011, pp. 1-8.

[37] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE transactions on information forensics and security*, vol. 11, pp. 2268-2283, 2016.

[38] J. Galbally, S. Marcel, and J. Fierrez, "Biometric Antispoofing Methods: A Survey in Face Recognition," *Access, IEEE*, vol. 2, pp. 1530-1552, 2014.

[39] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using texture and local shape analysis," *IET biometrics*, vol. 1, pp. 3-10, 2012.

[40] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *Computer Vision–ECCV 2010*, ed: Springer, 2010, pp. 504-517.

[41] B. Peixoto, C. Michelassi, and A. Rocha, "Face liveness detection under bad illumination conditions," in *18th IEEE International Conference on Image Processing (ICIP)*, 2011, pp. 3557-3560.

[42] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: a public database and a baseline," in *international joint conference on Biometrics (IJCB)*, 2011, pp. 1-7.

[43] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," in *international joint conference on Biometrics (IJCB)*, 2011, pp. 8-15.

[44] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human detection using partial least squares analysis," in *12th international conference on Computer vision*, 2009, pp. 24-31.

[45] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics,* vol. 3, pp. 610-621, 1973.

[46] W. R. Schwartz, R. D. Da Silva, L. S. Davis, and H. Pedrini, "A novel feature descriptor based on the shearlet transform," in *18th IEEE International Conference on Image Processing (ICIP)*, 2011, pp. 1033-1036.

[47] H. Wold, "Partial least squares," *Encyclopedia of statistical sciences,* vol. 9, pp. 125-138, 2004.

[48] A. Pinto, W. R. Schwartz, H. Pedrini, and A. d. R. Rocha, "Using Visual Rhythms for Detecting Video-Based Facial Spoof Attacks," *IEEE Transactions on Information Forensics and Security,* vol. 10, pp. 1025-1038, 2015.

[49] T. de Freitas Pereira, J. Komulainen, A. Anjos, J. M. De Martino, A. Hadid, M. Pietikäinen, and S. Marcel, "Face liveness detection using dynamic texture," *EURASIP Journal on Image and Video Processing,* vol. 2014, p. 2, 2014.

[50] K. Wonjun, S. Sungjoo, and H. Jae-Joon, "Face Liveness Detection From a Single Image via Diffusion Speed Model," *IEEE Transactions on Image Processing,* vol. 24, pp. 2456-2465, 2015.

[51] S. Rahimzadeh Arashloo, J. Kittler, and B. Christmas, "Face Spoofing Detection Based on Multiple Descriptor Fusion Using Multiscale Dynamic Binarized

Statistical Image Features," *IEEE Transactions on Information Forensics and Security*, vol. PP, pp. 1-1, 2015.

[52] S. R. Arashloo and J. Kittler, "Dynamic Texture Recognition Using Multiscale Binarized Statistical Image Features," *IEEE Transactions on Multimedia*, vol. 16, pp. 2099-2109, 2014.

[53] Q. Zhen, D. Huang, Y. Wang, and L. Chen, "LPQ Based Static and Dynamic Modeling of Facial Expressions in 3D Videos," in *Biometric Recognition*. vol. 8232, Z. Sun, S. Shan, G. Yang, J. Zhou, Y. Wang, and Y. Yin, Eds., ed: Springer International Publishing, 2013, pp. 122-129.

[54] J. Bihan, M. Valstar, B. Martinez, and M. Pantic, "A Dynamic Appearance Descriptor Approach to Facial Actions Temporal Modeling," *IEEE Transactions on Cybernetics*, vol. 44, pp. 161-174, 2014.

[55] D. Caetano Garcia and R. L. de Queiroz, "Face-Spoofing 2D-Detection Based on Moiré-Pattern Analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, pp. 778-786, 2015.

[56] B. Weyrauch, B. Heisele, J. Huang, and V. Blanz, "Component-based face recognition with 3D morphable models," in *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*, 2004, pp. 85-85.

[57] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pp. 684-698, 2005.

[58] Frontal Face Dataset From the Computational Group at Caltech [Online]. Available: http://www.vision.caltech.edu/archive.html, accessed Jan. 2018.

[59] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security,* vol. 10, pp. 746-761, 2015.

[60] Y. Jianwei, L. Zhen, L. Shengcai, and S. Z. Li, "Face liveness detection with component dependent descriptor," in *International Conference on Biometrics (ICB),* 2013, pp. 1-6.

[61] 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," in *2012 5th IAPR International Conference on Biometrics (ICB),* 2012, pp. 67-72.

[62] R. Raghavendra, S. Venkatesh, K. B. Raja, P. Wasnik, M. Stokkenes, and C. Busch, "Fusion of Multi-Scale Local Phase Quantization Features for Face Presentation Attack Detection," in *2018 21st International Conference on Information Fusion (FUSION)*, 2018, pp. 2107-2112.

