Biometric Liveness Detection Using Gaze Information

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

Asad Ali

March 2015

Abstract

This thesis is concerned with liveness detection for biometric systems and in particular for face recognition systems. Biometric systems are well studied and have the potential to provide satisfactory solutions for a variety of applications. However, presentation attacks (spoofing), where an attempt is made at subverting them system by making a deliberate presentation at the sensor is a serious challenge to their use in unattended applications. Liveness detection techniques can help with protecting biometric systems from attacks made through the presentation of artefacts and recordings at the sensor. In this work novel techniques for liveness detection are presented using gaze information.

The notion of natural gaze stability is introduced and used to develop a number of novel features that rely on directing the gaze of the user and establishing its behaviour. These features are then used to develop systems for detecting spoofing attempts. The attack scenarios considered in this work include the use of hand held photos and photo masks as well as video reply to subvert the system. The proposed features and systems based on them were evaluated extensively using data captured from genuine and fake attempts.

The results of the evaluations indicate that gaze-based features can be used to discriminate between genuine and imposter. Combining features through feature selection and score fusion substantially improved the performance of the proposed features.

Acknowledgements

I would like to thank my supervisors Dr. Farzin Deravi and Dr. Sanaul Hoque for their guidance and support. I thank them especially for giving me the freedom to explore new research ideas. My conversations with them have been a source of great encouragement, inspiration and learning. Their words of wisdom and advice for my work has always been invaluable and has been instrumental to finishing my PhD. I would also like to thank the school of Engineering and Digital Arts and the University of Kent for providing me the financial support.

I would also like to thank Mr. Harvey Twyman and Mr. Clive Birch of the technical team at the School of Engineering and Digital Arts for their timely help whenever.

Special thanks go to Mr. Richard Douglas and Helen Winder, who patiently made every effort to support me in every way they could. Without their encouraging words and support, finishing the PhD would not have been possible.

Finally, I would like to thank my family in Pakistan and in the UK. I would not have finished this PhD without the support, love and encouragement from my elder brother. Their wisdom and insight has been a great source of strength.

Contents

List of Figures

CHAPTER 1

Introduction

This thesis is concerned with liveness detection for biometric systems. Biometric systems have the potential to provide security for a variety of applications. Biometric systems are vulnerable to certain attacks. Security countermeasure is, therefore, required to be incorporated to protect a biometric system from attacks. Examples of security measures include liveness detection which can detect fake biometric samples. Liveness detection is a challenging area and requires an in-depth understanding of the subtle differences between genuine and fake attempts, and the exploitation of that information to prevent such impostor attacks [1]. This thesis suggests a challenge/response gaze-based novel features scheme which can overcome sophisticated impostor attacks in face recognition systems. Unlike existing work described in the literature, this study does not focus on a particular type of attack but aims to deal with a collection of attack modes (photo, photo mask and video).

In this thesis terms such as impostor attack, presentation attack, fake attempt, spoofing attack are used interchangeably. In the context of liveness detection

[1] Any person who, intentionally or otherwise poses as an authorised user is named an impostor [18]

these terms are used when artifacts are presented to the biometric system at sensor level to subvert its normal operation.

In the rest of this chapter the motivation for this research is further expanded upon. Section 1.1 will present a brief introduction to biometrics systems along with potential vulnerability related to biometric systems. Section 1.2 will present the motivation of this research along with system block diagram. Aims and objectives are listed in Section 1.3. A scope for the work will be explored in Section 1.4. Finally Section 1.5 will present the structure of the thesis.

1.1 Biometric systems

Security systems which require high accuracy are becoming more important than ever in our technologically dependent world. In the modern interconnected society we live in, to be able to reliably recognise a person at the remote end of a computer network is becoming critical. Many questions can be raised, for example,"Is she/he really who she/he claims to be?" or "Is this person at the other end of the network authorized to use this facility?", and so on.

The search for techniques that can improve the performance of an automatic person recognition system is very important [19]. Traditionally a person is automatically recognised based on "what s/he remembers", for example, passwords, PIN, etc. Similarly such recognition can also be based on tokens possessed by someone such as ID cards and keys, etc. All these methods have been used to control access to premises or systems, etc. [20]. However, if the ID card is stolen or a password is known to an unauthorised user, security can be breached, especially if the system is unattended. Recognition based on what a person is or does can address the problems related to these traditional methods [21]. Technologies

for person recognition based on their physiological or behavioural characteristics is known as biometrics [22].

The biometrics-based systems can have their functioning based on various parts of the human body or human behaviour. For example, voice [23–27]; signature [28–31]; gaze [32–36]; gait [37–40] etc. are usually classified as behavioural biometrics and iris [41–45]; hand geometry [46–50]; fingerprint [51–55]; face [56–61]; etc. are examples of physiological biometric, all of which are used in the real world for security.

However all modalities typically aim to fulfil the following criteria [22]:

- Universality: It is present in every person.
- Uniqueness: Two persons should be sufficiently different in terms of the characteristic.
- Permanence: It must not change with time.
- Measurability: It has to be possible to measure it.

Biometrics-based methods have several advantages over the traditional security methods, such as PIN codes, passwords, keys, cards, IDs, tokens, etc. For example, the old classical security methods require the user to remember PIN codes or long passwords that could easily be forgotten, or to carry cumbersome bunches of keys, tokens or cards that can be easily lost or stolen [62–65]. Biometrics, however, guarantees that the user who accesses certain facilities cannot deny using it (non-repudiation) [24] and does not require the possession of any physical tokens, nor rely on uncertainty of the human memory. Biometric information has been widely used with a satisfactory performance in criminal investigation, access control, etc. [66].

Despite these advantages, biometric systems do have some disadvantages, and can be vulnerable to external attacks which can compromise security [65]. Ratha et al. [67] highlighted several possible attack points to biometric recognition systems as depicted in Figure 1.1. These can be grouped into two main categories: direct attacks also called presentation attacks, where the impostor attacks the biometric system at the sensor by presenting synthetic biometric samples, e.g. gummy fingers [68]. Matsumoto et al. [68] showed how easily gummy fingers can be made with a common material like gelatine which can be used to spoof various fingerprint devices with optical or capacitive sensors. Ruiz-Albacete at el. [69] reported the vulnerabilities of iris-based recognition systems. They printed high quality photo of the iris to present to the iris recognition system.

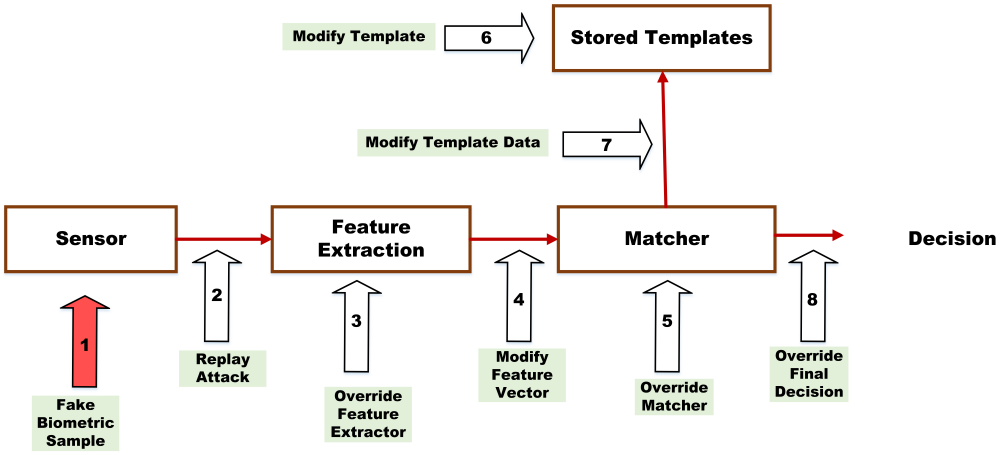

Figure 1.1 Possible points where biometrics system can be attacked. (The focus of the thesis is shown by the red arrow - presentation attacks)

The remaining points of attack shown in Figure 1.1 can be considered indirect attacks. This study only deals with presentation attacks. This is a type of attack where the the impostor does not need any prior knowledge about the underlying working principles of the system.

1.2 Motivation

Humans most commonly use facial appearance to recognise a person, therefore, it is the natural choice as a modality in biometric technology [70, 71]. As the face is normally visible, it is easy to capture the facial image of a person with or without the cooperation of the individual [72]. A considerable amount of work has been reported in the literature on this modality [56–59]. Among all the biometric systems, face recognition is especially convenient in areas where immediate, correct recognition of individuals at unattended access control points, such as entrances to buildings, or security at border crossings, security in the street, and where being able to uniquely recognize individual humans without user cooperation is a vital aspect of achieving effective security.

Many companies have already implemented face recognition systems. These may be accurate and fast, but they can also be susceptible to various threats such as presentation attacks. Most facial recognition systems process facial images for identification without checking whether the sample is captured from the authorised user or from a photo or video of the authorised user. Therefore, an impostor can use a high quality image or a video of authorised users in order to gain unauthorized access to premises, systems or data. A reliable facial recognition system should, therefore, prevent impostors from gaining unauthorized access to unattended places or data. Hence the suggestion that existing facial recognition systems require an effective liveness detection function in order to avert impostor attacks [73, 74].

The algorithms proposed in this study are based on the assumption that the spatial and temporal coordination of the movements of eye, head and hand involved in the task of following a visual stimulus are significantly different when a genuine attempt is made compared with certain types of spoof attempts. The

task requires head/eye fixations on a simple shape that appears on a screen in front of the user, and in the case of a photo spoofing attack, visually guided hand movements are also required to orientate the photographic artifact to point in the correct direction towards the challenge item on the screen.

It is likely that the head pose and direction of gaze will be different when photo spoofing is attempted as coordination may be maintained by delaying the hand movements until the eye is available for guiding the movement [75]. The introduction of hand movements is also likely to change the relationship between head and eye movements, as the coordination of the eye and head in gaze changes is usually a consequence of synergistic linkage rather than an obligatory one [75–77]. Therefore, it is assumed that accurately directing the photograph to a particular orientation indicated by the visual stimulus on the screen is likely to be less repeatable than merely looking at the stimulus. Hence, the variance in measured gaze parameters is used to distinguish genuine from fake attempts as described in the rest of the research. So the features which will be investigated in this research will be based on the gaze of the human. By gaze we mean head/eye movements.

Many algorithms have been proposed in the literature [78–80] to address the difficult problem of face liveness detection. A variety of techniques have been proposed for liveness detection including detecting eye blinks, sensing response to stimuli, and detection of facial gestures. Various approaches and constraints are used to enhance the reliability of these systems. However, finding good novel features which can detect all types of attack scenarios is a challenging task.

The general architecture of the face liveness detection system that will be followed in this thesis is shown in Figure 1.2. Novel features from facial landmarks will be extracted from the images which are then analysed to determine whether the captured images are acquired from a genuine source or not.

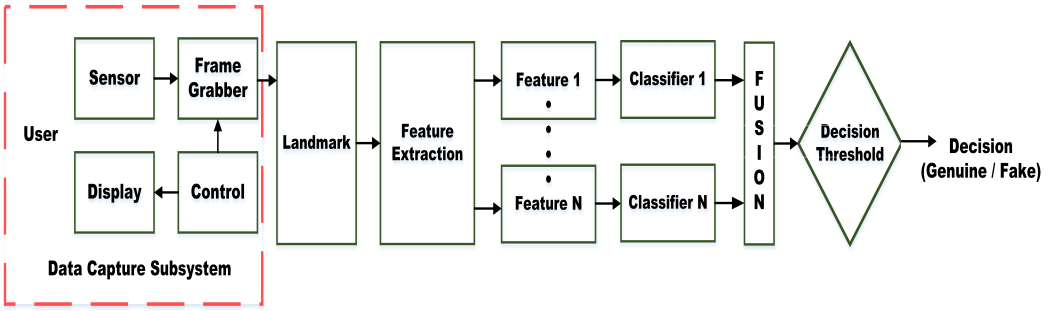

Figure 1.2 System structure flow diagram

Most smart phones have built-in cameras and have the option to use face recognition for logging into the phone [81] instead of using a password. The proposed liveness detection system can be added to such devices to enhance the security of the existing face recognition system.

People normally use a password to log into a PC, laptop or note book. This can be replaced by face recognition together with a liveness detection mechanism to avoid entering a user name and password each time a system is used. In fact this can be used to lock the system automatically as well if the authorised user is not sitting in front of the PC. This will not only make it easier for the user, but it will also enhance the system security as one can steal a client's password.

There is another, perhaps more important aspect to the face liveness detection method that can be implemented e.g. when the user is making an online transaction or withdrawing money from an ATM machine. The proposed system can be embedded to the existing process of the online payment procedure, where the system will ensure that the live face of the authorised user is available at the time of completing the transaction to avoid unauthorised payment transaction and unauthorised withdrawal using a card cloned from the card belonging to the authorised user.

There are many potential applications of the proposed system and here we have highlighted only a few important ones which demonstrate the potential impact

of the proposed approach.

Note that the proposed liveness detection approach may be combined with other biometric modalities too - notably iris recognition.

1.3 Aims and Objectives

The general aim of this research is to develop robust and efficient liveness detection algorithms to enhance the trust in and the security of biometric recognition systems. This work will explore the effectiveness of such systems to counteract impostor attacks for a number of attack scenario schemes. The specific objectives of the research are to review the state of the art of biometric liveness detection methods and to propose a liveness detection framework that can deal with a multitude of presentation attack types. This study will also aim to collect the appropriate biometric databases and propose an evaluation framework to facilitate investigative analysis of presentation attacks. The research objectives include the exploration of gaze-based features to achieve liveness detection. This work will also explore the use of fusion techniques to improve the efficiency of the scheme and to optimise the proposed algorithms and carry out comparative analysis.

1.4 Scope of the Project

The list of the work which will be carried in this study is summarized below an explains the areas which will be covered in this research. The areas which will not be covered in this research are also listed. This study will only explore facial liveness detection methods. The research will deal with three types of

presentation attacks, e.g photo, mask and video replay. In this study 2-D photo mask attack detection will be explored. This work will not explore 3-D mask attack detection. One would need several photos of the target and a 3-D printer to produce the 3-D mask which may not be available promptly. The research will investigate features based on gaze stability. Though liveness detection techniques are normally used in conjunction with biometric person recognition systems, this research does not address novel techniques for biometric recognition in general and face recognition systems in particular. However, the interaction between biometric systems and the liveness detection function is of relevance to the work presented here.

1.5 Structure of the Thesis

The organization of the thesis is given below.

In Chapter 2, the background of various key concepts for understanding the related previous work on face liveness detection is presented. A detailed comparative study is presented that focusses on various liveness detection methods that have been proposed in recent years. The related previous work is grouped into two main categories. This is further divided into several groups based on the type of feature used for liveness detection.

Chapter 3 provides the detail of the database that was collected for training and testing purposes. It also provides further details on the evaluation strategy used for this work as well as the hardware and software used to conduct the experiments.

Chapter 4 introduces the gaze colocation feature. The use of ROC curves to analyze and assess the performance is proposed and demonstrated.

Another novel feature, the gaze collinearity, is presented in Chapter 5 and is aimed at improving the performance of the proposed face liveness detection system. This chapter also explores combining collinearity and colocation to produce more effective measures for liveness detection.

Chapter 6 presents another novel gazed-based homography feature, to further enhance the accuracy of the proposed face liveness detection system. This chapter also explores combining collinearity, colocation and homography to produce more effective measures for liveness detection.

Conclusions, a summary of the contributions of this work and suggestions for future work are provided in Chapter 7.

The goal of this thesis has been to perform an extensive experimental study of various novel features, classification and combination rules applied to the problem of face liveness detection for biometric systems.

CHAPTER 2

Literature Review

2.1 Introduction

The biometric technology involving face recognition has developed rapidly in recent years as it is user friendly and convenient, and is used for many security purposes, but is vulnerable to abuse, such as spoofing photographic or video substitution and many others as discussed in Chapter 1. However, by adding liveness detection the effectiveness of security systems can be substantially improved. The differences between a photograph or video of an individual and the real person can be used to establish liveness.

Various approaches have been presented in the literature to establish liveness for detecting presentation attacks. Liveness detection approaches can be grouped into two broad categories: active and passive. Active approaches require user engagement to enable the facial recognition system to establish the liveness of the source through the sample captured at the sensor. Passive approaches do not require user co-operation or even user awareness but exploit involuntary physical movements, such as spontaneous eye blinks, and 3D properties of the image.

Challenge response is a type of intrusive approach, the user is asked to perform specific activities to ascertain the liveness. Uttering digits, changing head pose are the examples of the challenge response. Passive anti-spoofing techniques are usually based on the detection of signs of life, e.g. eye blink, facial expression, etc. Here the face liveness detection methods are grouped based on the feature and the methods that were used to estimate the liveness.

2.2 Literature

Several approaches which are implemented to solve the the problem of face liveness detection for the face recognition system. In this section, these approaches are grouped and explored based on the nature of the feature. Following methods are proposed in literature for liveness detection.

2.2.1 Eyeblink Based Liveness detection

Blinking is a natural biological function of the closing and opening of the eyelid. The blink helps spread fluid from the tear ducts across the eye and removes irritants from the surface of the cornea and conjunctiva [82]. Blinking can vary with fatigue, emotional stress, amount of sleep, eye injury, medication, and disease [83]. It has been reported [84, 85] that the blink rate of a human is between 15 to 30 times per minute. The average blink lasts for about 250 milliseconds [86].

Blinking has been used as a means of human interaction with computer [87, 89]. Various researchers have used blink detection for face liveness detection.

One method to detect blinks, is to classify each image in the video sequence independently as one state (closed eye or opened eye), for example, using the

Viola-Jones cascaded Adaboost approach to detect the face and eye [90]. AdaBoost is a learning algorithm that selects a small set of weak classifiers from the large number of potential features. This method assumes that all of the images in the temporal sequence are independent. In real, the neighbouring images during blinking are dependent, since the blink is a procedure of eye going from open to closed, and back to open. The temporal information, which may be very helpful for recognition, is ignored in this method. Lin Sun et al. [1] presented an eye blink detection approach for detecting face liveness using Conditional Random Fields (CRFs). Using Hidden Markov Model (HMM) [91] which can not accommodate long-range dependencies on the observation, they used conditional random fields(CRFs) which are probabilistic models for segmenting and labeling sequence data and mainly used in natural language processing for its ability to accommodate long-range dependencies on the observed sequence [92–94]. Lin Sun et al. employed a linear chain of CRFs in their method [1].

Lin Sun et al. [1] demonstrated that the blinking consisted of two continuous sub-actions, from open to closed and from closed to open. These (open to closed and from closed to open) activities could be sampled into an image sequence with eye in various states. The state of the eye in images are classified as open, half open and closed. Each state of the blink should not be considered independently for blinking recognition. They stressed the need to model blinking on the contextual dependencies in an eye blinking sequence. At particular points of time, it is hard to predict blinking activity using the previous state and the current observation.

Lin Sun et al. [1] used symbols C for closed state and NC for non-closed (including open and half-open), to label eye states. The graphical structure of their CRF-based blinking model is shown in Figure 2.1.

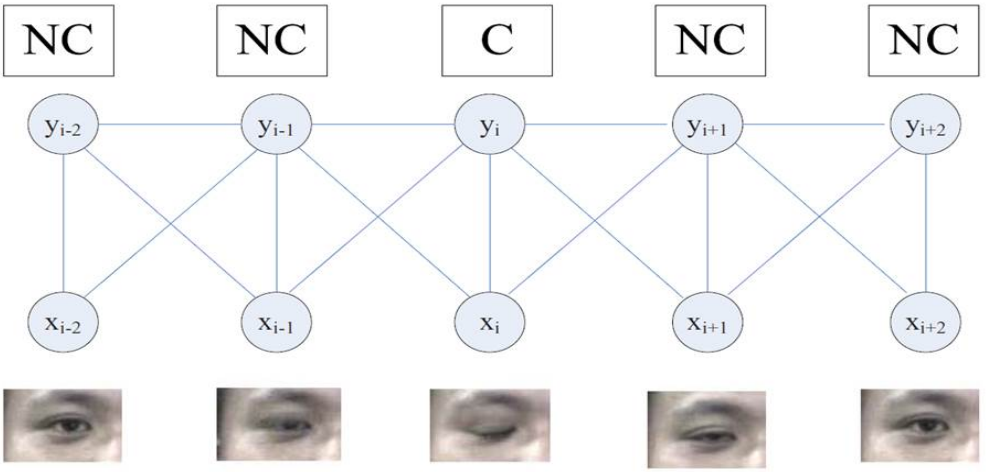

Figure 2.1 Graphic structure of CRF-based blinking model. C and NC are for closed state and non-closed state respectively [1]

They collected video database from genuine users using a webcam. There were 20 participants and 4 video clips were recorded of each creating 80 video clips of about 5 seconds length. The number of blinking varies from 1 to 6 times in each video. They also collected impostor database which contain 180 impostor video capture using photo. The authors reported 98.3% imposter detection rate.

Pan et al. in [2] further enhanced the Lin Sun et al. work [1]. The main eye states modeled are opening and closing. In addition, there is an ambiguous state when blinking from open state to closed or from closed state to open. In this method they extract the temporal information from the process of the eye blink, namely the consecutive stages of open, half closed and closed, followed by half open and fully open all of which are sequential eye blink movements and constitute a complete eye blink pattern which was used to determine liveness.