[63] R. Raghavendra and C. Busch, "Robust 2D/3D face mask presentation attack detection scheme by exploring multiple features and comparison score level fusion," in *17th International Conference on Information Fusion (FUSION),* 2014, pp. 1-7.

[64] J. Kannala and E. Rahtu, "Bsif: Binarized statistical image features," in *21st International Conference on Pattern Recognition (ICPR)*, 2012, pp. 1363-1366.

[65] T.-W. Lee, G.-H. Ju, H.-S. Liu, and Y.-S. Wu, "Liveness detection using frequency entropy of image sequences," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 2013, pp. 2367-2370.

[66] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks,* vol. 13, pp. 411-430, 2000.

[67] L. Wu, X. Xu, Y. Cao, Y. Hou, and W. Qi, "Live Face Detection by Combining the Fourier Statistics and LBP," in *Biometric Recognition*, ed: Springer, 2014, pp. 173-181.

[68] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Transactions on Information Forensics and Security*, vol. 11, pp. 1818-1830, 2016.

[69] Z. Boulkenafet, J. Komulainen, and A. Hadid, "On the generalization of color texture-based face anti-spoofing," *Image and Vision Computing*, vol. 77, pp. 1-9, 2018.

[70] Z. Boulkenafet, J. Komulainen, X. Feng, and A. Hadid, "Scale space texture analysis for face anti-spoofing," in *International Conference on Biometrics (ICB)*, 2016, pp. 1-6.

[71] A. Witkin, "Scale-space filtering: A new approach to multi-scale description," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1984, pp. 150-153.

[72] E. H. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proceedings of the national academy of sciences*, vol. 83, pp. 3078-3080, 1986.

[73] A. Sepas-Moghaddam, F. Pereira, and P. L. Correia, "Light Field-Based Face Presentation Attack Detection: Reviewing, Benchmarking and One Step Further," *IEEE Transactions on Information Forensics and Security*, vol. 13, pp. 1696-1709, 2018.

[74] R. Raghavendra, K. B. Raja, and C. Busch, "Exploring the Usefulness of Light Field Cameras for Biometrics: An Empirical Study on Face and Iris Recognition," *IEEE Trans. Information Forensics and Security*, vol. 11, pp. 922-936, 2016.

[75] A. Sepas-Moghaddam, P. L. Correia, and F. Pereira, "Light field local binary patterns description for face recognition," in *IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3815-3819.

[76] A. Sepas-Moghaddam, V. Chiesa, P. L. Correia, F. Pereira, and J.-L. Dugelay, "The IST-EURECOM light field face database," in in *5th International Workshop on Biometrics and Forensics (IWBF)*, 2017, pp. 1-6.

[77] S. Kim, Y. Ban, and S. Lee, "Face liveness detection using a light field camera," *Sensors*, vol. 14, pp. 22471-22499, 2014.

[78] R. Raghavendra, K. B. Raja, and C. Busch, "Presentation attack detection for face recognition using light field camera," *IEEE Transactions on Image Processing*, vol. 24, pp. 1060-1075, 2015.

[79] Z. Ji, H. Zhu, and Q. Wang, "LFHOG: a discriminative descriptor for live face detection from light field image," in *IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 1474-1478.

[80] A. Sepas-Moghaddam, L. Malhadas, P. L. Correia, and F. Pereira, "Face spoofing detection using a light field imaging framework," *IET Biometrics*, vol. 7, pp. 39-48, 2017.

[81] R. Raghavendra and C. Busch, "Presentation attack detection algorithm for face and iris biometrics," in *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*, 2014, pp. 1387-1391.

[82] S. Çakir and A. E. Çetin, "Mel-cepstral methods for image feature extraction," in *17th IEEE International Conference on Image Processing (ICIP)*, 2010, pp. 4577-4580.

[83] K. Kollreider, H. Fronthaler, and J. Bigun, "Verifying liveness by multiple experts in face biometrics," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2008, pp. 1-6.

[84] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in “liveness” assessment," *IEEE Transactions on Information Forensics and Security,* vol. 2, pp. 548-558, 2007.

[85] L. Sun, G. Pan, Z. Wu, and S. Lao, "Blinking-based live face detection using conditional random fields," in *Advances in Biometrics,* ed: Springer, 2007, pp. 252-260.

[86] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences,* vol. 55, pp. 119-139, 1997.

[87] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE,* vol. 77, pp. 257-286, 1989.

[88] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," 2007.

[89] G. Pan, L. Sun, Z. Wu, and Y. Wang, "Monocular camera-based face liveness detection by combining eyeblink and scene context," *Telecommunication Systems,* vol. 47, pp. 215-225, 2011.