In this work the authors defined a three-state set for eyes, α : open, γ : closed, β : ambiguous and a typical blink activity was described as a state change pattern

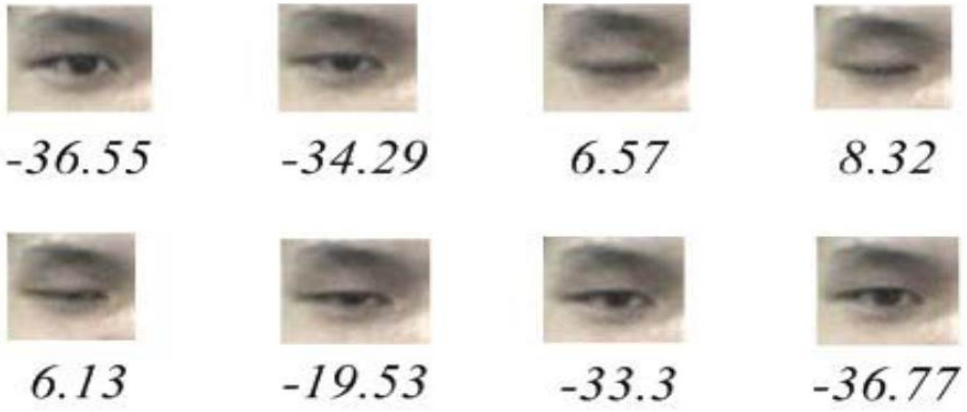

Figure 2.2 Illustration of the blinking activity sequence. The value of the closeness for each frame is below the corresponding frame. The bigger the value, higher the degree of closeness [2]

of $\alpha \rightarrow \beta \rightarrow \gamma \rightarrow \beta \rightarrow \alpha$. They used the same database which is discussed in [2] for evaluating the performance of their method. They compared three methods, cascaded Adaboost, HMM and and their method against photo spoofing using the photo-imposter video database.

Pan et al. [3] further explored the method to enhance their face liveness detection system discussed [1] and [2]. In this work they fused eyeblink and scene context. The authors assumed that the face recognition system camera is fixed while the system is operating. The first frame was captured from the scene without a person in front of the camera. This frame was designated as the reference scene; an impostor video would be of a different scene.

They extracted reference points or activity in the captured frame and named them clues. Clues extracted from the face region were named inside clue, example of a face clue is eye blinks. Similarly clues extracted from the background of the captured image near a face were named outside clues (scene context clue). These

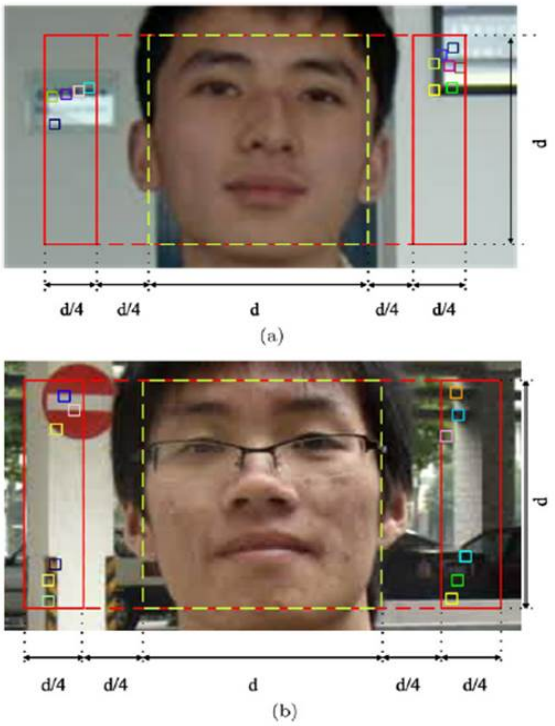

Figure 2.3 Examples of scene region of interest and fiducial points extraction. Yellow dashed line rectangles are face regions and red solid line rectangles are scene regions of interest. Fiducial points are labeled by colorful squares [3]

can be any item located in the background which can be captured and extracted for recognition.

They extract scene context clues from the right and the left parts of the detected face region as shown in Figure 2.3. The eye blink may stop the photographs and 3D models spoofing, while the scene context is used for anti-spoofing by video replay. They combined these clues of eyeblinks and scene context to improve the performance of the liveness detection system. The proposed fusion system is shown in Figure 2.4.

For training and testing purposes, they collected their own data. The database consist of 100 video clips from 20 volunteers. A high-quality photo was taken from each volunteer. The five categories of photo attacks below were simulated.

- Keep the photo still.

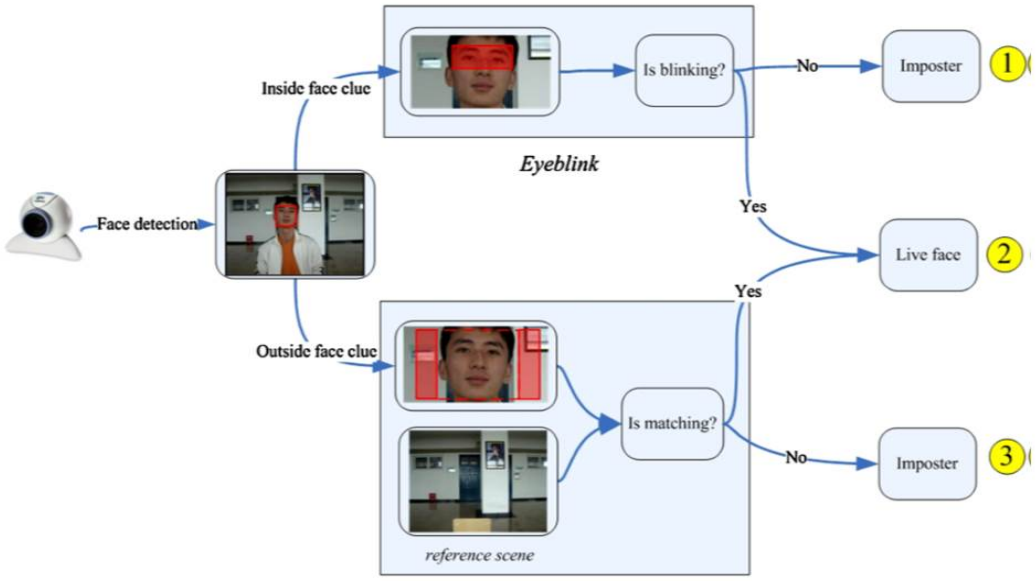

Figure 2.4 Illustration of liveness detection system using a combination of eyeblinks and scene context [3]

- Move the photo vertically, horizontally, back and front.
- Rotate the photo in depth along the vertical axis.
- Rotate the photo in plane.
- Bend the photo inward and outward along the central line.

For each attack, one video clip is captured with a length of about 10 to 15 seconds. Some samples are shown in Figure 2.5.

The live face video database contains 196 clips for 14 individuals. There were 2 indoor and 5 outdoor scenes. Each scene has a scene reference image. Each individual appears before the camera twice and stays there about 5 seconds for liveness verification. Examples of the data are shown in Figure 2.6.

Jee et al. [4] introduced a memory efficient method for face liveness detection for embedded face recognition system. The method is based on the analysis of

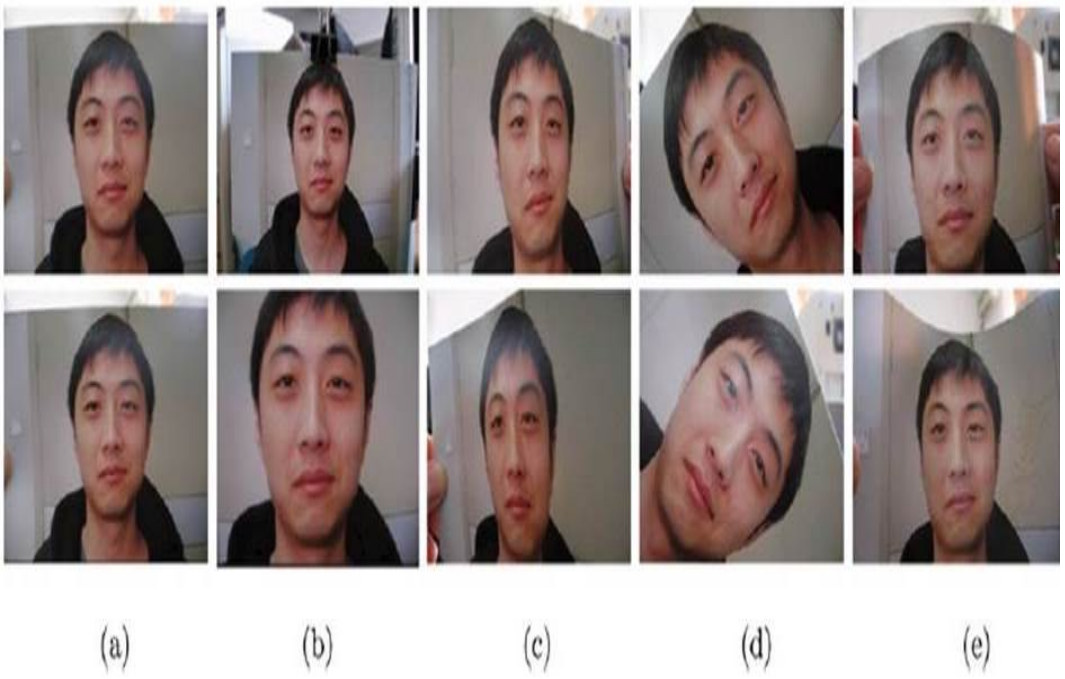

Figure 2.5 (a) Keep photo still. (b) Move vertically, horizontally, backward and forward. (c) Rotate in depth. (d) Rotate in plane. (e) Bend inward and outward [3]

Figure 2.6 The first row is four scene reference images. The second row is live faces in video [3]

the eye movement. They used Viloa-Jones [90] methods to detected the eye in the facial images.

They normalized the input face images as they can vary in size and orientation. After normalizing the face region, the eye regions were extracted. Then the eye regions were binarized in order to achieve the pixel value of 0 and 1 by using a threshold. The threshold is adaptively obtained from the mean pixel value of each eye region. Figure 2.7 shows the eye regions of genuine and fake face which change very little in case of fake attempt, and a much larger variation in shape in case of the genuine attempt because of the blink or the movement of the pupil.

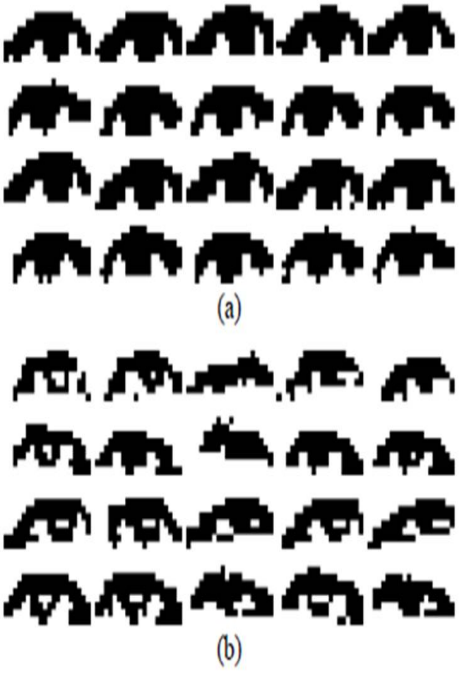

Figure 2.7 Example of binarized eye regions of (a) fake face and (b) live face [4]

The authors used the hamming distance method to calculate the liveness detection score for the eye region. They extracted 10 liveness scores of both the left and right eyes and added them, and used the average of the scores. If the average liveness score was bigger than the set threshold, the input image is classified as a genuine face.

Wang et al. [5] presented a liveness detection method in which the physiological motion was detected by estimating the eye blink and using an eye contour extraction algorithm. They used an active shape model [95] with a random forest classifier trained to recognize the local appearance around each landmark. They showed that if any motion in the face region is detected the sample is considered to be captured from an imposter.

The proposed method is composed of two parts; the attempt passed through both parts of the approach for the liveness check. The first part detects the physiological motion in three modules for eye detection, eye contour extraction, and eye blinking detection. The second part extracts the motion cues and seeks to hold the head still. If any motion in the face region was detected, the attempt was classified as fake. The flow chart of the system is shown in Figure 2.8

The eye blink estimation detects the eye blinks in the face sequences. The authors related the blinking to the degree of the eye opening estimated from the distances d_1 and d_2 in Figure 2.9. where $d_1$is the distance between the upper and lower eyelids of the left eye and d_2 is the distance between the upper and lower eyelids of the right eye. Where as d_e is the distance between the center of the left and right eyes. The eye opening is then calculated as in Equation 2.1

$$D = \frac{min(d_1, d_2)}{d_e} \tag{2.1}$$

When the opening degree is deduced from larger to small, an eye blink is detected.

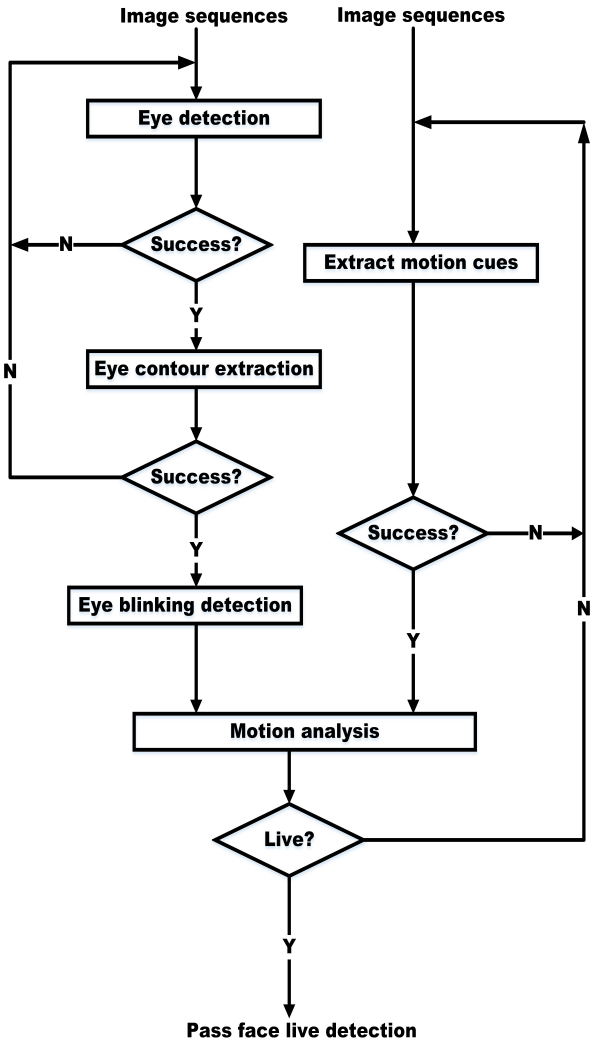

Figure 2.8 Algorithm flowchart

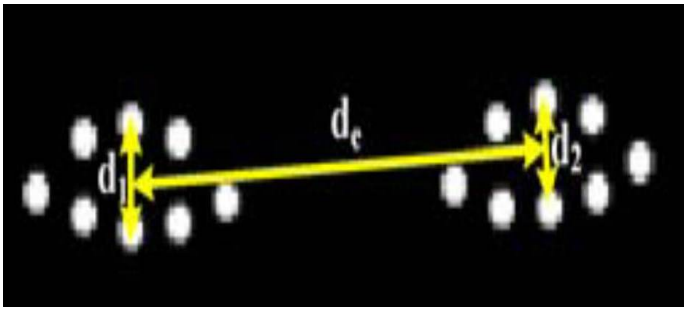

Figure 2.9 Eye opening estimation [5]

2.2.2 Face Liveness Detection using Frequency and Texture Analysis

Kim et al. [6] proposed a face liveness detection method for distinguishing 2-D paper masks from the genuine faces. They used a multi-classifier method

for detecting fake attempts by combining frequency information from the power spectrum and texture information using Local Binary Pattern (LBP) features.

3-D shapes lead to the difference in the low frequency regions which is closely related to the illuminance component induced by the overall shape of a face. The detail between the live attempts and the mask attempts trigger the disparity in the high frequency information [96] [9]. The images taken from the 2-D objects have less texture richness compared to the images taken from the 3-D objects. The texture information obtained using Local Binary Pattern (LBP) features.

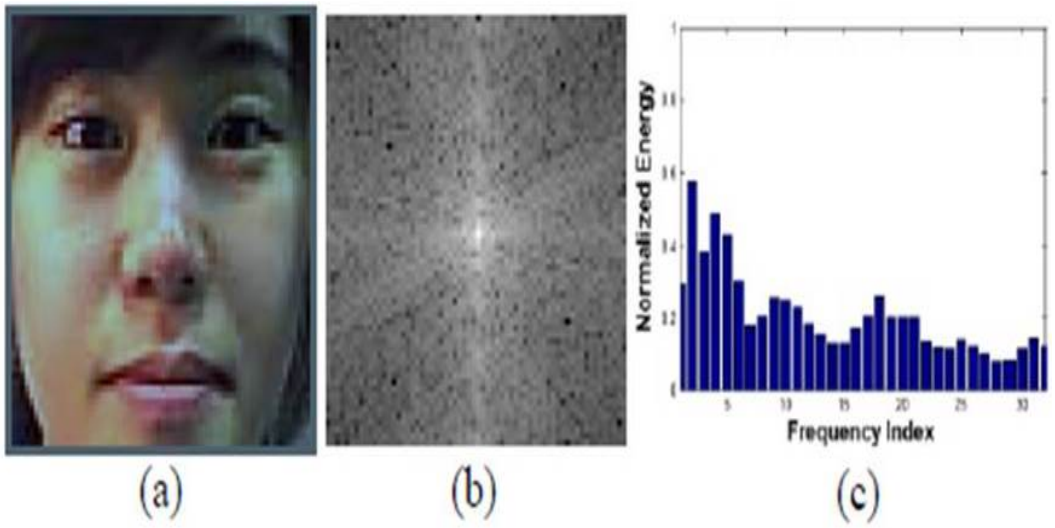

Figure 2.10 Frequency-based feature extraction (a) original facial image (b) Log-scale magnitude of the Fourier-transformed image (power spectrum) (c) 1-D frequency feature vector extracted from the normalized power spectrum [6]

Frequency information from the facial images were extracted and were transformed to the frequency domain using the 2-D discrete Fourier transform. The face image is shown in Figure 2.10(a), Figure 2.10(b) shows the Fourier-transformed image and Figure 2.10(c) shows the resulting frequency feature. Local Binary Patterns (LBP) [7] is one of the most popular methods to describe the texture information of the images. The authors used LBP to analyze the texture characteristics of the image taken from genuine and fake attempt. Figure 2.11 explain

the process of acquiring the LBP feature vector from a given facial image. Figure 2.11(a) is the original facial image while Figure 2.11(b) shows the LBP-coded image of Figure 2.11(a). Figure 2.11(c) shown histogram of the LPB-coded image, which will be exploited as the feature vector for the classification.

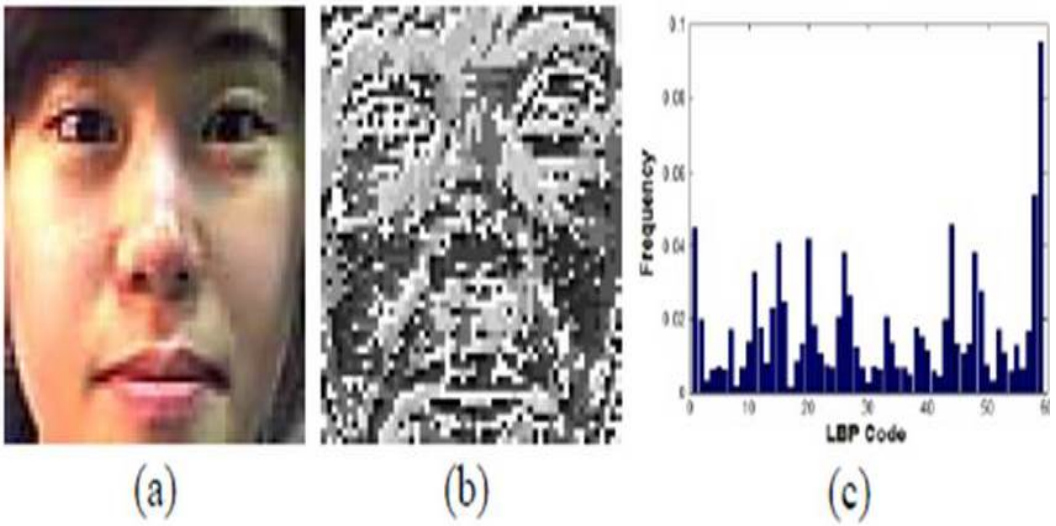

Figure 2.11 Feature vector extraction process based on LBP (a) original facial image (b) LBP-coded image (c) histogram of the LBP-coded image [7]

They extracted frequency-based feature, texture-based feature and implemented own their own and fused them together. They used Support Vector Machine (SVM) classifiers using the two types of feature vectors extracted. The decision values of these two SVM classifiers were then used as 2-D feature vectors for the subsequent fusion.

Komulainen et al. [97] explored the use of dynamic texture information for spoofing detection. They argued that masks and 3D head models are rigid, whereas genuine faces are non-rigid with contractions of facial muscles resulting in temporal deformation of facial features, such as moving eyelids, lips, etc. The structure and dynamics of the micro-textures that characterise real faces were used in their proposed approach to spoof detection. They used spatiotemporal (dynamic texture) extensions of the local binary pattern in this approach.

The authors considered local binary patterns from three orthogonal planes (LBP-TOP) which have been shown to be very effective in describing the horizontal and vertical motion patterns in addition to appearance. The original LBP operator was defined to only deal with the spatial information. It has been extended to a spatiotemporal representation for dynamic texture analysis (DT). This has resulted in the so called Volume Local Binary Pattern operator (VLBP) [98]. The idea behind VLBP consists of looking at dynamic texture as a set of volumes in the (X,Y,T) space where X and Y denote the spatial coordinates and T denotes the frame index (time). The neighbourhood of each pixel is thus defined in a three dimensional space. Then, similarly to basic LBP in spatial domain, volume textons can be defined and extracted into Face Spoofing Detection Using Dynamic Texture 149 histograms. Therefore, VLBP combines motion and appearance into a dynamic texture description.

They carried out experiments on the CASIA Face Anti-Spoofing Database [99] and Print-Attack Database [100]. The CASIA data set contains 50 real clients and the corresponding fake faces are captured with high quality from the original ones. Three imaging qualities, low, normal and high were then extracted. They used SVM classifier for training and testing the method.

Komulainen et al. [8] explored fusion of motion and texture based countermeasures under several types of face attacks. They explored the fusion potential of different visual cues and show that the performance of the individual methods can be vastly improved by performing fusion at score level.

Figure 2.12 is the diagram of the authors' proposed fusion strategy. The video sequences were divided into overlapping windows of frames. Each observation generate scores from motion and micro-texture are combined to achieve a single score using score based fusion using linear logistic regression (LLR). They carried out experiments on the Replay Attack database [101]. The database was divided

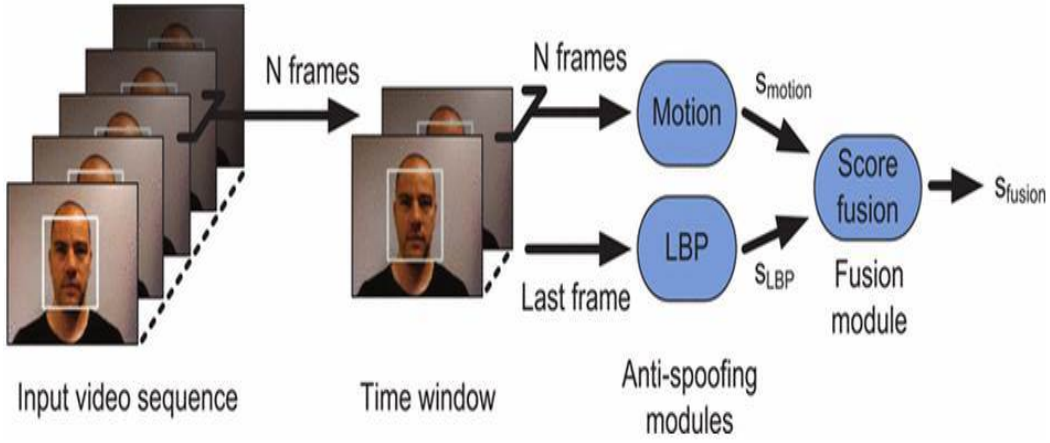

Figure 2.12 Block diagram of the used fusion strategy [8]

into three sets for training, development and testing to evaluate the method. Jiangwei Li et al. [9] explored a technique based on the analysis of 2-D Fourier spectra of the face image. They proposed that the size of a photograph is smaller than the real image and the photograph is flat, it therefore has fewer high frequency components than real face images. They further explored that if a photo is held before a camera and is in motion, since the expressions and poses of the face contained in the photo are invariant, the standard deviation of frequency components in the sequence must be very small. This can only be valid for low resolution photographs but it is possible print bigger and high resolution photographs in which the high frequency components will be closer to those of a real image.

The most common presentation attack are printed photo, video replay on lcd screen. The media of these sources are 2-D planar structure, whereas genuine face is 3-D structures. The intensity contrast of a genuine facial image is more obvious than that of a fake image and such differences lead to greatly different reflectivity of light, which is shown in the frequency distribution of an image. Figure 2.13, shows the comparison of the Fourier spectra of a genuine face image

to that of a fake face image. (a) genuine face image; (b) A fake face image; (c) 2D Fourier spectra of (a); (d) 2D Fourier spectra of (b).

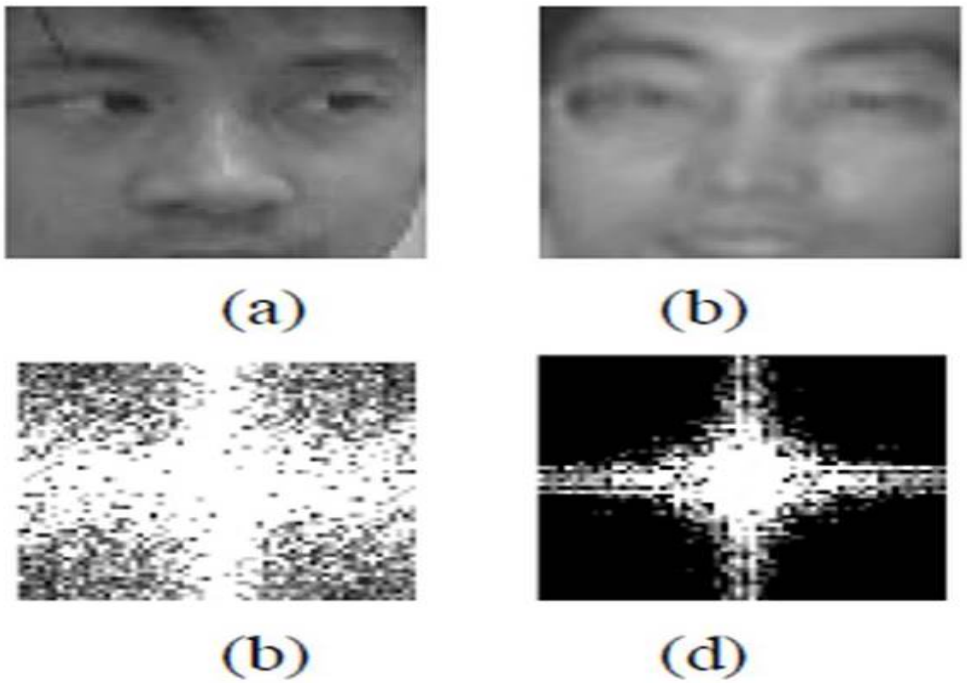

Figure 2.13 Difference between live face and fake face in frequency domain [9]

Bharadwaj et al. [102] explored the utility of Local Binary Patterns (LBP) based features along with motion magnification. The authors explored two types of feature extraction algorithms. They presented a configuration of LBP that provided better performance compared to other computationally expensive texture based approaches. Motion feature estimation is also explored.

The authors localized required motion and then magnified it under the Taylor expansion assumption. The approach enhances facial movements including subtle motion such as blinking, saccadic and conjugate eye motion that may otherwise only be visible on close inspection of the video.

Motion magnified video of a subject can be classified for spoofing detection using either texture or motion based features. As mentioned, texture features are widely explored in spoofing detection literature as compared to motion based

features. They proposed the texture and motion based features for spoofing detection.

Bharadwaj et al. [102] exploited various texture based spoofing detection approaches [101, 103–107] to explored the utility of LBP based features along with motion magnification for liveness detection system. To encode texture information at multiple scales, they proposed to use feature concatenation of the three LBP configurations.

Wu et al. [10] suggest a liveness detection scheme, combining Fourier statistics and local binary patterns. Both techniques, Fourier spectra and local binary patterns, have been investigated on their own and the authors fuse them together in order to improve the liveness detection performance. Figure 2.14 shows the their proposed system where the score from features of local binary patterns and Fourier spectra are combined using support vector machine.

They carried out experiments on the NUAA [96] database which consists of genuine and photo of impostor. Their method classified genuine and fake with 100% and 92.33% respectively.

Das et al. [108] proposed method based frequency analysis and texture analysis by using frequency descriptor and Local Binary Pattern respectively. They exploited frequency and texture based analysis to differentiate between images captured from genuine and fake attempts. Images captured from fake attempts have low frequency regions and less texture richness. However, genuine samples have high frequency information and high texture richness. They also used NUAA database for training and testing purposes.

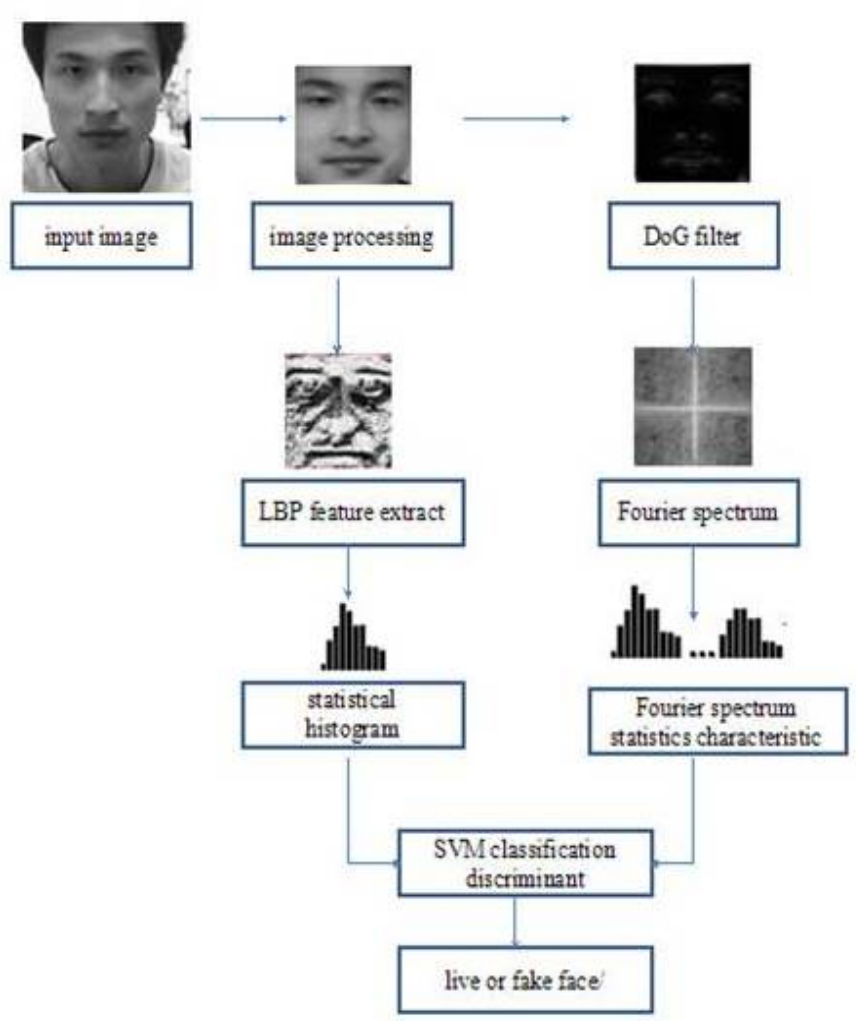

Figure 2.14 The framework of the proposed approach for liveness detection by Wu [10]

2.2.3 Challenge Response Mechanisms

Systems based on the challenge-response approach belong to the active category, where the user is asked to perform specific activities to ascertain liveness such as uttering digits or changing his or her head pose. For instance Frischholz et al. [11] investigated a challenge-response approach to enhance the security of the face recognition system. They developed a head pose estimation technique using a single camera. The users were required to look in certain directions, which were chosen by the system randomly. The system estimated the head pose and

compared the real time movement (response) to the instructions asked by the system (challenge) to verify the user authenticity. After responding to several of these challenges, the user is asked to look straight into the camera and the final image is captured for face recognition.

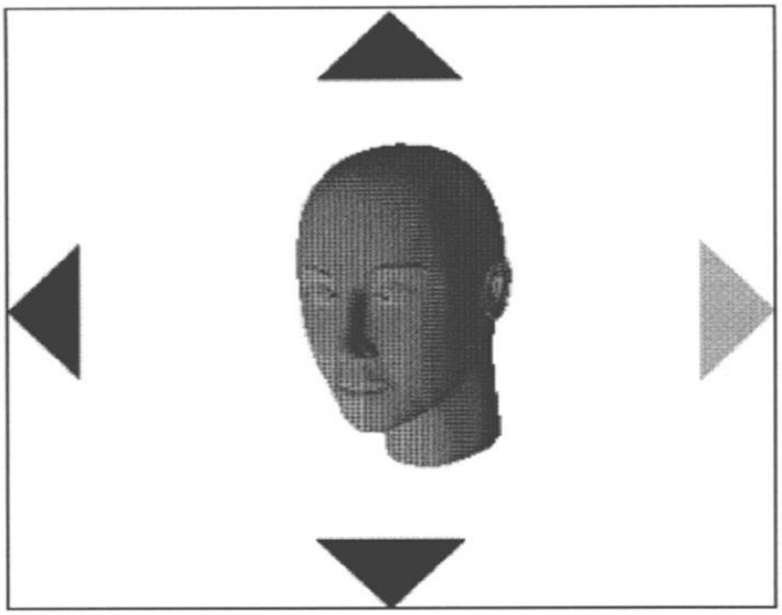

Figure 2.15 Random challenge directions [11]

Kollreider et al. [12] explored the liveness detection approach where users are required to interact with the face liveness detection system. This interaction occurred through the utterance of a specific digit sequence, either known previously by the person or prompted randomly. The authors favored the latter scheme, since a sound utterance can be recorded easily. They explored the changes in facial expressions during the utterances.

In this technique they located the mouth regions and process every frame and extract OFL in real time. They have used the XM2VTS database for evaluation. Volunteers' videos were recorded pronouncing digits (from 0 to 9). The aim was to recognize the digits of the volunteer through lip-motion only. For a digit, they have used 100 short videos. For training, there were a total of 60 videos and

for testing, a total of 40 videos. For each of the digit videos, feature vectors are extracted from mouth regions and given to SVM classifier for purpose of classification. Out of 100 individuals, recognition rate is 0.73 (73%).

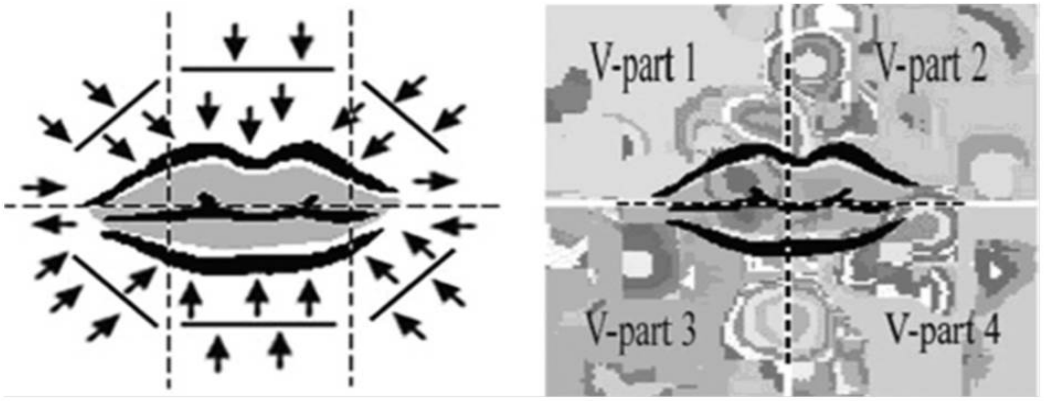

Figure 2.16 Dimension reduction of the extracted velocities in the mouth region [12]

Eveno et al. [109] proposed a liveness detection method by measuring the correlation between the movement of the lips and the speech produced. Linear predictive coding (LPC) was used to parameterize the speech signal. The LPC filters parameters are extracted at the video frame rate of 25fps. The video parameters are derived from the outer contour of the lips. The algorithm requires the manual selection of a single point above the mouth in the first frame and then the remaining segmentation is automatically achieved by fitting a deformable template. Five audio parameters and three video parameters are associated with each video frame. Canonical correlation analysis is the statistical method used to measure the relationship between two sets of the multi dimensional data. This method was used to find the linear combination of the audio and video variable

correlation. They also considered another method called Coinertia Analysis, a statistic tool used to solve problems in ecology.

Bredin et al. [110] explored the synchronisation between the motion of the lips and the sound of the speech of the talking face. Talking faces contain more information which is available for verification, and not only contain the voice signal and video signal but the most important dynamic detail which is the correlation of the movement of the lips and the sound produced by the speech. They cross fused the audio and video signals to estimate the correlation between them.

Kant et al. [111] proposed a technique in which the user was asked to perform some activities to check liveness of the attempt. They were asked to act like chewing, smile or forehead movement. The camera captured sequences of images at certain frame rate while the users were responding to the request. They extracted the feature from the facial using correlation coefficient and image extension feature. Using some discriminant analysis method, images are discriminated and skin elasticity is calculated. The output is compared with the stored database to discriminant between fake and genuine attempts.

Saad [13] explored challenge response mechanism to avert spoofing attempt. They located face in the captured images and calculated the center of the face in the images as a reference point to track it once the challenge begins. The users were asked randomly to look toward right, left, up and down features were then estimated. Figure 2.17 shows the head movement in all four direction. The collected data of 21 users providing both still and interactive attempts. The videos were replayed using phone and tablet for spoofing attempts.

Singh et al. [112] suggested a liveness detection method where random challenge was generated to the user. The user's response was observed. The users were

Figure 2.17 Head motion actions examples [13]

ask to open and close the mouth or eyes.

2.2.4 3D face

Some authors use methods based on the 3D structure of the face for face liveness detection. Andrea Lagorio et al. [14] proposed a novel liveness detection method, shown in Figure 2.18, based on the 3D structure of the face. The method computed the 3D features of the captured facial image data to detect whether a human face has been presented to the acquisition camera. They collected a 3D facial database using a stereo camera system (VERTRA3D CRT) for performance evaluation.

Wang et al. [15] explored novel liveness detection approach to counter spoofing attacks by recovering sparse 3D facial structure using a single camera. They detected facial landmarks and selected key frames from a face video or several

Figure 2.18 Proposed anti spoofing system [14]

images which are captured from several viewpoints. The sparse 3D facial structure is recovered from the selected key frames and from the selected key frames, the sparse 3D facial structures are recovered. The SVM classifier was used to test the efficiency of the proposed method in classifying genuine and fake attempts. For experiments, the authors had collected three databases using different quality cameras to inspect the anti-spoofing performance across different devices. The proposed approach achieves 100% for both classification results and face liveness detection accuracy. Genuine and Fake attempt examples are shown in Figure 2.19.

2.2.5 Face Liveness Assessment Using Motion Analysis

Kollreider et al. [113] combine face parts (nose, ears) detection and optical flow estimation to determine a liveness score. They assumed that a 3D face produces a 2D motion which is higher at central face parts (e.g. nose) compared to the outer face regions (e.g. ears). The parts nearest to the camera move differently to parts

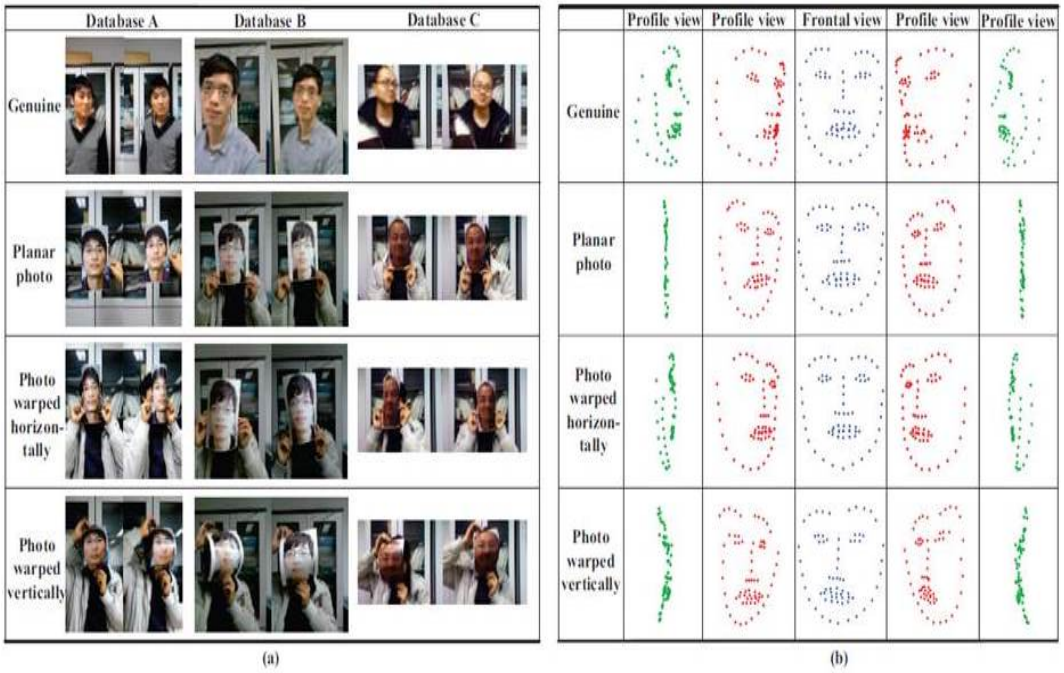

Figure 2.19 The Genuine and Fake attempt example [15]

which are further away in a live face. On the other hand, a translated photograph generates constant motion at various face regions. For the face part detection they employ a model-based Gabor decomposition and SVM. They locate the position of the face parts and compare their speed relative to each other. They also compare the direction of the motion of the same face part. This enables them to distinguish a live face from a photograph.

Continuing the work Kollreider et al. [114] exploited lightweight optical flow for face motion estimation using structure tensor and input frames. The authors presented a technique for computing and implementing the optical flow of lines (OFL). Here they again used the model-based local Gabor decomposition which are linear filters for edge detection and SVM.

The authors introduced two approaches for the face parts detection. First one was based on optical flow pattern matching and model-based Gabor feature classification. The second one extracted Gabor features in a non-uniform retinotopic grid and classifies them with trained SVM experts. The database which was

used contained 100 videos of Head Rotation Shot-subset (DVD002 media) of the XM2VTS database. Data was downsized to 300x240 pixels. Videos were cut (3 to 5 frames) and were used for live and non-live sequences. Each user's last frame was taken and was translated horizontally and vertically to get two non-live sequences per person.

Therefore, 200 live and 200 non live sequences were examined. Most of the live sequences achieved a score of 0.75 out of 1, whereas the non-live pictures achieved a score less than 0.5. It was also noticed that glasses and moustaches lowered the score, as they were close to the camera. The authors mentioned that the system will be error free if sequences containing only horizontal movements are used. By considering a liveness score greater than 0.5 as alive, the proposed system separates 400 test sequences with error rate of 0.75%.

Bao et al. [16] presented a method based on optical flow field. The difference of the optical flow filed generated through the movement of the two dimensional plane and three dimensional object was exploited for face liveness detection. The relative motion between the two dimensional plane and camera are four types named translation, rotation, moving forward or backward and swing. All other movements are combinations of these four basic types. The authors described four types of motion that can generate different optical flow field as shown in Figure 2.20. Any planar object's optical flow field can be represented as a linear combination of these four basic types with regularity.