[90] J. Yan, Z. Zhang, Z. Lei, D. Yi, and S. Z. Li, "Face liveness detection by exploring multiple scenic clues," in *12th International Conference on Control Automation Robotics & Vision (ICARCV)*, 2012, pp. 188-193.

[91] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *cvpr*, 1999, p. 2246.

[92] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *biometrics*, vol. 81, pp. 425-455, 1994.

[93] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. Ho, "Detection of Face Spoofing Using Visual Dynamics," *IEEE Transactions on Information Forensics and Security*, vol. 10, pp. 762-777, 2015.

[94] J. Komulainen, A. Hadid, M. Pietikainen, A. Anjos, and S. Marcel, "Complementary countermeasures for detecting scenic face spoofing attacks," in *International Conference on Biometrics (ICB)*, 2013, pp. 1-7.

[95] M. De Marsico, M. Nappi, D. Riccio, and J. Dugelay, "Moving face spoofing detection via 3D projective invariants," in *5th IAPR International Conference on Biometrics (ICB)*, 2012, pp. 73-78.

[96] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Visual tracking and recognition using probabilistic appearance manifolds," *Computer Vision and Image Understanding*, vol. 99, pp. 303-331, 2005.

[97] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, "Computationally efficient face spoofing detection with motion magnification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 105-110.

[98] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, "Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1932-1939.

[99] L. Wu, Y. Xu, M. Jian, W. Cai, C. Yan, and Y. Ma, "Motion Analysis Based Cross-Database Voting for Face Spoofing Detection," in *Chinese Conference on Biometric Recognition*, 2017, pp. 528-536.

[100] A. Lagorio, M. Tistarelli, M. Cadoni, C. Fookes, and S. Sridharan, "Liveness detection based on 3D face shape analysis," in *International Workshop on Biometrics and Forensics (IWBF)*, 2013, pp. 1-4.

[101] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection using 3D structure recovered from a single camera," in *International Conference on Biometrics (ICB)*, 2013, pp. 1-6.

[102] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," *International Journal of Computer Vision,* vol. 91, pp. 200-215, 2011.

[103] S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognition,* vol. 46, pp. 1415-1432, 2013.

[104] B. Billiot, F. Cointault, L. Journaux, J.-C. Simon, and P. Gouton, "3D image acquisition system based on shape from focus technique," *Sensors,* vol. 13, pp. 5040-5053, 2013.

[105] R. Veerender, K. Acharya, J. Srinivas, and D. Mohan, "Depth Estimation Using Blur Estimation in Video," *Int. J. Electron. Comput. Sci. Eng,* vol. 1, pp. 2350-2354, 2012.

[106] X. Xie, Y. Gao, W.-S. Zheng, J. Lai, and J. Zhu, "One-Snapshot Face Anti-spoofing Using a Light Field Camera," in *Chinese Conference on Biometric Recognition*, 2017, pp. 108-117.

[107] S. Kim, Y. Ban, and S. Lee, "Face Liveness Detection Using Defocus," *Sensors,* vol. 15, pp. 1537-1563, 2015.

[108] K. Sooyeon, Y. Sunjin, K. Kwangtaek, B. Yuseok, and L. Sangyoun, "Face liveness detection using variable focusing," in *International Conference on Biometrics (ICB)*, 2013, pp. 1-6.

[109] S. K. Nayar, "Shape from focus," *IEEE transactions on pattern analysis and machine intelligence,* vol. 16, no. 8, pp. 824-831, 1994.

[110] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence,* vol. 27, pp. 1615-1630, 2005.

[111] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery,* vol. 2, pp. 121-167, 1998.

[112] R. F. Nogueira, R. de Alencar Lotufo, and R. C. Machado, "Fingerprint Liveness Detection Using Convolutional Neural Networks," *IEEE Transactions on Information Forensics and Security,* vol. 11, pp. 1206-1213, 2016.

[113] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcao, and A. Rocha, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Transactions on Information Forensics and Security,* vol. 10, pp. 864-879, 2015.

[114] J. Yang, Z. Lei, and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," *arXiv preprint arXiv:1408.5601,* 2014.

[115] L. Li, X. Feng, Z. Xia, X. Jiang, and A. Hadid, "Face spoofing detection with local binary pattern network," *Journal of Visual Communication and Image Representation,* vol. 54, pp. 182-192, 2018.

[116] M. Vatsa, R. Singh, and A. Majumdar, *Deep Learning in Biometrics*: CRC Press, 2018.

[117] H. Li, W. Li, H. Cao, S. Wang, F. Huang, and A. C. Kot, "Unsupervised domain adaptation for face anti-spoofing," *IEEE Transactions on Information Forensics and Security,* vol. 13, pp. 1794-1809, 2018.

[118] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, and A. Majumdar, "Detecting silicone mask-based presentation attack via deep dictionary learning," *IEEE Transactions on Information Forensics and Security,* vol. 12, pp. 1713-1723, 2017.