During the investigation the authors found that translation, rotation and moving generated almost similar optical flow fields for both two and three dimensional objects whilst swing generated optical flow field that have much difference between two and three dimensional object. Their approach was based on the idea that the optical flow field for 2D objects can be represented as a projection transformation. The optical flow allowed them to deduce the reference field,thus

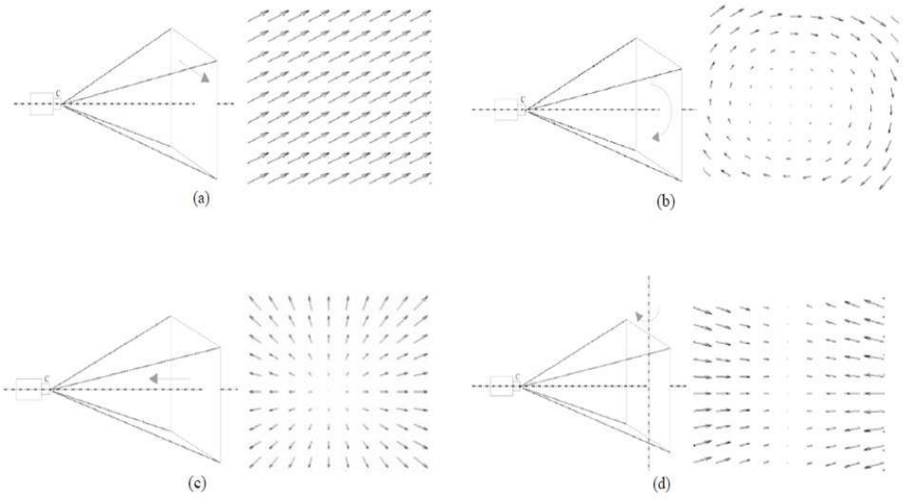

Figure 2.20 Optical flow fields generated by four basic types of relative motions (a) Translation (b) Rotation (c) Moving forward or backward (d) Swing [16]

allowed them to determine whether the test region is planar or not. For that, the difference among optical flow fields was calculated. To decide whether a face is a real face or not, this difference was noted as a threshold. The authors carried out experiments on three types of data. The first set contained 100 printed facial photo that were translated and rotated in front of the camera randomly. The second set of data contains 100 facial photos which were folded and curled before the photos were presented to the camera. The third set of data consisted of faces of real people doing random gestures like swinging, shaking, etc. Ten people participated the experiment in turn, with a total of 10 turns.

Tronci et al. [115] explored the information extracted from motion and clues from still images and video. They captured spatial information of the still images using different visual features as color. Video-based analysis was performed as a combination of information of motion such as blink, mouth movement and facial expression change among others.

Pinto et al. [116] proposed technique for video spoofing attempts. They explored that the addition of noise pattern in the sample is inevitable in acquisition

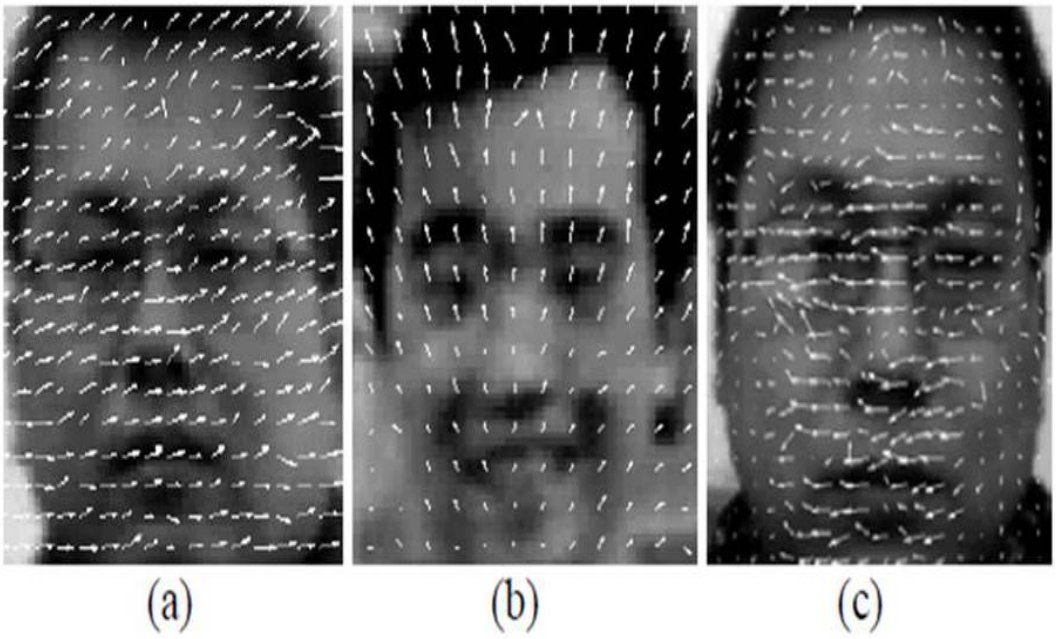

Figure 2.21 Examples of the optical flow fields with (a) Group 1 (b) Group 2 (c)Group 3 [16]

process. Fixed pattern noise and noise resulting from the sensor due non-uniform light-sensitive can be present in photo [117]. Noise pattern has been widely explored in digital document forensics [117, 118]. The exploited the noise signatures generated by the recaptured video to distinguish between genuine and fake. They used Fourier spectrum and then compute the video visual rhythms [119].

2.2.6 Miscellaneous Technologies

Chetty et al. [120] explored fusion of super resolved texture (SRT) features and 3D shape features with acoustic features for liveness checks. The proposed SRT features allowed information related to non-rigid variations on speaking faces, such as expression lines, gestures, and wrinkles, enhancing the performance of the system against impostor and spoof attacks. They transform each image into a new parametric vector space characterised by edge image and create a database of the source edges. The low resolution data of the target edge is replaced with

high resolution from the database. The super resolution (their proposed method) is divided into two stages absorption and synthesis. In the absorption phase, the source and target image frames are transformed and added to the edge database while in the synthesis phase the single target image frame is reconstructed at a higher resolution.

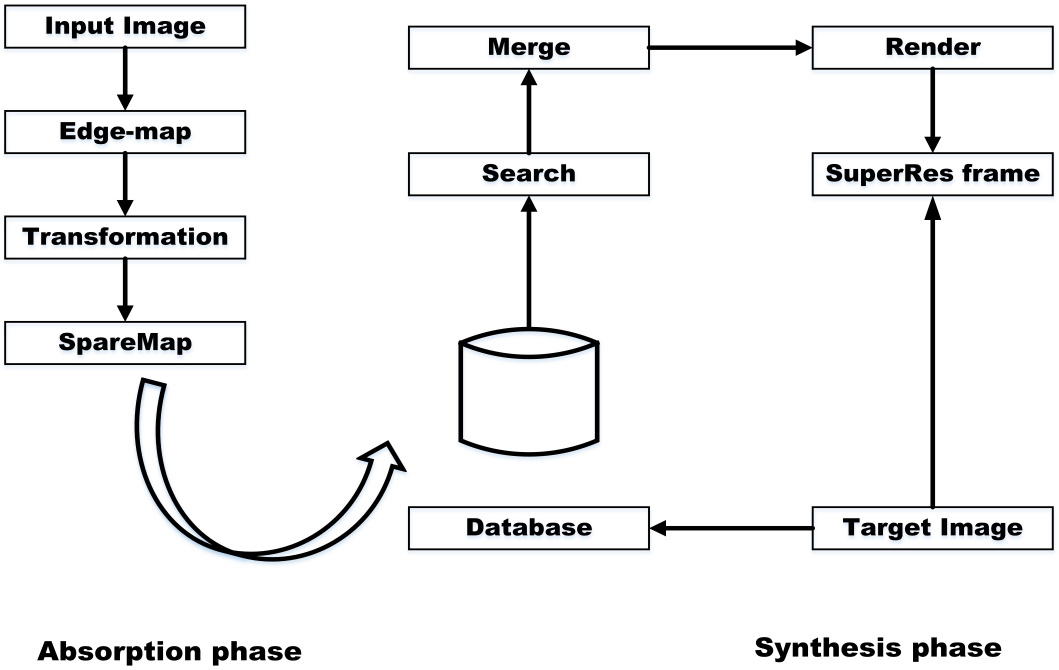

Figure 2.22 Stages of SuperRes algorithm

Synthesis: In this stage they key generation and hierarchical decomposition. Key generation is to construct keys of for each edge. These keys were then entered into the database and searched. Then, a hierarchical decomposition model was used for mapping edges to keys with the start by creating a single key for an entire edge, then recursively splitting the edge into two segments. Figure 2.22 illustrate this scheme.

Bai et al. [121] explored a physics-based method, the key idea of their approach is that when an image is displayed on paper or screen and captured the image again, the recapture image is an image of the medium (paper or screen) only.

However, because the medium has a target image, the new image appears to be the target image itself.

Chetty et al. [122] present the multi-level liveness verification (MLLV) framework for realizing the face-voice authentication system to avert audio and video replay attacks. The MLLV framework is based on feature extraction and multilevel fusion. The fusion approaches are bimodal feature fusion, cross model fusion and 3D multimodal fusion. The bimodal fusion level the system detects the still photo and pre-recorded audio but can be cheated with video replay. At the level of cross-modal fusion the system averts video replay attacks but it can be cheated with 3D synthetic talking head. The 3D multimodal fusion performs liveness check based on modelling the speaker in 3D space with the 3D shape and texture features.

Chan et al. [17] presented a method where the images of the user with and without flashlight and estimate the brightness of the face region and background and compared with each other. They believed that the genuine and fake should have different brightness difference. Figure 2.23 example of genuine and impostor video replay attack with and without flashlight. They collected data from 21 subjects.

Peng et al. [123] proposed a method based on high frequency descriptor. Instead of simply calculating the high frequency descriptor and comparing it with the set threshold, they calculated a high frequency descriptor for two images. One image was taken as normal and the other one taken with additional illumination provided by flashlight.

Kim [124] suggested a method for liveness detection using variable focusing. They captured two images sequentially taken at different focus. They discovered that in real faces, focused regions are clear and other regions are blurred due

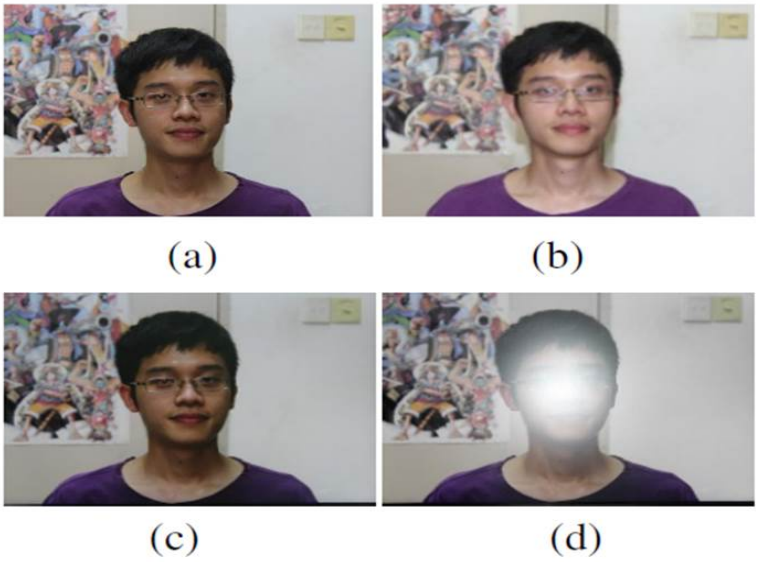

Figure 2.23 Examples of genuine and video replay with and without flashlight. (a) Genuine user without flashlight (b) Genuine user under flashlight (c) Video replay without flashlight (d) Video replay under flashlight [17]

to depth information. They also found that this did not happened in images taken from a printed photo. The extracted information based on the variation of the sum modified Laplacian [125] that represents the degrees of focusing. This information was used to discriminate between the genuine and impostor attack.

Yang [126] revised the method suggested in [124]. In this method the author proposed by investigating the focus distance between the face and background. The suggested that the focus distance should be same for photo and video and should very for genuine attempt. Kim et al. [127] further enhanced the method using the defocusing techniques this time. They argued that real face is 3D and the ear region may or may not be clear while there was little difference in clarity in case of photo using focus. This was used to classify genuine and fake face.

CHAPTER 3

Experimental Framework

3.1 Introduction

This chapter provides details of the experimental framework developed for the evaluation of the proposed approach. It presents the proposed implementation and the hardware set up used for the experiments. It also covers the definitions of different types of attack scenarios and the performance measures used to assess detection rates. The challenge design which is used for collecting the data is also explained in this chapter. The database that is developed for the purpose of evaluation of facial liveness detection methods is also described in this chapter. Facial liveness detection is a relatively new area of research, nevertheless, there are already some databases available for evaluation of liveness detection systems. However, due to the specific nature of the proposed challenge-response approach for liveness detection, none of the existing public databases are appropriate and a new database has been collected to evaluate the proposed system.

To enable the proposed system to evaluate the liveness, the user is required to interact with the system in a specific way, hence the visual stimulus providing the

challenge to the user is designed in a particular order to extract the required gaze-based novel features. The database included genuine and fake attempts (photo attacks, photo mask attacks and video replay attacks) collected from male and female volunteers.

The remainder of the chapter is organised as follows: Section 3.2 introduces the proposed system. The attack scenarios are presented in Section 3.3. Section 3.4 provides details of the system implementation and test setup. It covers system hardware, challenge design and landmark extraction. Section 3.5 provides details of the data collection covering hardware and software parameters, the number of volunteers, ethics approval and data storage. Finally, Section 3.6 presents the objective evaluation methods and metrics used in this study and Section 3.7 presents a brief conclusion of the chapter.

3.2 Proposed System

To explore the approach based on gaze stability introduced in the introductory chapters a system is proposed based on a challenge-response mechanism as outlined in Figure 3.1. The challenge is presented to the user as a visual stimulus appearing on a display screen. The client is asked to follow the shape that appears on the screen 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.

A control mechanism is used to ensure the placement of the display target shape and the image acquisition are synchronized. The system extracts facial landmarks (centre of eyes/corner of eyes) in the captured frames and computes various features from these landmarks, which are then used to classify the presentation

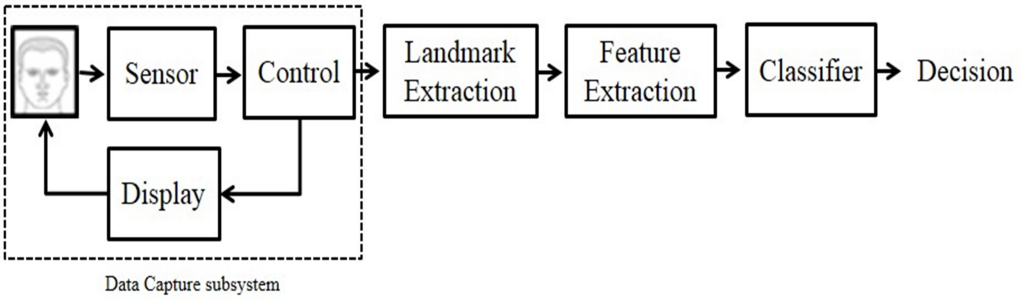

Figure 3.1 Proposed system block diagram

attempt as either genuine (i.e. coming from a live sample) or fake (i.e. coming from an impostor using a photo/mask or video attack instrument).

3.3 Attack Scenarios

Various types of attack scenarios were investigated in this study. The scenarios considered here include that of an impostor attempting authentication by holding a photograph, a simple photo mask or by replaying a recorded video of a genuine client in front of the camera of the face recognition system.

The photo presentation attack scenario uses a high quality colour photo of a genuine user held in front of the camera, whilst the attacker attempts to follow the stimulus by orienting the photograph to "face" the position of the challenge shape on the screen. In the case of photo mask presentation attacks, a high quality colour photo of a genuine user with holes made in the pupil of the eyes was held by the user in front of the eyes as a mask, and used to follow the stimulus. In the case of the video replay attack the videos of the genuine users were recorded while the users were following the challenge in a genuine attempt. These videos were then replayed later for different challenges to attempt to spoof the proposed system.

Figure 3.2 Example of Genuine attempt

Figure 3.3 Example of Photo Spoof attempt

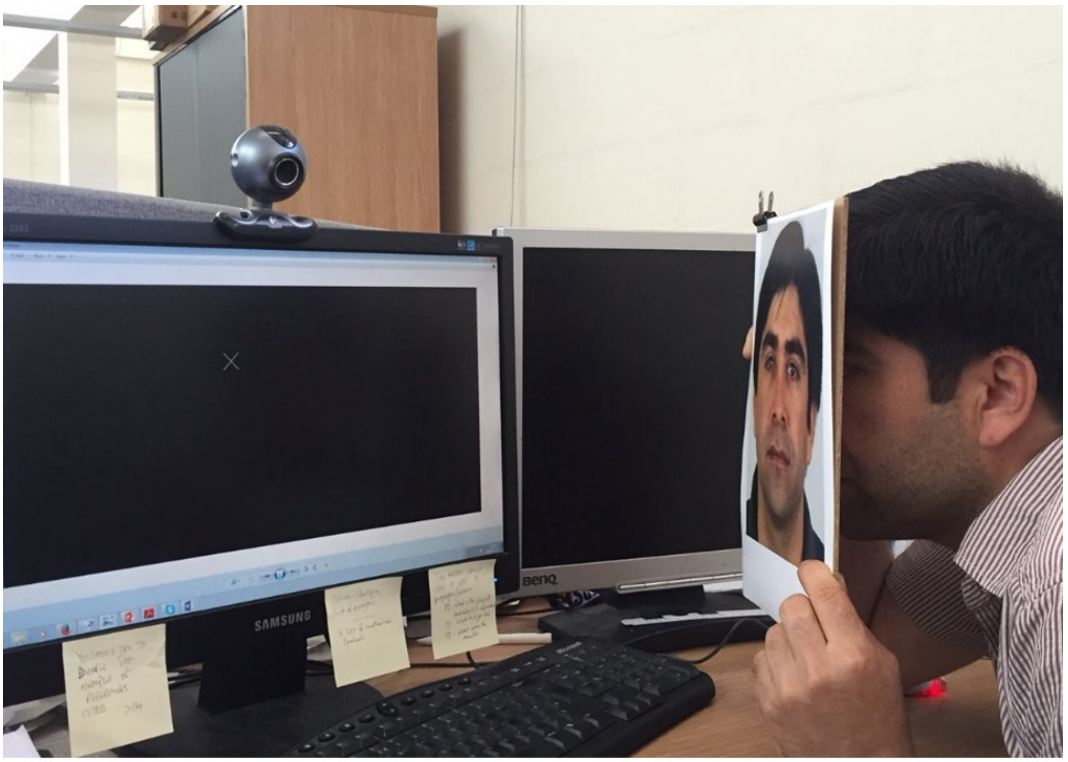

Figure 3.4 Example of Photo Mask Spoof attempt

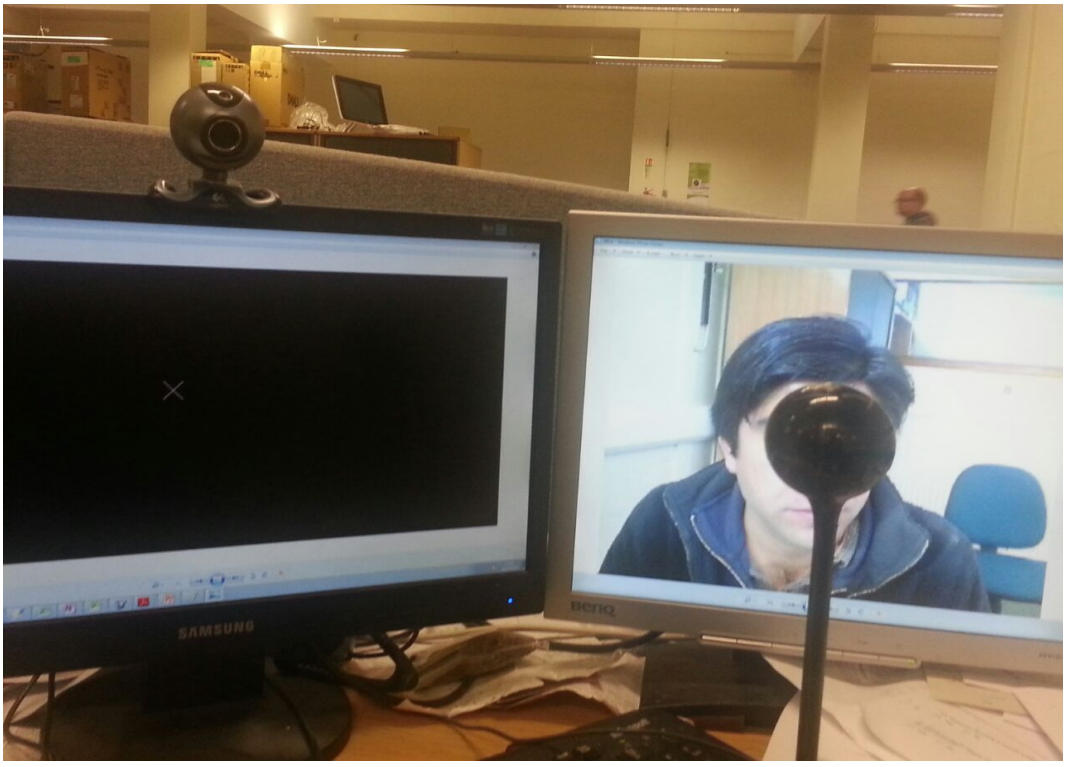

Figure 3.5 Example of Video replay Spoof attempt

Figure 3.3 shows an impostor (attempting a photo presentation attack) during the data collection process. Figure 3.4 shows an impostor (attempting a photo mask presentation attack) holding a photo mask to follow the challenge while in Figure 3.5 a recorded video is replayed to attack the system. Figure 3.2 shows a genuine user tracking the challenge to establish liveness.

3.4 Implementation

The following section describes the implementation of the proposed system. It explains the details of the hardware setup, challenge design and landmark extraction.

3.4.1 Hardware Setup

The hardware setup consists of a webcam, a PC and a display monitor of the PC. The challenge was displayed on the LCD screen for the user to follow the challenge to establish liveness. The webcam captures an image at each location of the challenge and stores these on the PC. Figure 3.6 is the sketch of the setup for the proposed system.

3.4.2 Challenge Design and Response Acquisition

A small shape (stimulus) is presented to the subjects on the LCD screen whilst subjects are seated in front of the computer screen and instructed to follow stimulus as it changes its location on the screen with natural head/eye movements. The data recording session starts after a brief cooling down period during which

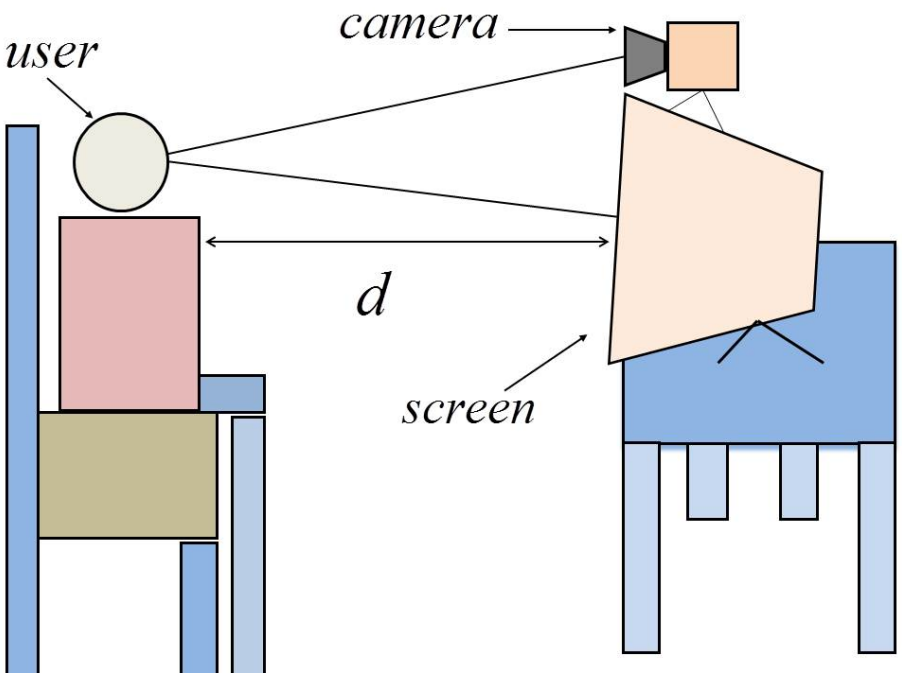

Figure 3.6 Data Acquisition Setup

the screen is black and the stimulus stationary for brief period. Figure 3.7 shows the possible predefined locations where the stimulus shape may be presented.

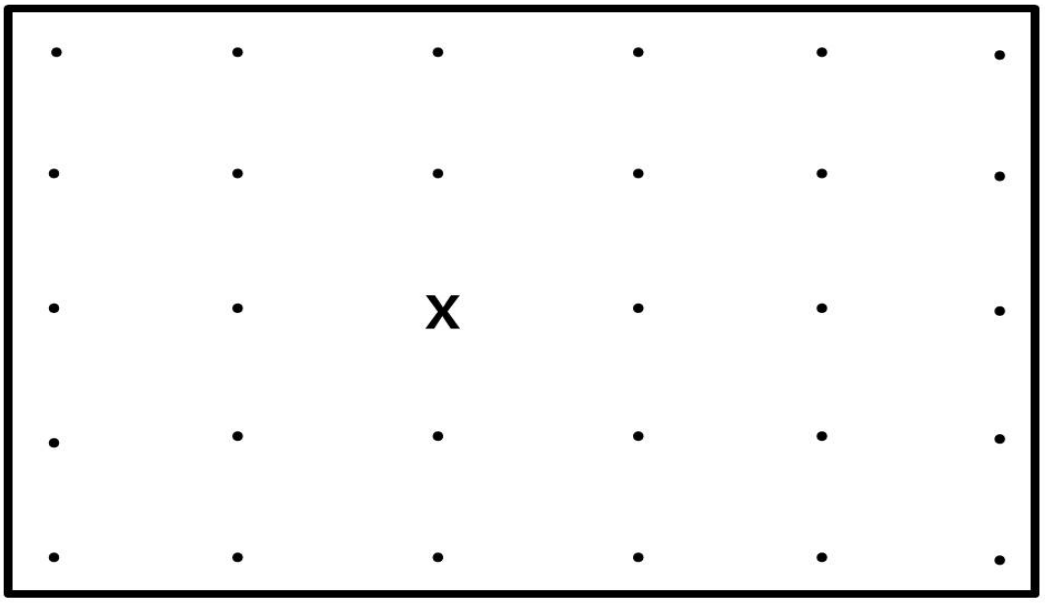

Figure 3.7 Challenge Locations

$$\begin{gathered} C = \{c_1, c_2, \ldots, c_d, \ldots, c_D\} \\ \text{where, } c_d = (x, y)\,;\ d = 1, \ldots, D \end{gathered} \tag{3.1}$$

Ideally stimulus locations should not be too close to one another, to encourage a greater range of possible head/eye movements. The set of possible locations is arranged so that some of these locations are used for the presentation of the stimulus several times during a challenge session. Let P be the sequence of M such presentations.

$$\begin{gathered} P = \{p_1, p_2, \ldots, p_i, \ldots, p_M\} \\ \text{where, } p_m \in C; i = 1, \ldots, M \end{gathered} \tag{3.2}$$

Let R be the corresponding set of landmark locations in the captured images. For a given landmark k (e.g. corner of the left eye)

$$R = \{r_{p1}, r_{p2}, \ldots, r_{pi}, \ldots, r_{pM}\} \tag{3.3}$$

$$\text{where, } r_{pi} = \{(u_{ik}, v_{ik})\} \quad 1 \leq i \leq M, 1 \leq k \leq K$$

3.5 Data Collection

To explore the performance of the proposed system initially two small data sets (Initial Database) were collected. These only included the photo spoofing attack scenario. A larger data set was subsequently collected (Extended Database) including more subjects to explore other attack scenarios and features. These databases are further described in this section.

3.5.1 Initial Database

The initial evaluation of the proposed gaze collinearity feature was carried out with a small database. Data was collected from only 5 subjects in 3 sessions. Each attempt acquired 358 image frames, and the resolution of the images was 352×288 pixels. This resolution provided adequate picture quality to recognize the facial landmarks. Higher resolution may not improve the performance but will increase the processing time [128]. Each person performed 3 fake and 3 genuine attempts in total, creating 15 sets of fake and 15 sets of genuine attempts. In this set up the stimulus was moving in a straight line continuously for the user to follow. In total, 48 vertically collinear and 24 horizontal collinear point sets were extracted.

Similarly the colocation feature was initially explored using a small data set collected from 8 subjects in 3 sessions. Each subject provided data for both genuine and impostor attempts, creating 26 sets of genuine and 26 sets of fake attempts. Subjects provided different a number of genuine and fake attempts. In total, 30 sets of x-y coordinates of the pupil centres from colocated gaze targets were extracted resulting in a feature vector of size 60 for each eye.

During spoofing attacks a high quality colour photo of a genuine user was held in front of the camera while the subject was instructed to orientate the photo towards the position of the stimulus shape on the screen. There were a small number frames where the pupil centres were not detected by software and such frames and associated colocation points) were excluded from the feature extraction process. The software used for locating the landmarks in the facial images worked better with front facial images. If the stimulus was more on the extreme left or right side of the screen, the user face was captured from onside rather then the front. This was more obvious in the case of printed photo and

later in photo mask attack where a small tilt could result in missing the whole face in the captured frame. It also depended on the quality of the captured frame. In the case of video replay the images were captured from the video playing on an ordinary LCD screen. Some of the captured frames were of bad quality due to the resolution and reflection of the screen. Hence the software could not detect faces in a number of attempts in the captured frames.

3.5.2 Extended Database

The extended data was captured in order to increase the number of subjects and also to include mask and video replay attacks. The new database was composed of genuine, photo, mask and video replay attack subjects. Photos of both male and female subjects were chosen for the hand-held photo and the photo mask spoofing trials. The photos were printed on A4 matt paper, which bends easily. Hard cardboard was attached to the back of the photo to attempt to minimize any unintended deformation of the paper.

In the case of photo mask attack, photos of different sizes and different pupillary distances (PD) were printed. Pupillary distances vary from person to person. It was not practical to print photo masks of exact PD for each subject participating in the data collection. Therefore, three different size photos (small, medium and large) were printed and the best fitting one was chosen for the impostor to facilitate the finding and follows of the stimulus by the impostor. Before a mask was given to the impostor, the pupillary distance was measured with a pupillary distance ruler similar to that shown in Figure 3.8. The photo mask with the PD closest to the impostor PD was offered to the user for the impostor attempt. The diameter of the hole in pupil centre was 4mm. The 4mm hole was large enough to see through to follow the challenge. A bigger hole could have made the task of

gaze direction easier for the attackers but may have compromised other biometric indicators of the attacker (e.g. iris) that undermined their spoofing attempt [112]. The male masks pupillary distances were 27mm, 31mm and 35mm for small, medium and large respectively. The female masks pupillary distances were 25mm, 29mm and 33mm for small, medium and large marks respectively.

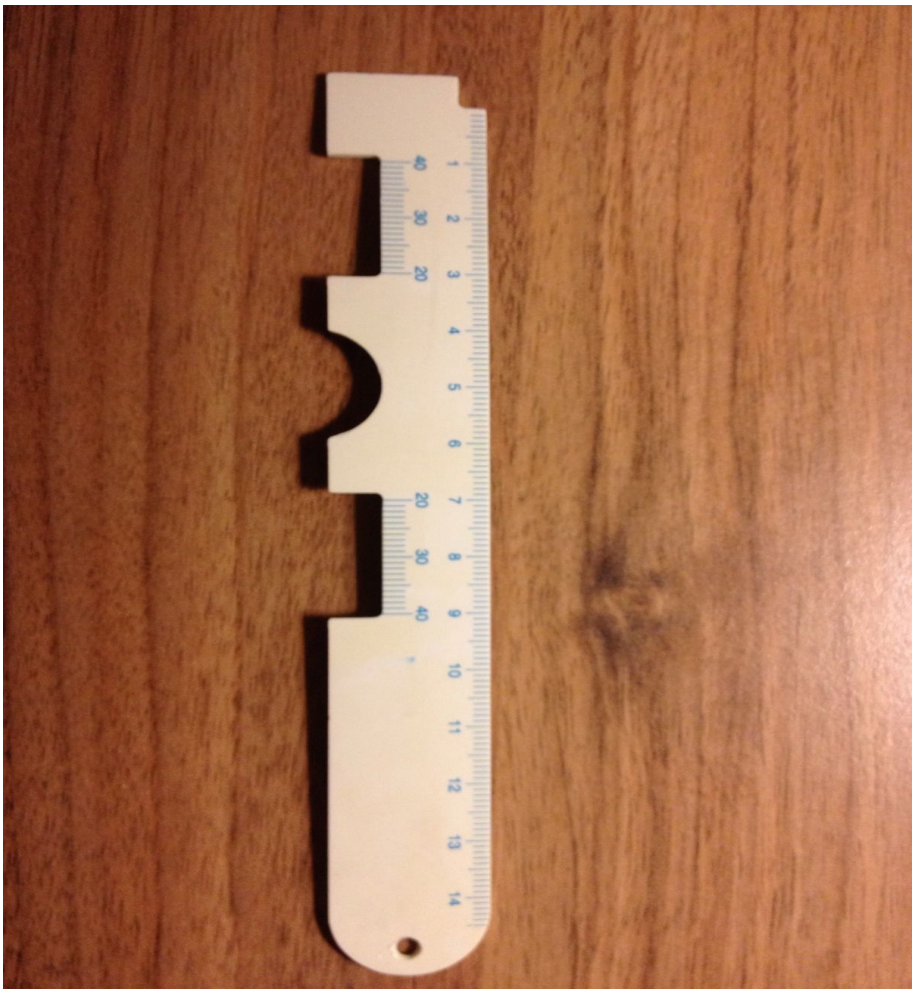

Figure 3.8 Pupillary distance ruler for measuring subjects PD

During a genuine attempt, the video of the genuine user was recorded for subsequent replay attacks. In this implementation, the stimulus was displayed on the screen at 30 distinct locations (i.e., D=30) in a random order visiting each position 3 times (thus, M=90). Typically 225-275 ms have been measured for gaze fixation in reading tasks [77]. In this work a 1 second delay was used between each presentation of the stimulus shape to provide ample time for the users

to fixate their gaze. Total duration of the challenge presentation was about 2 minutes. The locations are so arranged that there are 33 collinear sets and 30 colocation sets (i.e., L=33, W=30). For each presentation of the stimulus, the camera acquires a facial image.

Data collection involved material that are considered to be of a sensitive and personal nature, hence ethical approval was required. An ethical approval application was made to the Sciences Research Ethics Advisory Group (REAG) at the University of Kent together with a copy of the project protocol and other supporting documentations including a participant information sheet, an exit questionnaire sheet and a consent form. The Research Ethics Advisory Group approved the application with minor revisions.

3.5.3 Implementation Details

The hardware system setup was similar to the one shown in Figure 3.6. The setup consisted of a webcam, a PC and a display monitor. The camera used was a Logitech Quick Cam Pro 5000, and was centrally mounted on the top of a 21.5" LCD screen, a commonly used monitor type, having a resolution of 1920 × 1080 pixels and a 5ms response time. The reason for using a standard webcam was to establish the usability of a low-cost off-the-shelf sensor, which may be already embedded or available for many devices such as smart phones, laptops and note books.

The computer used had a quad core processor with a 3.2 GHz clock frequency, and 2 GB of RAM. The distance between the camera and the user was approximately 750 mm. There is flexibility in this distance provided that the facial features can be clearly acquired by the camera. The maximum expected gaze

deviation from the normal to the screen is approximately 15 degrees for the experimental setup. If the subject turns their head beyond this pose angle they are not following the instructions for using the system. In such a case landmarks detection may be compromised. Such frames are excluded in the feature extraction phase. If this occurs 5 times or more in a single presentation attempt, the whole attempt is excluded from the experiment, the user is asked to try again in a new attempt. The choice of this number is determined by the number of points and their placement in the stimulus. Such attempts were considered to be cases of Failure to Detect Liveness. If the number of such frames is 5 or less, then the missing landmark values were substituted by estimated data from the remaining landmarks. The data acquisition system setup was similar to that shown in Figure 3.6.

STASM Version 3.0 [128] is a software package for locating landmarks in the facial image using an Active Shape Model. This software is written in C++. STASM works on front view facial photo showing neutral expressions. Poor fits can be experienced on faces at angles or with expressions. STASM returns 68 different landmarks on the face region using an active shape model algorithm as shown Figure 3.10. If there are several faces in an image, STASM operates on the face with the largest dimension. STASM converts the captured image to monochrome before processing it. The coordinates of extracted landmarks were used for feature extraction in the proposed system and the landmarks were returned in an integer array. The first element of this array is the x coordinate and the second element is the y coordinate of the landmark.

However, in a small number of frames some of the desired landmarks were not detected by STASM. When the required landmarks were not detected in more than five frames in one attempt, the attempt was excluded from the experiments in this study. The choice of this number is determined by the number of points

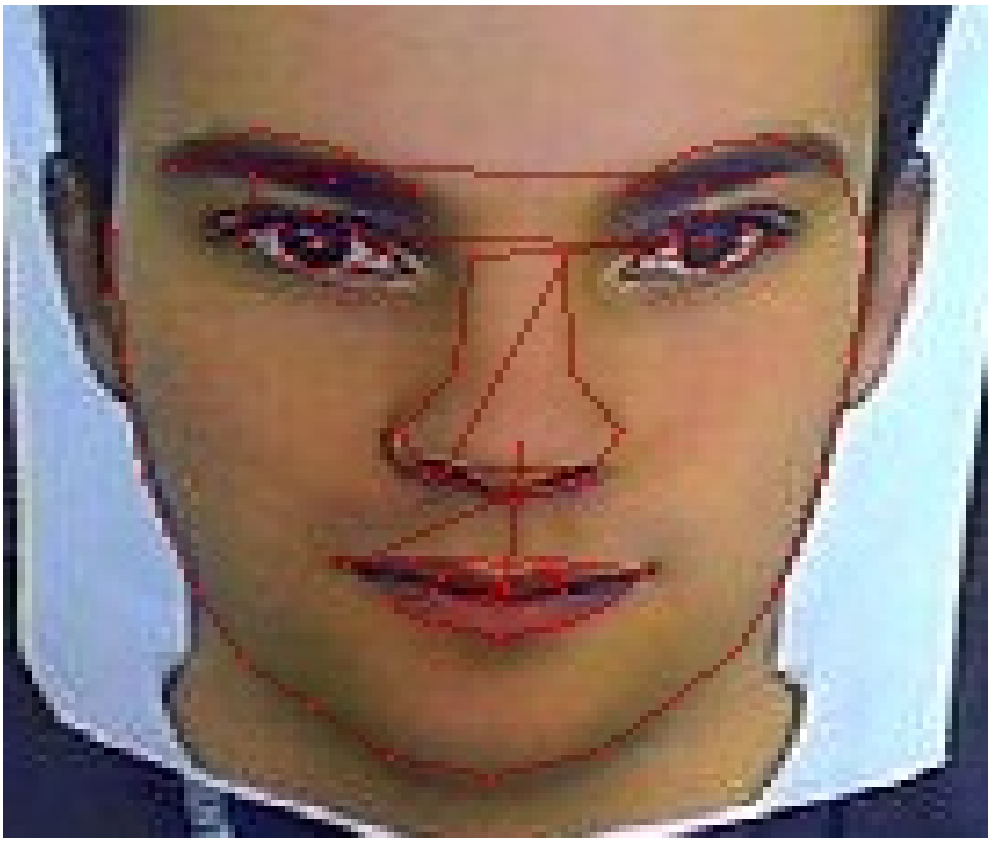

Figure 3.9 Landmarks extracted (best fit) using STASM

and their placement in the stimulus. Such attempts are considered to be cases of Failure to Detect Liveness. If the number of such frames is five or less, then the missing landmark values were substituted by estimated data from the remaining landmarks. Table 3.1 shows the detail of the database.

Table 3.1 Database details

Attempt Type	**Missing Face $>$ 5**	**Missing Face $\leq$ 5**	**Total**
Genuine	3	27	30
Photo Mat	6	24	30
Photo Mask Mat	5	25	30
Video Replay	7	23	30

For the extended experiments reported in this study, the selected database was composed of genuine attempts, hand held photos, photo masks and video replays of twenty three attempts from each scenario. The remaining seven attempts of each scenario were excluded from the experiment as these attempts had more

Figure 3.10 Landmarks extracted (poor fit)using STASM

than five photos in which the face (or facial landmarks) could not be detected by the software. Fifteen genuine samples and fifteen samples from each of the attack scenario were used for training and the remaining samples were used for testing the liveness detector.

The stimulus appears in a random sequence to prevent predictive video attacks, face images are then captured at every presentation of the stimulus.

3.5.4 Subjects

Data was collected from 30 volunteers of both male and female aged between 20-45 years. The gender balance was unequal in that there were fewer females available than males. The volunteers were from Asian, Arab, east and west

European decent. Most of the volunteers were not wearing glasses, if anyone was wearing glasses they were asked to remove them as STASM was only trained with images without glasses.

Three potential presentation attack scenarios were studied: photo attack, mask attack, and video replay attack. Each subject provided data for genuine attempts as well as for the three attack scenarios, thus creating 30 sets of data for each scenario. For the hand-held photo spoofing attacks session involved a high quality colour photo of a genuine user with holes made in the place of the eyes held in front of the camera, whilst the volunteer attempted to follow the stimulus.

Photos of both male and female subjects were chosen for the hand held photo and the photo mask spoofing trials. The photos were printed on A4 matt paper, which bends easily. Therefore, hard cardboard was attached to the back of the photo to attempt to minimize any unintended deformation of the paper. The photo mask attempt used three different photo sizes (small, medium and large) with different pupillary distances (PD) printed. The reason for producing a set of photos with pupillary holes at different distances was to match attackers with different PDs to facilitate their following of the target with relative ease.

3.5.5 Data Storage

The database materials were of a sensitive and personal nature, hence it was stored on a secure sever on a particular folder where access to the database was limited to the investigator. The size of the database was about 500MB.

3.6 Performance Analysis

Face liveness detection is a two-class classification problem. There are four possible outcomes of the classification process hereby referred to as: true positive, true negative, false negative and false positive, with "true" indicating a live/genuine presentation and "positive" indicating a live/genuine detection decision. When a genuine (live/non-spoof) presentation/attempt is classified as genuine and a false (fake/spoof) presentation/attempt is classified as genuine, these are termed true positive (TP) and false positive (FP) classifications respectively. Similarly, when a genuine presentation/attempt is classified as a fake and fake presentation/attempt is classified as fake these are called false negative (FN) and true negative (TN) cases respectively. FP and FN are the error outcomes of the process and the rates of their occurrence are reported as False Positive Rate (FPR) and False Negative Rate (FNR) in this report. The term True Positive Rate (TPR) is also used and is equal to (1-FNR) [129].

$$\begin{aligned} TPR &= \frac{TP}{(TP+FN)} \\ TNR &= \frac{TN}{(TN+FP)} \\ FPR &= \frac{FP}{(FP+TN)} \\ FNR &= \frac{FN}{(FN+TP)} \end{aligned} \tag{3.4}$$

The Total Error Rate (TER) is also used to quantify the overall performance of the system at a particular operating point and is defined in Eq. 3.5.The values for the TPR, TNR, FPR and FNR are based on Eq. 3.4.

$$TER = \frac{(FP+FN)}{(TP+TN+FP+FN)} \tag{3.5}$$

The Half Total Error Rate (HTER) is defined as a mean of False Positive Rate (FPR) and False Negative Rate (FNR). This trade-off between accuracy is through the receiver operating characteristic (ROC) curve in this study.

3.7 Conclusion

This chapter presents an experimental framework for the evaluation of the proposed approach. It covers the experimental system, attack scenarios, hardware setup, challenge design, landmarks extraction and the databases that were collected for evaluation of facial liveness detection algorithms.

The databases include data from genuine attempts and three presentation attack scenarios (photo attack, photo mask and video replay). The data which is described in this chapter will be used to explore colocation, collinearity and homography features for liveness detection that are explored in detail in the next three chapters.

The main contribution of this chapter has been,

- Collecting databases.
- Evaluation framework.

In the next chapter a first novel feature called gaze colocation will be presented.

CHAPTER 4

Gaze Colocation

4.1 Introduction

In this chapter (in part based on work published by the author in [130]) a novel feature for facial liveness detection in the presence of photo, photo mask and video replay attacks is presented and explored. This novel feature is named "gaze colocation". A similar experimental setup to the one shown in Chapter 3 has been used for data collection for the evaluation of this feature. The novel feature proposed here is based on the ability of the human gaze to return and fixate to the same location consistently.