[119] H. Li, P. He, S. Wang, A. Rocha, X. Jiang, and A. C. Kot, "Learning Generalized Deep Feature Representation for Face Anti-Spoofing," *IEEE Transactions on Information Forensics and Security,* vol. 13, pp. 2639-2652, 2018.

[120] L. Li, Z. Xia, L. Li, X. Jiang, X. Feng, and F. Roli, "Face anti-spoofing via hybrid convolutional neural network," in *International Conference on the Frontiers and Advances in Data Science (FADS),* 2017, pp. 120-124.

[121] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, "Face anti-spoofing using patch and depth-based CNNs," in *IEEE International Joint Conference on Biometrics (IJCB),* 2017, pp. 319-328.

[122] H. Yu, T.-T. Ng, and Q. Sun, "Recaptured photo detection using specularity distribution," in *15th IEEE International Conference on Image Processing (ICIP),* 2008, pp. 3140-3143.

[123] A. M. K. Saad, "Anti-spoofing using challenge-response user interaction," Ph.D. dissertation, Dept. Comp. Science, Eng., American Univ., Cairo, 2015.

[124] C. Kant and N. Sharma, "Fake face detection based on skin elasticity," *International journal of advanced research in computer science and software engineering, 3 (5),* 2013.

[125] K. Kollreider, H. Fronthaler, and J. Bigun, "Evaluating liveness by face images and the structure tensor," in *Fourth IEEE Workshop on Automatic Identification Advanced Technologies,* 2005, pp. 75-80.

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

[127] A. Ali, F. Deravi, and S. Hoque, "Spoofing attempt detection using gaze colocation," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2013, pp. 1-12.

[128] A. Ali, F. Deravi, and S. Hoque, "Liveness Detection Using Gaze Collinearity," in *Third International Conference on Emerging Security Technologies (EST)*, 2012, pp. 62-65.

[129] A. Ali, F. Deravi, and S. Hoque, "Directional Sensitivity of Gaze-Collinearity Features in Liveness Detection," in *Fourth International Conference on Emerging Security Technologies (EST)*, 2013, pp. 8-11.

[130] A. Ali, S. Hoque, and F. Deravi, "Gaze stability for liveness detection," *Pattern Analysis and Applications,* vol. 21, pp. 437-449, 2018.

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

[132] A. Boehm, D. Chen, M. Frank, L. Huang, C. Kuo, T. Lolic, I. Martinovic, and D. Song, "Safe: Secure authentication with face and eyes," in *International Conference on Privacy and Security in Mobile Systems (PRISMS)*, 2013, pp. 1-8.

[133] O. V. Komogortsev, A. Karpov, and C. D. Holland, "Attack of mechanical replicas: Liveness detection with eye movements," *IEEE Transactions on Information Forensics and Security,* vol. 10, pp. 716-725, 2015.

[134] L. Cai, C. Xiong, L. Huang, and C. Liu, "A Novel Face Spoofing Detection Method Based on Gaze Estimation," in *Computer Vision--ACCV 2014*, Springer, 2015, pp. 547-561.

[135] D. F. Smith, A. Wiliem, and B. C. Lovell, "Face Recognition on Consumer Devices: Reflections on Replay Attacks," *IEEE Transactions on Information Forensics and Security,* vol. 10, pp. 736-745, 2015.

[136] P. P. Chan, W. Liu, D. Chen, D. S. Yeung, F. Zhang, X. Wang, and C.-C. Hsu, "Face liveness detection using a flash against 2D spoofing attack," *IEEE Transactions on Information Forensics and Security,* vol. 13, pp. 521-534, 2018.

[137] W. Liu, "Face liveness detection using analysis of Fourier spectra based on hair," in *International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, 2014, pp. 75-80.

[138] J. Peng and P. P. Chan, "Face liveness detection for combating the spoofing attack in face recognition," in *International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, 2014, pp. 176-181.

[139] D. Tang, Z. Zhou, Y. Zhang, and K. Zhang, "Face Flashing: a Secure Liveness Detection Protocol based on Light Reflections," *arXiv preprint arXiv:1801.01949,* 2018.

[140] J. M. Di Martino, Q. Qiu, T. Nagenalli, and G. Sapiro, "Liveness Detection Using Implicit 3D Features," *arXiv preprint arXiv:1804.06702,* 2018.

[141] P. P. Chan and Y. Shu, "Face Liveness Detection by Brightness Difference," in *International Conference on Machine Learning and Cybernetics*, 2014, pp. 144-150.

[142] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *International Conference on Image Analysis and Signal Processing (IASP)*, 2009, pp. 233-236.

[143] L. Sun, G. Pan, Z. Wu, and S. Lao, "Blinking-based live face detection using conditional random fields," in *International Conference on Biometrics*, 2007, pp. 252-260.