Here a small shape is randomly presented repeatedly, at distinct locations on the screen. The user's gaze is directed to these locations on the display, and features are extracted from facial images captured at these colocated targets. The underlying hypothesis is that the variance in eye landmark positions for colocated positions should be small in genuine user attempts and greater for spoofing attempts. This phenomenon is then exploited to differentiate between genuine and fake users.

The chapter is organized as follows. The principle behind the gaze colocation feature is introduced in Section 4.2. Section 4.3 introduces the colocation feature and provides a formal definition of it. Initial experiments and further evaluations of this feature are included in Section 4.4. Section 4.5 presents extended experimental results. Finally some concluding remarks for the chapter are provided in Section 4.6.

4.2 Gaze Colocation Motivation

The idea behind the proposed new feature is that the imposter cannot align the photo back to the same location as accurately as the genuine user. The feature proposed in this chapter is also based on the assumption that the spatial and temporal coordination of the movements of the eye, head and hand involved in the task of following of a visual stimulus are significantly different when a genuine attempt is made compared with certain types of spoof attempts. The task requires head/eye fixations on a simple shape that appears on a screen in front of the user, and in the case of a photo spoofing attack, visually guided hand movements are also required to orientate the photographic artifact to point in the correct direction toward the challenge item on the screen.

To investigate the feasibility of this approach, we used the small data set presented in Section 3.5. The (x,y) coordinates of the pupil centres from frames captured while users are looking at the same stimulus location are plotted in Figure. 4.1 and Figure. 4.2 displaying deviations $\Delta x, \Delta y$ from the mean for all the genuine and fake attempts respectively. These coordinates were normalised and the figures display deviation in the x and y directions from the location where the user should be fixating the eye.

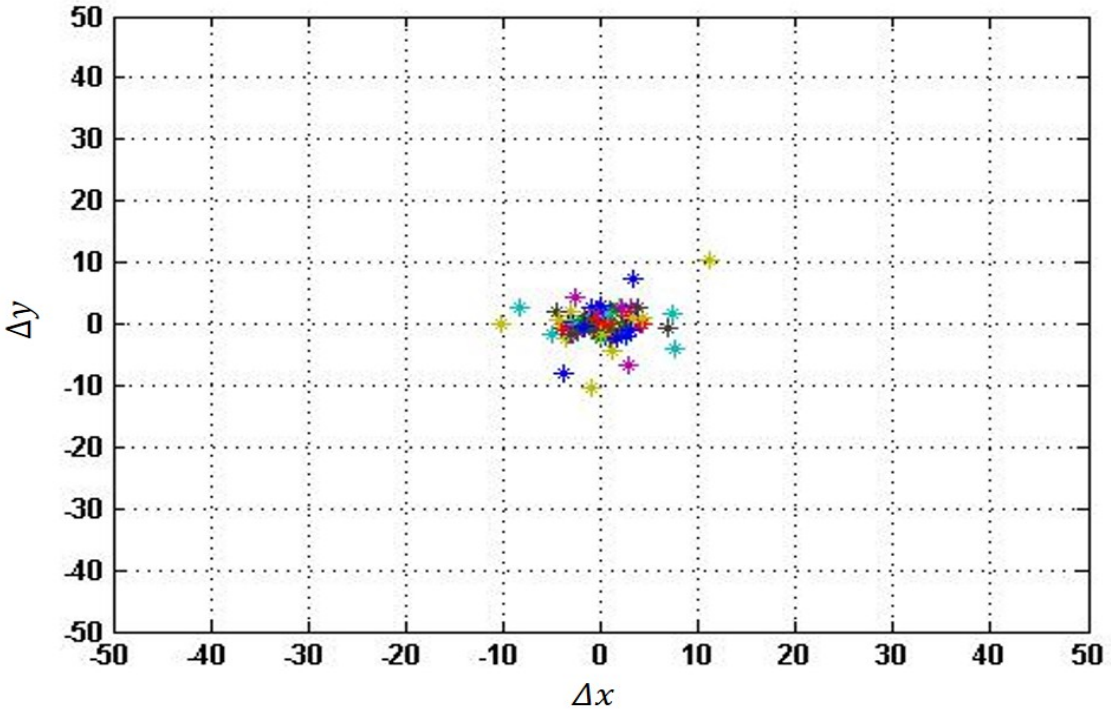

Figure 4.1 Pupil coordinates deviations from mean during genuine attempt for a location of the stimulus

The spread of the points in genuine attempts is compact compared to that of the fake attempts. This supports the gaze stability hypothesis and indicates that a impostor, relying on hand-eye coordination, is unable to align the photo back to the same spot as accurately as a genuine user. As can be seen from these figures, the deviation of the points captured during the fake attempts is approximately 4 times greater on the x-axis and 3 times greater along the y-axis than those of the genuine attempts.

4.3 Gaze Colocation Features

The gaze colocation features are extracted from images when the stimulus is at a given location. The x and y coordinates of the object on the display are the same when they reappear at a given place at different times during this exercise.

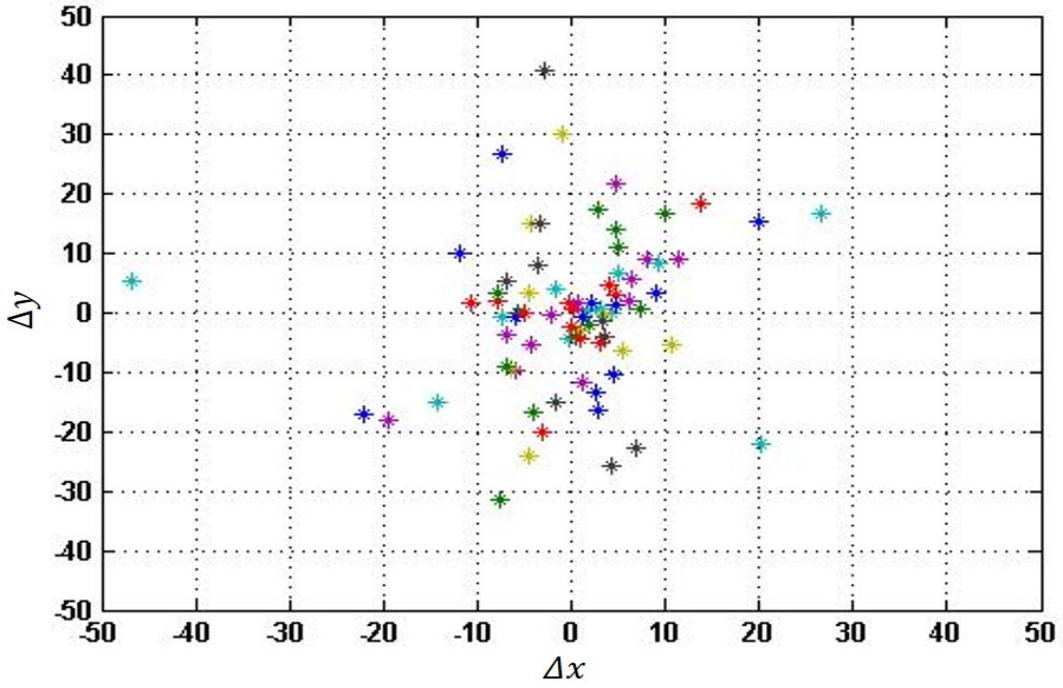

Figure 4.2 Pupil coordinates deviations from mean during spoof attempt for a location of the stimulus

It can, therefore, be assumed that the x and y coordinates of the pupil centres in the corresponding frames should also be very close. This should result in a very small variance in the observed x and y coordinates of the pupil centres in genuine attempts. A feature vector is thus formed from the variances of pupil centre coordinates for all the locations where the stimulus is colocated.

Fig. 4.3 illustrates the observed coordinates (u_{ik}, v_{ik}) of a given landmark k in response to the stimulus presented at the same location at different times. To quantify the deviation from perfect colocation, the variances in the observed landmarks are calculated.

Let Q_w be a subset of P where the stimuli appeared at the same location c_w on the screen at different times.

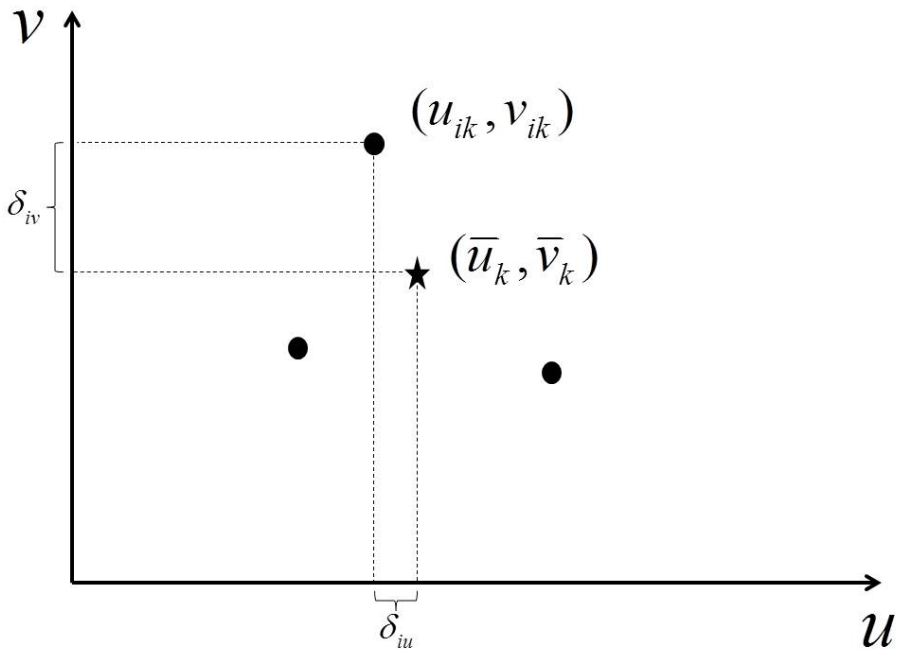

Figure 4.3 Observed (•) and expected (⋆) landmark positions

$Q_w \subseteq P$, $w = 1, \ldots, W$ where W is the number of such colocation sets used in the challenge. Let T_{wk} be the corresponding subset of R .

$$T_{wk} \subseteq R, w = 1, \ldots, W \tag{4.1}$$

Let σ^2_{uk} and σ^2_{vk} denote the variances of the observed landmarks along u and v directions respectively.

$$\begin{aligned} \sigma^2_{uk} &= \frac{1}{N} \sum\nolimits_i (u_{ik} - \bar{u}_k)^2 \\ \sigma^2_{vk} &= \frac{1}{N} \sum\nolimits_i (v_{ik} - \bar{v}_k)^2 \end{aligned} \tag{4.2}$$

where $(u_{ik}, v_{ik}) \in T_{wk}$, $(\bar{u}_k, \bar{v}_k)$ is the mean of the observed landmark locations and N is the cardinality of T_w. Let $\Gamma_{wk} = [\sigma^2_{uk}, \sigma^2_{vk}]$.

As there are K different landmarks as well as W colocation subsets, a colocation feature vector, F_{coloc}, can be constructed from the concatenation of these values

and used for liveness detection.

$$F_{coloc} = [\Gamma_{11}, \Gamma_{12}, \ldots, \Gamma_{1K}, \Gamma_{21}, \ldots, \Gamma_{wk}, \ldots, \Gamma_{WK}] \tag{4.3}$$

Many other features can be extracted from these facial landmarks which can further enhance the distinction between genuine and fake attempts. All these can be combined into a global feature vector,

$$F = [F_{coloc}, F_{other}, \ldots]. \tag{4.4}$$

If the impostor holds the photo or mask still (i.e., makes no attempt to respond to the challenge) this will produce very small changes in the measured landmarks location. However, this can also happen when a genuine user is non-cooperative or non responsive. The opposite extreme can also occur, when the genuine user's response to the challenge involves extreme head and eye movements. Some may produce large variances in response to the challenge. The maximum expected gaze deviation from the normal to the screen is approximately 15 degrees for this experimental setup. Therefore, the gaze variation for the entire attempt should also be taken into account during liveness detection. The overall variance must not be too small to ensure that the impostor was moving the photo or mask in response to the challenge movement. Such extreme behaviour can be treated as suspect and may be flagged as fake attempts. Overall variance for entire attempt was calculated as below

$$\begin{aligned} \sigma_u^2 &= \frac{\sum_{i=1}^{M} (u_i - \bar{u})^2}{M} \\ \sigma_v^2 &= \frac{\sum_{i=1}^{M} (v_i - \bar{v})^2}{M} \end{aligned} \tag{4.5}$$

Where u_i, v_i are the facial landmarks, $\bar{u}, \bar{v}$ are mean of u_i, v_i and M is total number of stimuli locations. So conditions such as $\theta_1 \leq \sigma_u^2 \geq \theta_2$ and $\theta_1 \leq \sigma_v^2 \geq \theta_2$ can be placed in order to eliminate such attempts, where θ_1 and θ_2 are some thresholds.

4.4 Experimental Evaluation

For an initial evaluation of the proposed feature the small data set presented in Section 3.5.1 was used. Features were extracted from single eye, both eyes and their feature and score fusion were investigated. In the extended set of experiments only score fusion was investigated using the larger database presented in Chapter 3, as score fusion provided the best results in the preliminary analysis.

4.4.1 Preliminary Experimental Results

The colocation features extracted from the single eye were investigated in Section 4.4.1.1. While the colocation features from each eye may be used in isolation it is interesting to explore whether there is complementarity in these feature sets and if a greater accuracy can be achieved by their combination. Therefore, both feature fusion and score fusion schemes were explored in Section 4.4.1.2, and Section 4.4.1.3 respectively to find if there would be any gain in accuracy by combining information from features extracted from both eyes.

4.4.1.1 Single Eye Feature

Colocation feature vectors were extracted from the facial images captured when the stimulus appeared at colocated locations. The features extracted from the

eye were passed to a classifier for training and testing purposes. The scheme is illustrated in Figure 4.4.

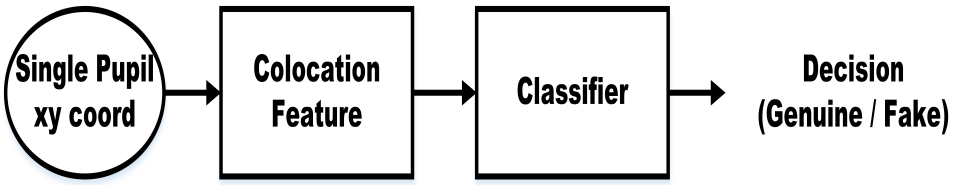

Figure 4.4 Scheme, where feature extracted from single-eye

In this experiment all 60 features from the single eye were given to a k-NN classifier for training and testing the system. The k-NN classifier from the prtools package used for this work automatically determined an optimum k-value for each experiment. Each experiment was run 400 times with random sets of data for training and testing, resulting in a different optimum k values for each run. These optimal k values were combined to obtain an overall mean k value to be used for operational systems. In the case of colocation features this mean optimal k value was 6. The ROC curve using features from the single eye is presented in Figure 4.5. It is apparent that the performance of the system is not very accurate. At 10% FPR, 40% TPR is achieved using the entire feature vector.

4.4.1.2 Feature Fusion

The features extracted from both eyes were concatenated to form a larger feature vector using feature fusion. In this scheme the entire feature sets from the eyes were combined. The combined feature set was passed to a classifier for training and testing. This scheme is shown in Figure 4.6.

All 60 features from the left eye and the 60 features from the right eye were combined in a feature-level fusion scheme. A k-NN classifier was used to investigate the performance of the system.

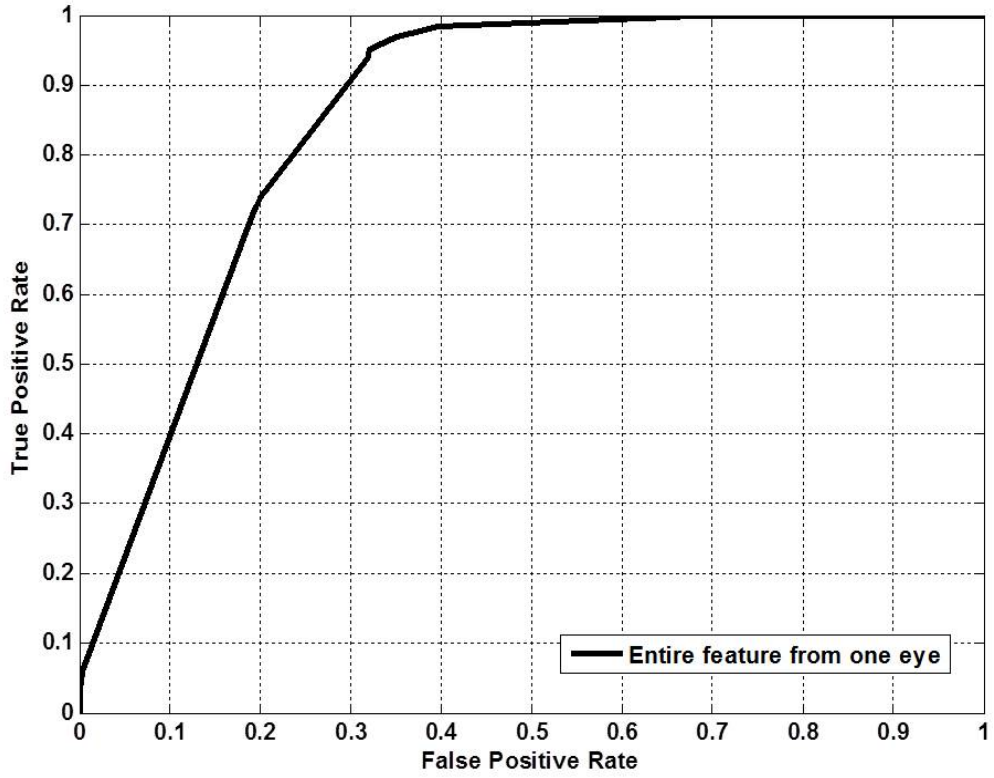

Figure 4.5 Performance with single eye feature using the entire feature vector for photo attack

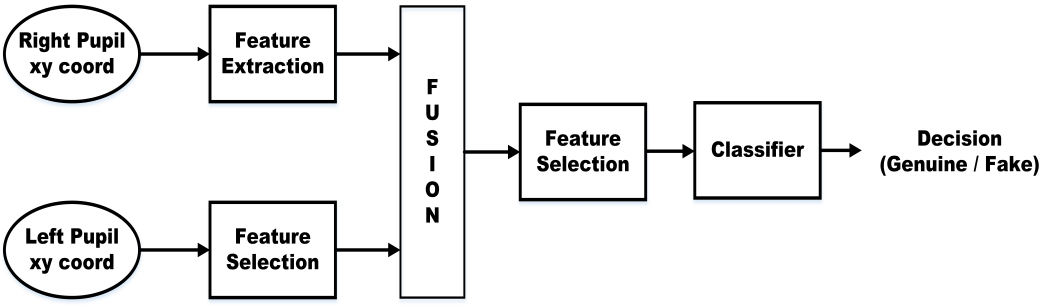

Figure 4.6 Feature fusion using left and right eye

The ROC curve of the proposed scheme using features from both eyes is presented in Figure 4.7. The performance of the system is slightly improved compared to the single-eye case, when using the entire feature vectors for both eyes. The TPR (at 10% FPR) of the system was 44% .

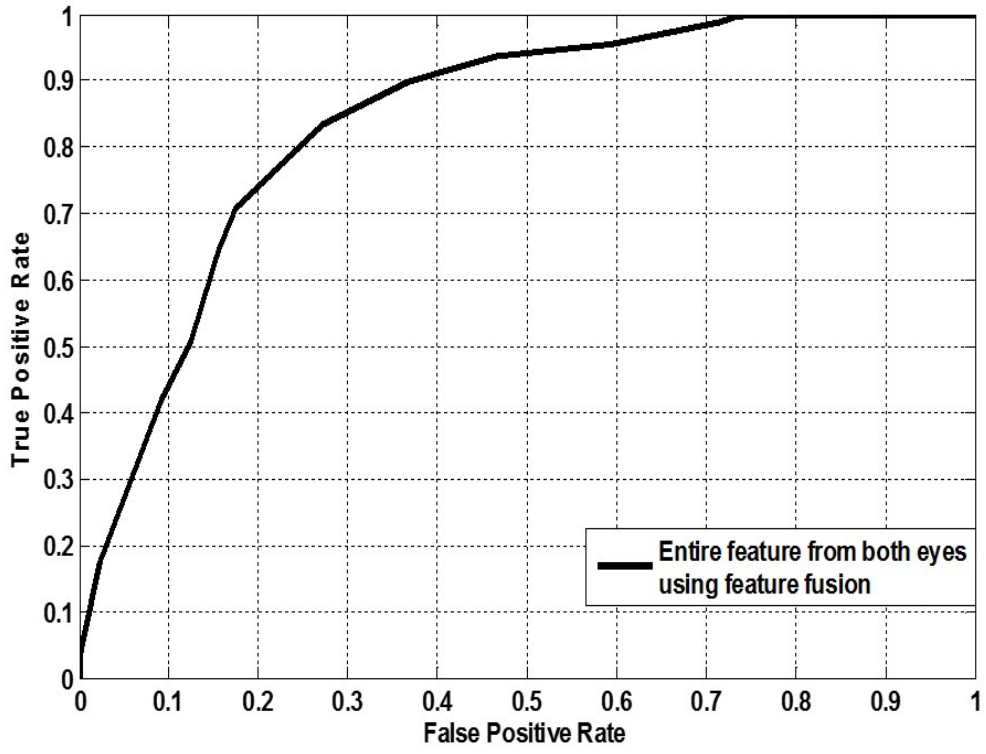

Figure 4.7 Feature fusion performance from both eyes for photo attack

4.4.1.3 Score Fusion

In the score fusion scheme, features were extracted from both eyes and independent classifiers are used to obtain classification scores for each eye. The scheme is illustrated in Figure 4.8.

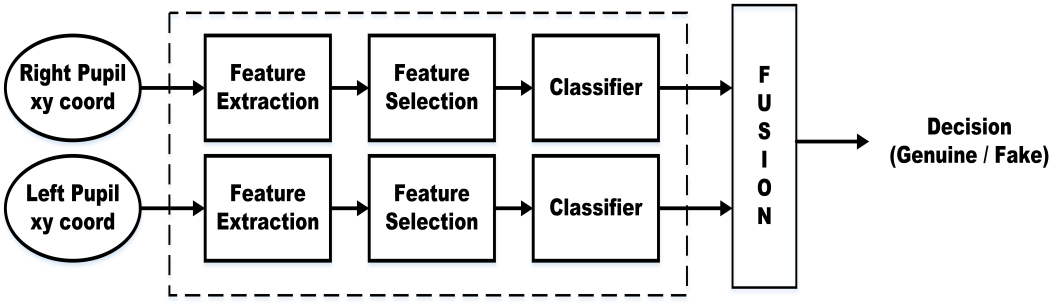

Figure 4.8 Score fusion scheme

In this multi-classifier system two k-NN classifiers are used, one for each eye. The a-posteriori probabilities from the two classifiers were combined using fusion. Figure 4.9 shows the ROC curve for the photo spoofing attack scenario. The scheme achieved a TPR of 68% at FPR of 10%. This performance is much improved compared to the single-eye scheme and the feature fusion scheme using both eyes.

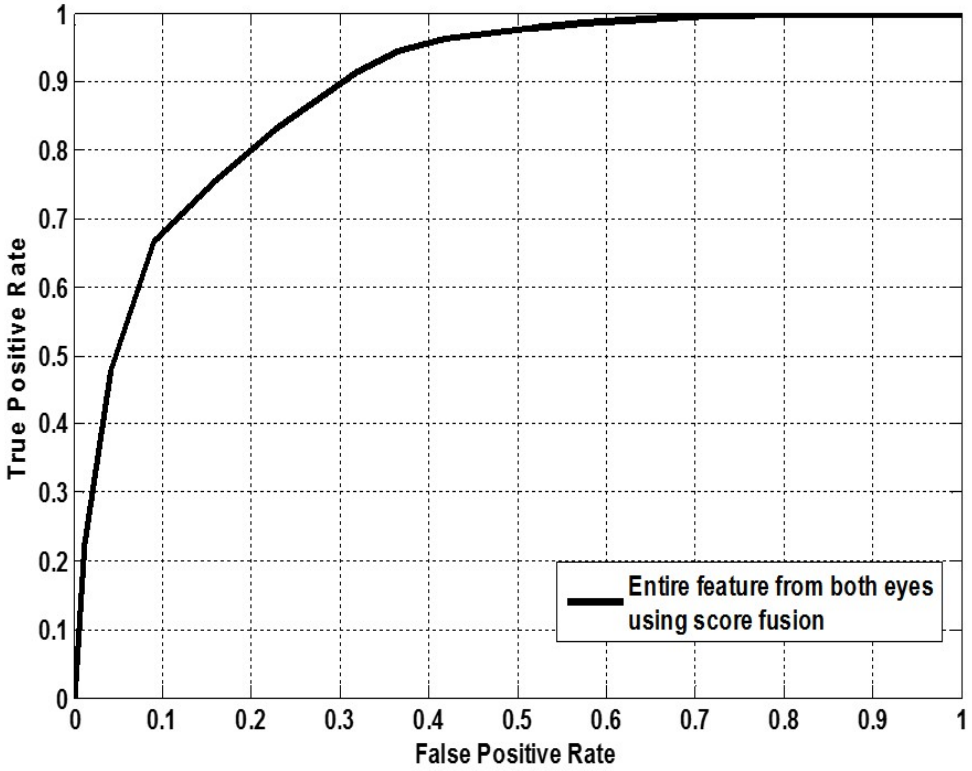

Figure 4.9 Score fusion performance of both eyes for photo attack

To investigate the tradeoff between the feature dimensionality and liveness detection accuracy, several experiments were performed to explore the performance of the proposed method. The forward feature selection [131] method was run 400 times with random sets of data for training and testing. This resulted in different rankings of features for each run. These rankings were combined to obtain an overall ranking as follows. The feature that most frequently had the first rank

was given the first overall ranking. This procedure was repeated for all the other ranks so that the feature that appeared most frequently at rank N was given rank N in the overall ranked list. As the number of feature elements (dimensionality of the feature vector) were steadily reduced to a certain level, the performance of the method was improved.

Figure 4.10 illustrates total error rates for different feature dimensions selected.

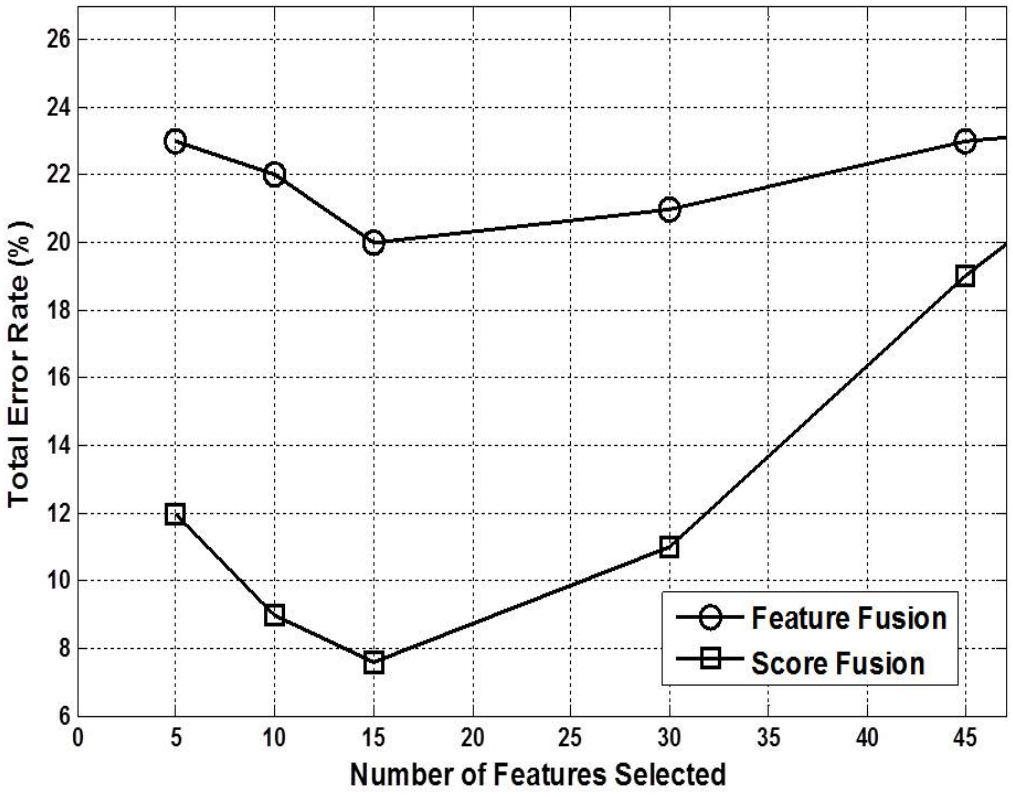

Figure 4.10 Variation in accuracy with feature dimension for feature and score fusion

It can be seen in the Figure 4.10 that the error is reducing as the feature dimension was reduced. The lowest total error rate was observed when the feature dimension was reduced to around 15. The total error rate started increasing again when the feature dimension was further reduced. The system produced higher total error rates when the feature dimension was larger.

Figure 4.11 shows the ROC curves for different feature a single-eye for photo attack. Reducing the number of features improved the performance but the best results were achieved when using a subset of best 15 features (as shown in Figure 4.11). At 10% FPR, the TPR exceeded 90% which was only around 40% when using the entire feature set.

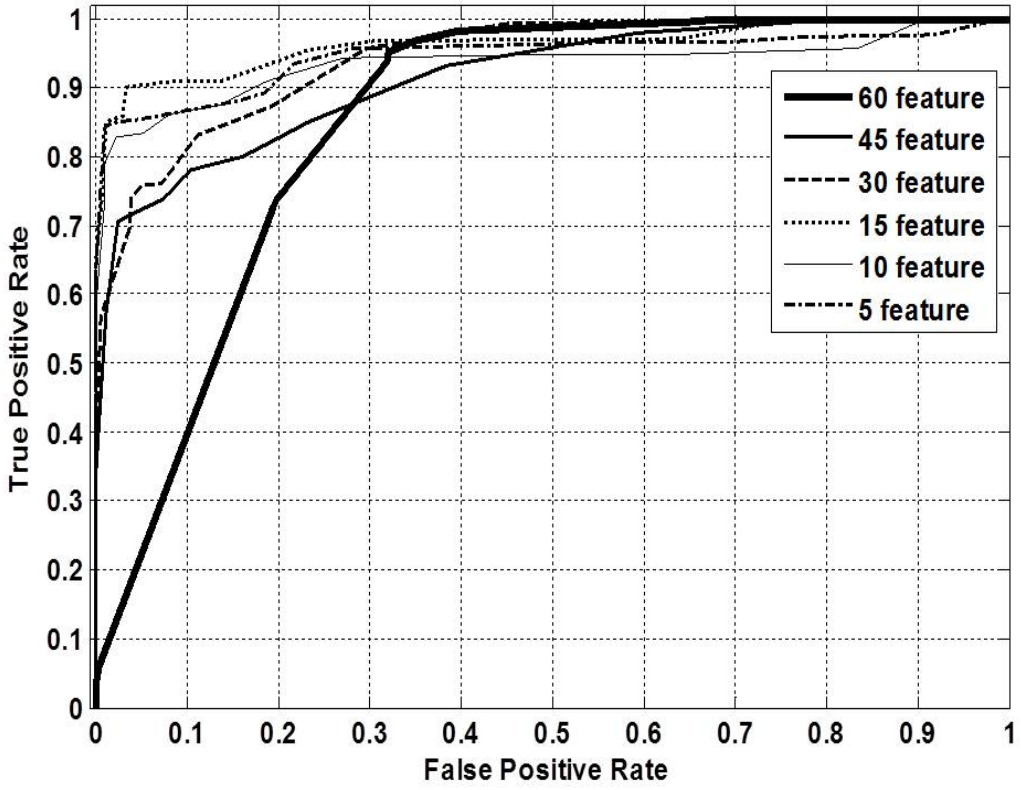

Figure 4.11 Feature performance with single-eye feature

Similarly Figure 4.12 shows the performance of the feature fusion of both eyes for optimum feature and other various feature sets.

The results improved using the optimum feature set. In Figure 4.12 ROC curves for various subsets of the feature set are presented. At 10% FPR, the TPR was approximately 80% using 15 best features. This was only around 44% when using the entire feature set.

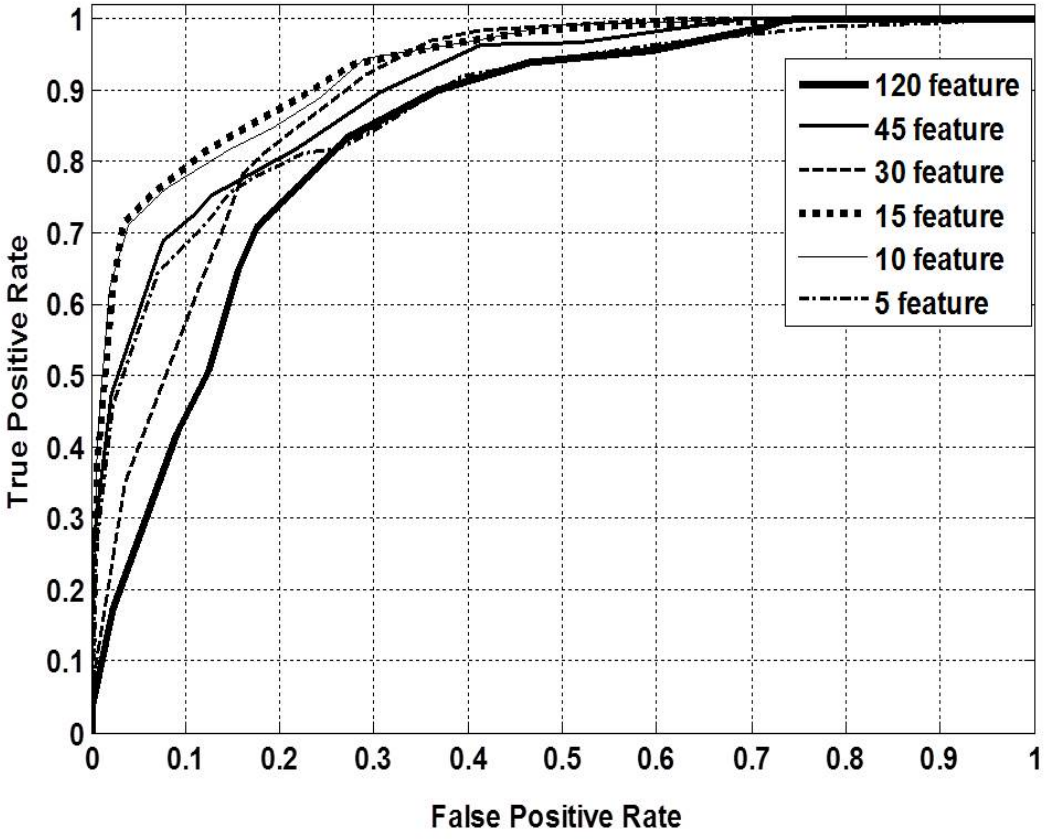

Figure 4.12 Feature fusion performance

The ROC curve of the score fusion scheme using optimum features from left and right eye is presented in Figure 4.13. It is apparent that the performance of the system is improved significantly using the optimum feature set. At 10% FPR, the TPR exceeded 99%. This was only around 68% when using the entire feature set.

Table 4.1 presents a comparative performance of the proposed methods at various levels of FPR. The feature fusion scheme gave the highest error rates in all cases. While using features from only single-eye, the TPR was 91%. This improved considerably when the score fusion approach was implemented. At 1% FPR, a TPR of 93% was achieved using score fusion. At 10% FPR, this rose to 99%.

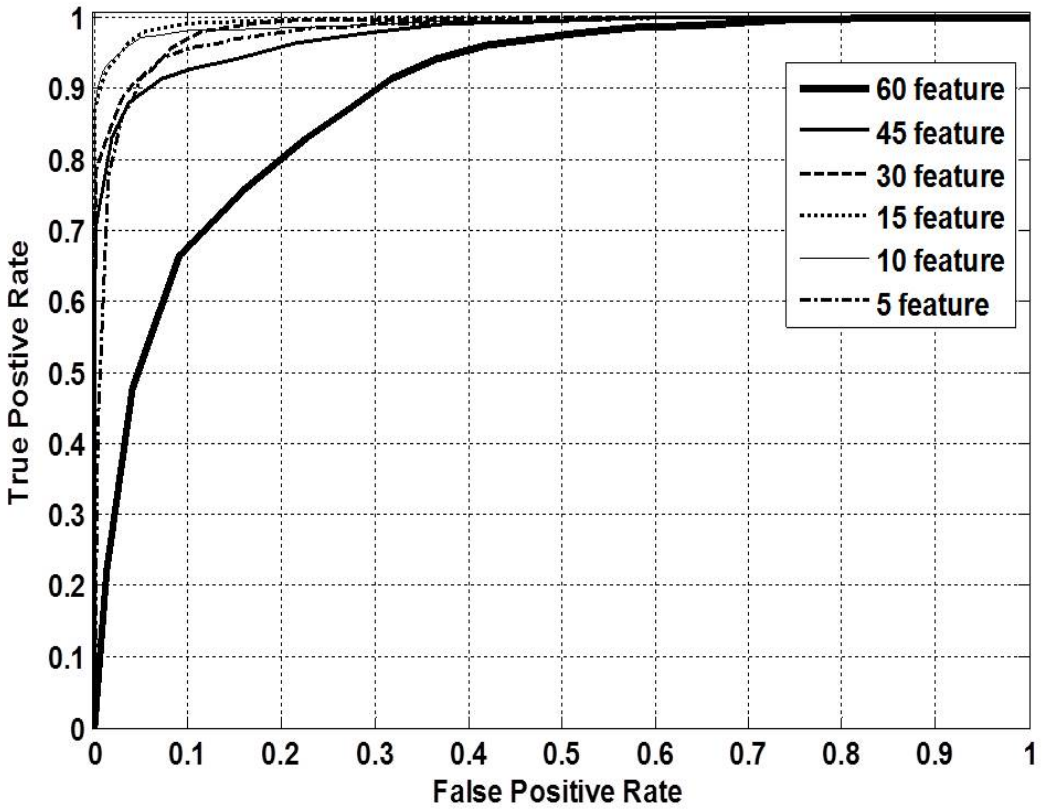

Figure 4.13 Score fusion performance

Table 4.1 TPR comparison of the three schemes at various FPRs

	FPR=0.01	**FPR=0.02**	**FPR=0.05**	**FPR=0.10**
Signle Eye	0.84	0.86	0.90	0.91
Feature Fusion	0.47	0.62	0.74	0.78
Score Fusion	0.93	0.94	0.97	0.99

4.5 Extended Experimental Results

Initial experiments using the small database provided an encouraging indication of the potential of the proposed colocation feature and fusion schemes for photo attacks. The initial experiments also indicated that the score-based fusion scheme outperformed the other schemes (single eye, feature fusion). It was decided to explore whether the choice of features and fusion scheme that would work for

detecting other types of presentation attacks (using masks or video replay).

Further experiments were carried out to test the proposed novel feature on a larger database and other types of attack scenarios to confirm the claims made in the preliminary experiments. These experiments included photo, mask and video replay attacks.

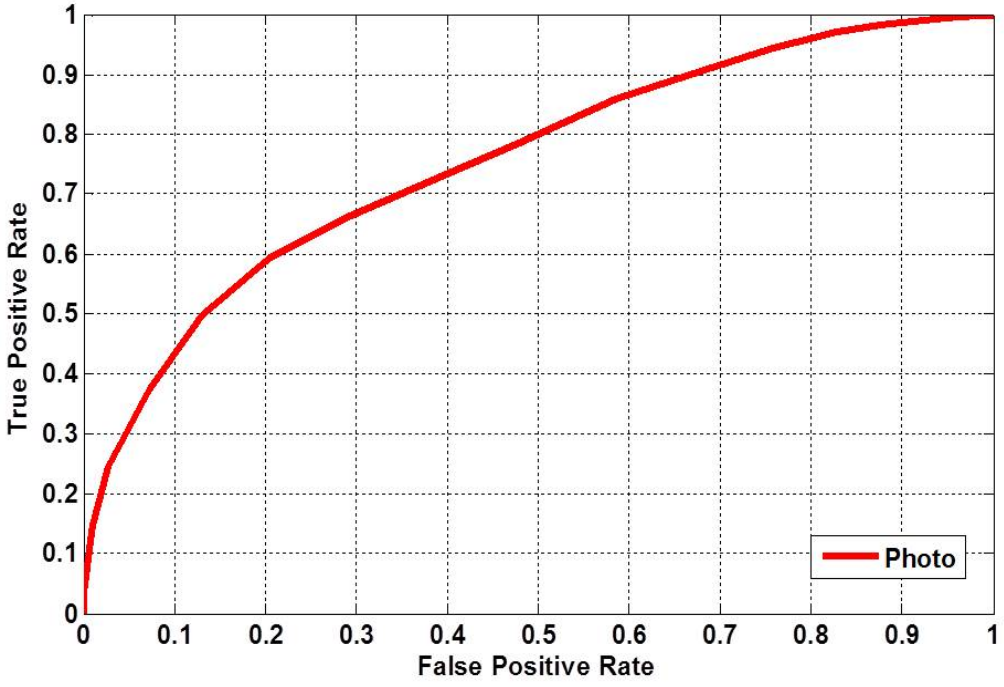

Figure 4.14 ROC curve of the colocation feature using entire feature set for photo attack

The colocation feature performance for photo spoofing attacks using the entire feature set is illustrated in Figure 4.14. The performance of the system is poor but it still classifies some spoofing attempts correctly. This could be improved further with feature selection or in combination with other features.

Figure 4.15 shows the performance of the mask spoofing attack using the entire feature set. The system performance is poor when compared to photo spoofing

attacks. Nevertheless, the feature may be able to contribute to improve the system performance to some extent when combined with other features.

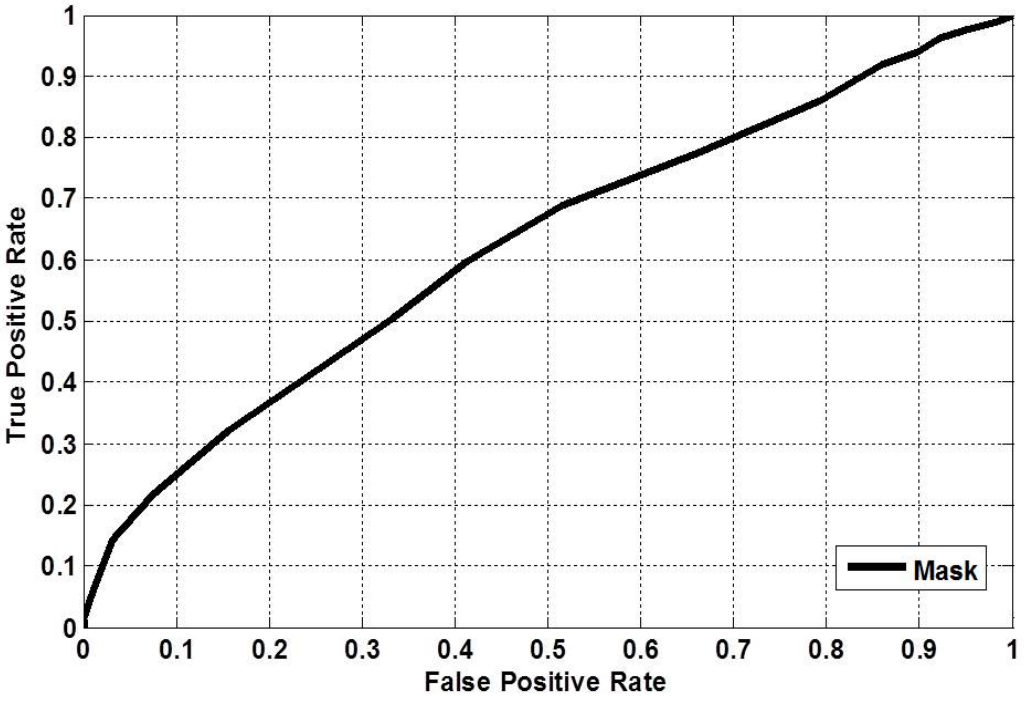

Figure 4.15 ROC curve of the colocation feature using entire feature set for mask attack

Figure 4.16 shows the ROC curves for the entire feature sets for video replay attacks. The performance of the system was found to be better compared with other attack scenarios as shown in Figure 4.14 and Figure 4.15.