[144] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Face liveness detection by learning multispectral reflectance distributions," in *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011)*, 2011, pp. 436-441.

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

[146] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, "OULU-NPU: A Mobile Face Presentation Attack Database with Real-World Variations," in *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp. 612-618.

[147] C. Sanderson, *Biometric person recognition: Face, speech and fusion* vol. 4: VDM Publishing, 2008.

[148] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE transactions on pattern analysis and machine intelligence,* vol. 23, pp. 643-660, 2001.

[149] N. Erdogmus and S. Marcel, "Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect," in *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1-6.

[150] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Incremental face alignment in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1859-1866.

[151] *PRTools*. Available: http://prtools.tudelft.nl/, accessed Jan. 2019.

[152] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern recognition,* vol. 38, pp. 2270-2285, 2005.

[153] M. Vidal, A. Bulling, and H. Gellersen, "Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets," in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, 2013, pp. 439-448.

[154] S. Sirohey, A. Rosenfeld, and Z. Duric, "A method of detecting and tracking irises and eyelids in video," *Pattern Recognition,* vol. 35, pp. 1389-1401, 2002.

[155] T. Soukupova and J. Cech, "Real-time eye blink detection using facial landmarks," in *21st Computer Vision Winter Workshop (CVWW'2016)*, 2016, pp. 1-8.

[156] M.-T. Puth, M. Neuhäuser, and G. D. Ruxton, "Effective use of Pearson's product–moment correlation coefficient," *Animal Behaviour,* vol. 93, pp. 183-189, 2014.

[157] J. Zhao and L. Itti, "shapeDTW: shape Dynamic Time Warping," *arXiv preprint arXiv:1606.01601,* 2016.

[158] P. Sanguansat, "Multiple multidimensional sequence alignment using generalized dynamic time warping," *WSEAS Transactions on Mathematics,* vol. 11, pp. 668-678, 2012.

[159] M. Shokoohi-Yekta, B. Hu, H. Jin, J. Wang, and E. Keogh, "Generalizing DTW to the multi-dimensional case requires an adaptive approach," *Data mining and knowledge discovery,* vol. 31, pp. 1-31, 2017.

[160] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "LBP− TOP based countermeasure against face spoofing attacks," in *Asian Conference on Computer Vision*, 2012, pp. 121-132.

[161] A. Anjos, L. El-Shafey, R. Wallace, M. Günther, C. McCool, and S. Marcel, "Bob: a free signal processing and machine learning toolbox for researchers," in *Proceedings of the 20th ACM international conference on Multimedia*, 2012, pp. 1449-1452.

[162] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *22nd International Conference on Pattern Recognition (ICPR)*, 2014, pp. 1173-1178.

[163] Android Developers. (2018, August. 7). *Motion sensors* [Online]. Available: https://developer.android.com/guide/topics/sensors/sensors_motion, accessed May 2018.

[164] T.-M. Gronli, J. Hansen, G. Ghinea, and M. Younas, "Mobile application platform heterogeneity: Android vs Windows Phone vs iOS vs Firefox OS," in *Advanced Information Networking and Applications (AINA), 2014 IEEE 28th International Conference on*, 2014, pp. 635-641.

[165] GSMArena.com. M*OTOROLA GOOGLE NEXUS 6* [Online]. Available: https://www.gsmarena.com/motorola_nexus_6-6604.php, accessed May 2018.

Appendix A

Participant Information Sheet

Project title: Face Presentation attack Detection for Mobile Devices

Version 1.0 Date: ___________________________

What is the purpose of the research?

The aim of the project is to establish a new mechanism for detecting whether a fraudulent attempt is made at accessing information through the use of artefacts (e.g. photos or masks) to spoof biometric authentication systems. The findings from this study will be used to design and develop algorithms for biometric recognition systems for mobile applications to counteract such attacks. These technologies will help secure future online transactions and accessing devices such as mobile phones.

Why have I been invited to take part?

You have been selected as a part of a volunteer sample. Any individual with normal vision and over the age of 18 years only is welcome to attend the data collection and try using the proposed systems both as a genuine user and as a potential attacker using various artefacts such as masks and photo projection using a mobile device.

Do I have to take part?

Your participation is entirely voluntary. It is up to you to decide.

What will I be asked to do if I take part?

You will be seated in front of the computer screen following a stimulus with natural head/eye movements, wearing/holding 2D/3D masks or projecting a photo. You will be shown the stimulus at a number of different places on the screen which you should follow as best as you can. Each attempt will be followed by a short cool down period before the next attempt begins. A video recording is made of your attempt at following the stimulus. This video data is used for the evaluation of the proposed systems. The data collection session is designed to last no more than 45 minutes. You only need to make one visit to complete your contribution to this project. We will give you an exit questionnaire to be completed at the end of the data collection session. You can miss out any questions that you do not feel comfortable answering.

What are the benefits of taking part?

A £10 Amazon Voucher will be given as reward for your participation after the end of the data collection session.

Are there any risks involved?

There are no risk associate with participating in this study.

However, face images are captured and although stored anonymously, may be identifiable. Image are stored in a secure server and their use is subject to strict terms to ensure the preservation of anonymity unless explicit permission is given by you to use your images in scientific publications.

Will my participation be confidential?

Yes, your participation will be confidential. Data will be anonymised. All participants will be assigned a code prior to data collection and all data will link to this code. No identifiable personal information will be recorded but due to the nature of the data, the volunteer may be identifiable from their facial images through recorded videos. All data will be stored in a secure server with access restricted to authorised users only.

You will have the option to consent at three levels: (1) to provide your data, (2) to allow use of your face images in academic presentations or publications, and (3) to allow inclusion of your images in any database that may be released for academic research purposes in the future.

What will happen if I want to withdraw from the research?

Your participation is voluntary. You may withdraw your consent (at each of the three levels) up to 3 months after the data collection session. There is no penalty for so doing.

Where can I get more information?

Nawal Alsufyani

School of Engineering and Digital Arts

University of Kent

Canterbury, Kent

CT2 7TN

Email: na381@kent.ac.uk

.

Appendix B

Consent Form

Title of project: Face Presentation attack Detection for Mobile Devices

Name of Researcher: Nawal Alsufyani

Participant Identification Number for this project:

Please initial the box(es) if you agree with the statement(s):

Please initial box

<u>PARTICIPATION IN THE DATA COLLECTION EXERCISE</u>

1. I confirm I have read and understood the information sheet (Version1) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. □

2. I understand that my participation is voluntary and that I am free to withdraw at any time up to 3 months after the end of data collection without giving any reason. Contact Nawal Alsufyani at : na381@kent.ac.uk □

3. I confirm that I am healthy and I do not have any known disability. □

4. I understand that my responses will be anonymised before analysis. I give permission for members of the research team to have access to my anonymised responses. □

5. I agree to take part in the above research project. □

Please tick the appropriate box(es):

CONSENT FOR USE OF IMAGES IN PUBLICATIONS/PRESENTATIONS

	Agree	**Disagree**
I hereby consent to the use of my facial images, or the measures derived from these images in academic publications or presentations. I understand that this may mean that a picture of my face may appear on screen or in print, and may appear online if the paper or presentation is uploaded to a web-based repository.	□	□
I understand that if my facial image, or the measures derived from these images are used in this way, **NO** other information which may link my facial image directly to my personal confidential information will be attached.	□	□
I understand that I may change my mind at any time up to 3 months after the end of data collection without my legal rights being affected.	□	□

CONSENT FOR INCLUSION OF IMAGES IN A POTENTIAL DATABASE FOR ACADEMIC PURPOSES

	Agree	**Disagree**
I understand that the database may be prepared for release to the academic community for research purposes. I hereby consent to the inclusion of my face images, or the measures derived from these images in this data release.	□	□
I understand that the release of the database would take place only under strict conditions controlled by a licence for a limited period (renewable on request).	□	□
I understand that I may change my mind at any time up to 3 months after the end of data collection without my legal rights being affected.	□	□

Name of participant Date Signature

Name of person taking consent Date Signature

(if different from lead researcher)

To be signed and dated in presence of the participant

Lead researcher Date Signature

Copies:

When completed: 1 for participant; 1 for researcher site file; 1 (original) to be kept in main file.

Appendix C

Participant Registration Form

Project Title: Face Presentation attack Detection for Mobile Devices

Participant Registration Form

Participant Identification Number for this project:…………………..

Please fill in the form below

Full Name: ……………………………………………………

Email address: ………………………………………………

Gender: Male □ Female □

Age band: 18-25 □ 26-35 □ 36-45 □ 46-55 □ 55+ □

Glasses/Contact Lenses: Yes □ No □

Handedness: Right hand □ left hand □

.

Appendix D

Exit Questionnaire

Project Title: Face Presentation attack Detection for Mobile Devices

Q1: How did you find following the visual stimulus on the screen?

Genuine attempts:	□ Very hard	□ Hard	□ OK	□ Easy	□ Very easy
Photo Projection:	□ Very hard	□ Hard	□ OK	□ Easy	□ Very easy
Paper Mask:	□ Very hard	□ Hard	□ OK	□ Easy	□ Very easy
3D Mask:	□ Very hard	□ Hard	□ OK	□ Easy	□ Very easy

Comments: ---

Q2: Would you be willing to use such a system for:

Unlocking your phone:	Yes □	No □	Not sure □
Online shopping:	Yes □	No □	Not sure □
Online banking:	Yes □	No □	Not sure □
Unattended passport control:	Yes □	No □	Not sure □

Comments: --

Q3: What would be acceptable maximum duration of the challenge for you to use this system for:

Unlocking your phone:	< 3 Seconds □	3 to 5 Seconds □	> 5 Seconds □	Other ----
Online shopping:	< 3 Seconds □	3 to 5 Seconds □	> 5 Seconds □	Other ----
Online banking:	< 3 Seconds □	3 to 5 Seconds □	> 5 Seconds □	Other ----
Unattended Passport Control:	< 3 Seconds □	3 to 5 Seconds □	> 5 Seconds □	Other ----

Comments: ---

Appendix E

Gaze-based Features for Detection of Presentation Attack

In this appendix, detailed results are provided for each of the gaze-based features considered in this work.

AE.1 Experimental Evaluation of Gaze Correlation Analysis

Tables AE.1 summarises overall performances (using the newly standardized APCER and BPCER metrics as defined in the ISO/IEC CD 30107-3 standard) for three attack scenarios—displayed photos, 2D masks, and 3D masks—for the **Tablet** format, using gaze correlation features tested on the entire KGDD database (80 participants) during **Lines**, **Curves,** and **Points** challenges of five-second duration.

Experiments similar to those with **Phone** format devices were performed to explore the same feature in a much smaller screen size (the **Phone** format). Table AE.3 specifies the overall performances (in ISO metrics) for the entire dataset, with 80 participants, during **Lines**, **Curves,** and **Points** challenges of five-second duration.

Table AE.1: Performance of the system at APCER = 0.05 for various challenge types using gaze correlation analysis and **Tablet format**.

Attack type	**Lines**		**Curves**		**Points**	
	APCER = 0.05*		**APCER = 0.05***		**APCER = 0.05***	
	BPCER	**ACER**	**BPCER**	**ACER**	**BPCER**	**ACER**
Photo	0.17	0.11	0.09	0.07	.016	0.11
2D mask	0.22	0.13	0.20	0.12	0.22	0.14
3D mask	0.30	0.17	0.25	0.15	0.32	0.19
Overall	0.30	0.17	0.25	0.15	0.32	0.19

*Using all 80 participants in the KGDD (48 for training and 32 for testing)

Table AE.2: Performance of the system at APCER = 0.05 for various challenge types using gaze correlation analysis and **Tablet format** with **tinted glasses**.

Attack type	Lines		Curves		Points	
	APCER = 0.05*		APCER = 0.05*		APCER = 0.05*	
	BPCER	ACER	BPCER	ACER	BPCER	ACER
Photo	0.68	0.36	0.65	0.35	0.61	0.33
2D mask	0.60	0.32	0.72	0.38	0.57	0.31
3D mask	0.73	0.39	0.71	0.38	0.68	0.36
Overall	0.73	0.39	0.72	0.38	0.68	0.36

*Using all 80 participants in the KGDD (48 for training and 32 for testing)

Table AE.3: Performance of the system at APCER = 0.05 for various challenge types using gaze correlation analysis and **Phone format**.

Attack type	Lines		Curves		Points	
	APCER = 0.05*		APCER = 0.05*		APCER = 0.05*	
	BPCER	ACER	BPCER	ACER	BPCER	ACER
Photo	0.28	0.16	0.20	0.12	0.23	0.14
2D mask	0.35	0.20	0.33	0.19	0.29	0.17
3D mask	0.40	0.22	0.43	0.24	0.46	0.25
Overall	0.40	0.22	0.43	0.24	0.46	0.25

*Using all 80 participants in the KGDD (48 for training and 32 for testing)

Table AE.4: Performance of the system at APCER = 0.05 for various challenge types using gaze correlation analysis and **Phone format** with **tinted glasses**.

Attack type	Lines		Curves		Points	
	APCER = 0.05*		APCER = 0.05*		APCER = 0.05*	
	BPCER	ACER	BPCER	ACER	BPCER	ACER
Photo	0.60	0.32	0.72	0.38	0.45	0.25
2D mask	0.69	0.37	0.74	0.39	0.49	0.27
3D mask	0.73	0.39	0.78	0.41	0.65	0.35
Overall	0.73	0.39	0.78	0.41	0.65	0.35

*Using all 80 participants in the KGDD (48 for training and 32 for testing)

AE.2 Experimental Evaluation of Gaze Alignment Consistency

Table AE.5: Performance of the system at APCER = 0.05 for various line segments using gaze alignment consistency and **Tablet format**.

Attack type	**5 Line Segments**		**7 Line Segments**		**10 Line Segments**	
	APCER = 0.05*		**APCER = 0.05***		**APCER = 0.05***	
	BPCER	**ACER**	**BPCER**	**ACER**	**BPCER**	**ACER**
Photo	0.08	0.06	0.06	0.06	**0.05**	**0.05**
2D mask	0.16	0.10	0.11	0.08	0.07	0.06
3D mask	0.21	0.13	0.15	0.10	0.10	0.08
Overall	0.21	0.13	0.15	0.10	**0.10**	**0.08**

*Using all 80 participants in the KGDD (48 for training and 32 for testing)

Table AE.6: Performance of the system at APCER = 0.05 for various line segments using gaze alignment consistency and **Tablet format** with **tinted glasses**.

Attack type	**5 Line Segments**		**7 Line Segments**		**10 Line Segments**	
	APCER = 0.05*		**APCER = 0.05***		**APCER = 0.05***	
	BPCER	**ACER**	**BPCER**	**ACER**	**BPCER**	**ACER**
Photo	0.66	0.36	0.54	0.29	0.61	0.35
2D mask	0.77	0.41	0.73	0.39	0.70	0.37
3D mask	0.72	0.39	0.58	0.31	0.71	0.38
Overall	0.77	0.41	0.73	0.39	0.71	0.38

*Using all 80 participants in the KGDD (48 for training and 32 for testing)

Table AE.7: Performance of the system at APCER = 0.05 for various line segments using gaze alignment consistency and **Phone format**.

Attack type	**5 Line Segments**		**7 Line Segments**		**10 Line Segments**	
	APCER = 0.05*		**APCER = 0.05***		**APCER = 0.05***	
	BPCER	**ACER**	**BPCER**	**ACER**	**BPCER**	**ACER**
Photo	0.18	0.11	0.15	0.10	0.07	0.06
2D mask	0.30	0.17	0.31	0.18	0.16	0.10
3D mask	0.33	0.19	0.26	0.15	0.18	0.12
Overall	0.33	0.19	0.31	0.18	0.18	0.12

*Using all 80 participants in the KGDD (48 for training and 32 for testing)

Table AE.8: Performance of the system at APCER = 0.05 for various line segments using gaze alignment consistency and **Phone format** with **tinted glasses**.

Attack type	5 Line Segments		7 Line Segments		10 Line Segments	
	APCER = 0.05*		APCER = 0.05*		APCER = 0.05*	
	BPCER	ACER	BPCER	ACER	BPCER	ACER
Photo	0.66	0.36	0.54	0.29	0.60	0.32
2D mask	0.77	0.41	0.73	0.39	0.65	0.35
3D mask	0.72	0.39	0.58	0.31	0.71	0.38
Overall	0.77	0.41	0.73	0.39	0.71	0.38

*Using all 80 participants in the KGDD (48 for training and 32 for testing)

AE.3 Experimental Evaluation of Gaze Time Dynamics Features

Table AE.9: Performance of the system at APCER = 0.05 for various challenge types using gaze time dynamics features and **Tablet format**.

Attack type	Lines challenge		Curves challenge	
	APCER = 0.05*		APCER = 0.05*	
	BPCER	ACER	BPCER	ACER
Photo	0.19	0.12	0.21	0.13
2D mask	0.28	0.16	0.33	0.19
3D mask	0.42	0.23	0.46	0.25
Overall	0.42	0.23	0.46	0.25

*Using all 80 participants in the KGDD (48 for training and 32 for testing)

Table AE.10: Performance of the system at APCER = 0.05 for various challenge types using gaze time dynamics features and **Tablet format** with **tinted glasses**.

Attack type	Lines challenge		Curves challenge	
	APCER = 0.05*		APCER = 0.05*	
	BPCER	ACER	BPCER	ACER
Photo	0.58	0.31	0.87	0.31
2D mask	0.56	0.30	0.62	0.34
3D mask	0.71	0.38	0.75	0.40
Overall	0.71	0.38	0.75	0.40

*Using all 80 participants in the KGDD (48 for training and 32 for testing)

Table AE.11: Performance of the system at APCER = 0.05 for various challenge types using gaze time dynamics features and **Phone format**.

Attack type	**Lines challenge**		**Curves challenge**	
	APCER = 0.05*		**APCER = 0.05***	
	BPCER	**ACER**	**BPCER**	**ACER**
Photo	0.29	0.17	0.24	0.15
2D mask	0.45	0.25	0.34	0.19
3D mask	0.47	0.26	0.38	0.21
Overall	0.47	0.26	0.38	0.21

*Using all 80 participants in the KGDD (48 for training and 32 for testing)

Table AE.12: Performance of the system at APCER = 0.05 for various challenge types using gaze time dynamics features and **Phone format** with **tinted glasses**.

Attack type	**Lines challenge**		**Curves challenge**	
	APCER = 0.05*		**APCER = 0.05***	
	BPCER	**ACER**	**BPCER**	**ACER**
Photo	0.59	0.32	0.85	0.45
2D mask	0.57	0.30	0.62	0.34
3D mask	0.69	0.36	0.75	0.40
Overall	0.96	0.36	0.85	0.45

*Using all 80 participants in the KGDD (48 for training and 32 for testing)