Although the video is recorded from the genuine person, responding to the challenge, the challenge is random each time it runs. When the impostor was replaying the video to spoof the method, the probability of repeating the same pattern of the challenge locations is very low. Hence the subject in the video was not responding to the corresponding locations of the simulated challenge. Therefore, there is no correlation between the challenge and and video response. That is why video attack detection performed better.

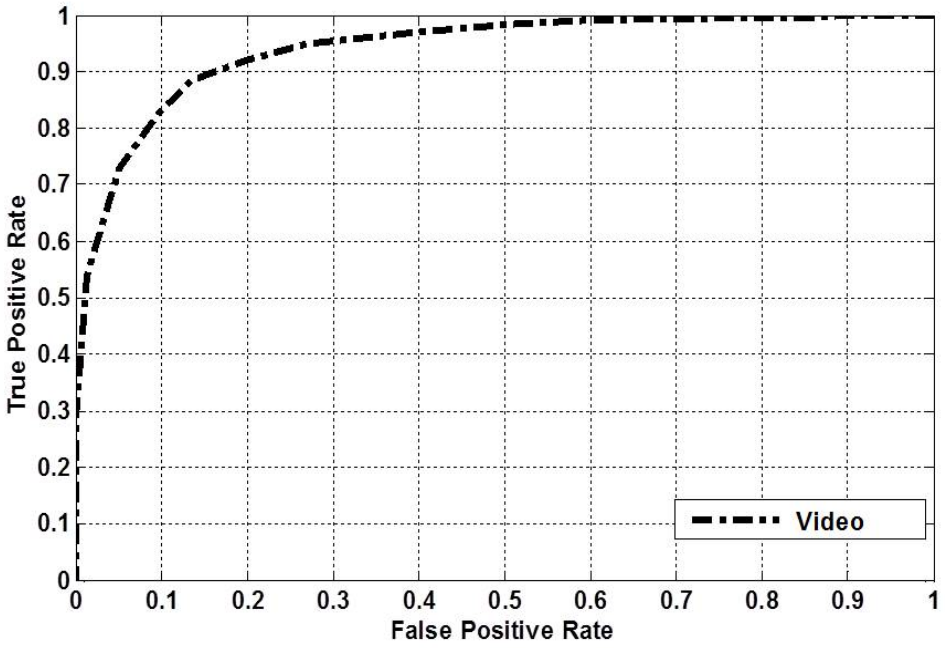

Figure 4.16 ROC curve of the colocation feature using entire feature set for video replay attack

At 10% FPR, video replay performance is 83% and photo attack TPR is about 43%. The mask attack detection performance is lower compared to the video and photo spoof detection performance. At 10% FPR, video replay performance is 25%.

This method performance could be improved to use the subset of features which are more optimum of the entire feature set. In order to find the feature dimension for such feature, forward feature selection method [131] was again used to rank the features. In similar fashion the feature selection method was run many times, at each run random sets of the data for training and testing were chosen. The results of these runs were combined to a single feature.

Figure 4.17 presents total error rates as a function of the number of features selected to find optimum feature sets for colocation features of photo and mask

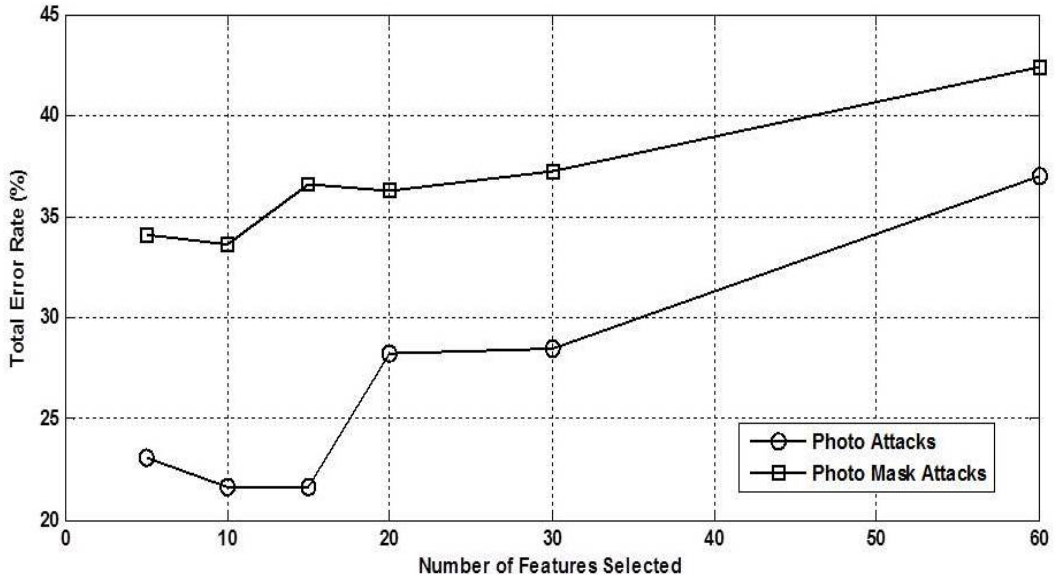

Figure 4.17 Variation in accuracy with feature dimension

spooking attempts. The total error rate was decreasing as the feature dimension was reduced until around 10 for mask attacks. In case of photo attacks the total error was almost same for 15 and 10 feature set. Further reducing the feature dimension caused the total error rate to increase.

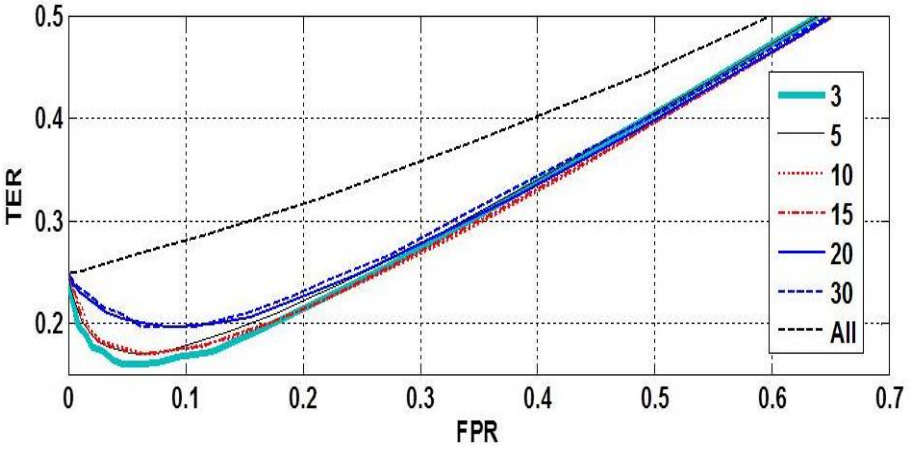

Figure 4.18 Variation in accuracy with FPR

Figure 4.18 presents the total error rates as a function of the false positive rate to find optimum sets of colocation features for a various set of features selected. In this experiment the photo and mask attacks modalities were combined and

treated together as single spoofing attempts against genuine attempts. The combination of these attack modalities allows the establishment of a single optimal feature vector that can be used for all of these major spoofing challenges. Video spoofing attack data was excluded from this feature ranking exercises as the system already has shown to perform well in detecting video spoofing attacks.

Again the feature selection method was run several times, choosing random sets of data for training and testing for each run. The results of these runs were combined to rank the features. It is shown in Figure 4.18 that the lowest total error rate was observed when the feature dimension was reduced to around 3.

The colocation feature performance for photo, mask and video spoofing attacks using this optimum feature set is illustrated in Figure. 4.19.

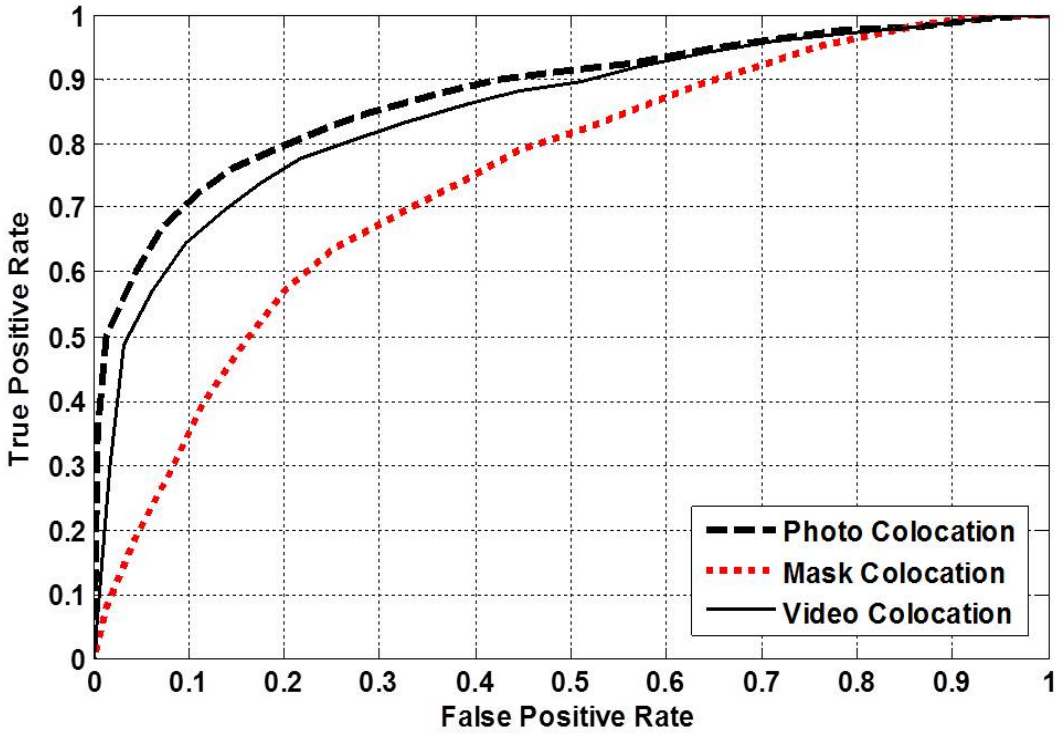

Figure 4.19 ROC curve of the colocation feature using optimum feature set

Photo attack detection gives the best performance, the video replay attack detection ranks second in performance, followed by the photo mask attack detection using colocation feature. At 10% FPR about 70%, 38% and 65% TPR are achieved for photo, mask and video replay detection respectively. In general video performance could have improved further using reduced feature set if the feature extracted from video were included in feature ranking procedure. But as is discussed in earlier section that the feature ranking was done only for photo and mask attacks because the video attack detection performed considerably better.

4.6 Conclusion

This chapter has presented a novel feature for liveness detection in the presence of photo, photo mask and video replay spoofing attacks for face verification systems. It is a challenge-response approach using a visual stimulus to direct the gaze. The test scenario did not constrain the users to move either their head or eyes exclusively. However, the proposed gaze colocation features provided a robust measure for discriminating between live and fake attempts.

Experiments support the potential viability of this approach.

The main contribution of this chapter has been,

- New feature

- Investigate single-eye, feature fusion and score fusion of both eyes.

- Investigate three types of presentation attack which were photo, mask and video replay attacks.

In the next chapter a new feature called collinearity will be presented.

CHAPTER 5

Gaze Collinearity

5.1 Introduction

The work presented in this chapter (in part based on work published by the author in [132] [133]) explores another feature set, hereby referred to as the gaze collinearity feature set, for the detection of presentation attacks. The same kind of visual stimulus which is presented in Section 3.5.3 is used here to direct the gaze of the user to sets of collinear points on the screen. The system records the gaze of the user with an ordinary webcam. Features based on the measured collinearity of the observed gaze are then used to discriminate between genuine attempts responding to the challenge and those conducted by impostors. Several sets of experiments reported later in this chapter indicate the effectiveness of the proposed method in detecting spoofing attacks. The scenario considered in the experiments reported in this chapter is that of a face recognition system using an ordinary webcam. The spoofing attack is assumed to be through an imposter holding a photograph, a photo mask or replaying a recorded video of

an individual target presented to the camera of the face recognition system and attempting authentication.

In this chapter we also explore the effect of stimulus alignments or orientations on the performance of the proposed system. The aim is to establish whether there is such a sensitivity and, if so, to explore how this may be exploited for improving the design of the stimulus. The results suggest that for liveness detection using this experimental setup collecting feature points along the horizontal direction may be more effective than the vertical direction.

The chapter is organized as follows: First the motivation for using gaze collinearity as a feature is presented in Section 5.2. In Section 5.3 a definition of the collinearity feature and a formal derivation of this feature is provided. Section 5.4 gives a summary of the experimental results. Section 5.5 explores the fusion of colocation and collinearity features. Some concluding remarks for the chapter are given in Section 5.6.

5.2 Gaze Collinearity Motivation

Collinearity feature vectors extracted from the facial images captured when the stimulus locations appeared on straight lines. When the stimulus locations on screen are along a straight line, the x or y-coordinate values of these locations are the same as shown in Figure 5.1. Facial images were captured when stimuli were presented to the user on these locations. Therefore, it may be assumed that the x and y coordinates of the corresponding centers of the pupils for these sets of stimulus locations should also be very similar in genuine attempts. This should result in a very small variance in the observed coordinate values for these sets

of collinear points compared to those obtained for fake attempts. The collinearity feature vector is therefore a set of variances of face landmark coordinates extracted from multiple sets of collinear challenges/targets.

5.3 Gaze Collinearity Features

A set of points lying on a straight line is referred to here as a collinear set of points and this property of this set of points is hereby referred to as collinearity. Collinearity features are, therefore, extracted from sets of images captured when the stimulus is on a given line. In the investigations reported here, only horizontal or vertical collinearity cases were studied.

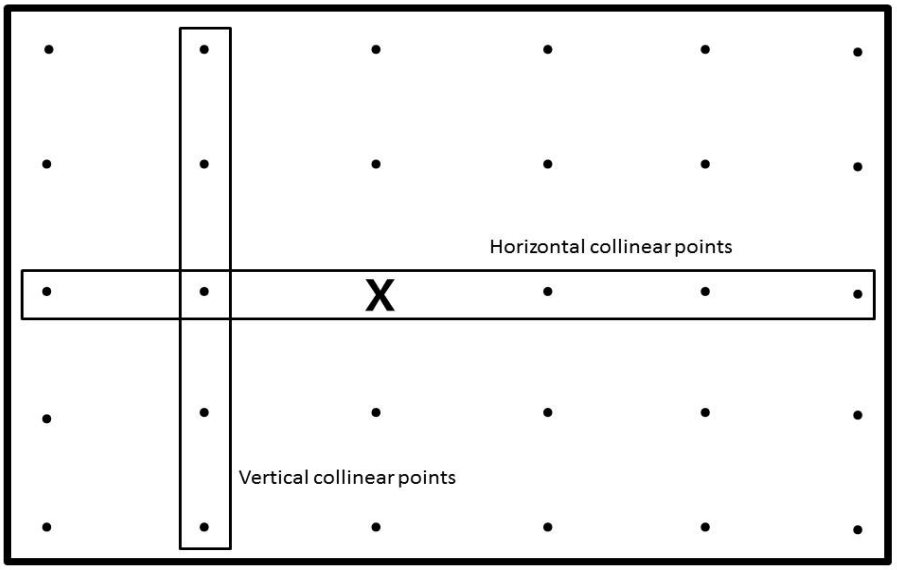

Figure 5.1 Vertical and Horizontal Collinear set of points

Let S_l be a collinear subset of P, where the stimuli are horizontally aligned. $S_l \subseteq C, \quad l = 1, \ldots, L$ where L is the number of horizontally aligned sets of stimulus locations. For $(x, y) \in S_l, y = a_l$ where a_l is constant. Let R be the set

of landmark locations in the captured images.

$$\begin{aligned} R &= \{r_{p1}, r_{p2}, \ldots, r_{pi}, \ldots, r_{pM}\} \\ \text{where, } r_{pi} &= \{(u_{ik}, v_{ik})\} \quad 1 \leq i \leq M, 1 \leq k \leq K \end{aligned} \tag{5.1}$$

and (u, v) are the pixel positions in the image coordinate system for a given landmark k (e.g. corner of the left eye) and K is the total number of such landmarks. Individual subjects moved their eyes and heads by different amounts in response to the movement of the stimulus. They may also be sitting in different positions relative to the screen and camera in each session. So in order to remove these user and session dependent factors in estimating gaze features the data was normalized. The spatial coordinates of the landmarks for each session were normalized using the Min-Max normalization technique [134] prior to feature extraction. The (u, v) coordinates used in this study refer to these normalized values.

For each S_{lk} there is a corresponding subset of R. Let this be denoted by T_{lk}.

$$T_{lk} \subseteq R, l = 1, ..., L \tag{5.2}$$

for any given landmark k. Let $v_{ik} = f(u_{ik})$ denote the trajectory of the facial landmark in response to the challenge. Since the trajectory of the challenge S_l is horizontal, a horizontal response can be assumed and this may be approximated by the equation of a horizontal line.

$$\hat{v}_k = b_k \quad \text{where } b_k \text{ is a constant} \tag{5.3}$$

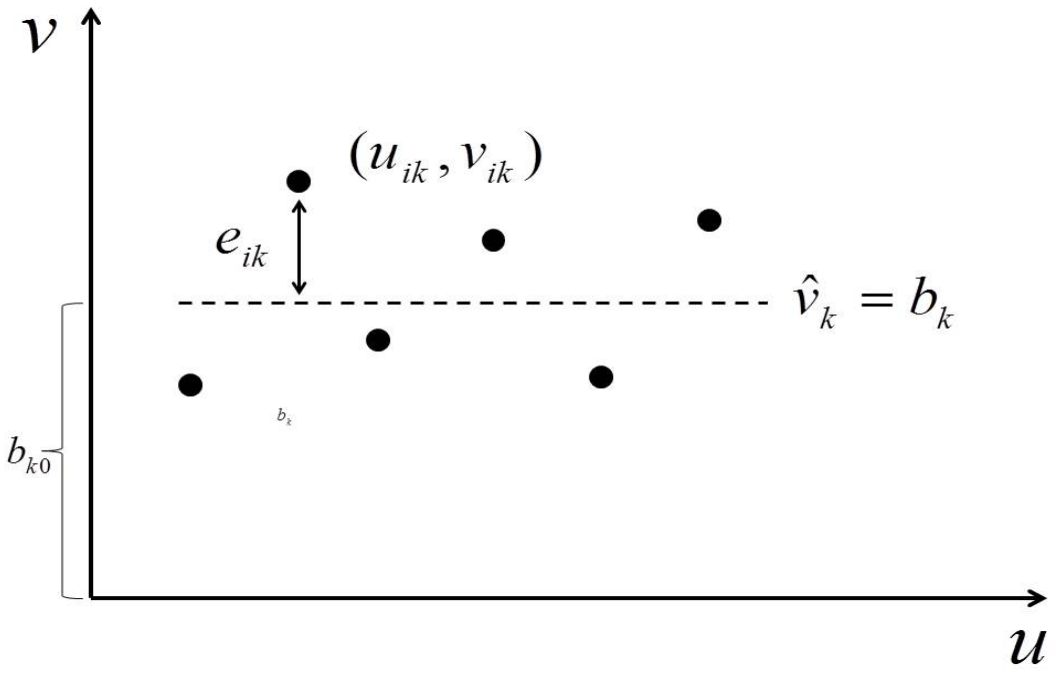

Figure 5.2 Observed locations (•) and expected locus of the landmark positions (–)

The particular value of b_k depends on the system setup. Let, e_{ik} denote the deviation between the estimated $\hat{v}_{ik}$ and observed v_{ik} (see Fig. 5.2), i.e.,

$$e_{ik} = v_{ik} - \hat{v}_{ik} \tag{5.4}$$

For simple horizontal collinearity, $\hat{v}_{ik}$ is calculated as the mean of the observed v_{ik}. So, the mean square error (MSE) for T_{lk} will be

$$E_{lk} = \frac{1}{N}\sum\nolimits_i e_{ik}^2 = \frac{1}{N}\sum\nolimits_i (v_{ik} - \hat{v}_{ik})^2 \tag{5.5}$$

where N is the cardinality of T_{lk}. A similar expression can be derived when the challenge is vertically aligned. As there are multiple face landmarks as well as several stimulus challenge trajectories, a feature vector, F_{colin}, can be constructed from the concatenation of these MSE values (and optionally other feature values)

and used for liveness detection.

$$F_{colin} = [E_{11}, E_{12}, \ldots, E_{1K}, E_{21}, \ldots, E_{ik}, \ldots, E_{LK}] \tag{5.6}$$

A generalised form of the expression for collinearity feature along any straight line is derived here. The collinearity feature provided above, derived for horizontal stimulus loci, may be generalised to include any linear trajectory. Let S_{lk} be a collinear subset of C, where the stimuli are linear. $S_{lk} \subseteq C, \quad l = 1, \ldots, L$ where L is the number of linear sets of stimulus locations. For $(x, y) \in S_{lk}, y = a_{l1}x + a_{l0}$ where a_{l1} is constant.

Let R be the set of landmark locations in the captured images.

For each S_{lk} there is a corresponding subset in R. Let this be denoted by T_{lk}

$$T_{lk} \subseteq R, l = 1, \ldots, ..., L \tag{5.7}$$

for any given facial landmark k, and let $v_{ik} = f(u_{ik})$ denote the trajectory of the landmark in response to the challenge. Since the trajectory of the challenge S_l is linear, a linear response can be assumed and this can be approximated by the equation of a line

$$\hat{v}_k = b_{k1}u_k + b_{k0} \quad \text{where } b_{k1}, b_{k0} \text{ are constants.} \tag{5.8}$$

b_{k1} should be the same as a_{l1} (the slope of the challenge trajectory) whereas b_{k0} depends on the system setup, user interaction, etc.

Let, e_{lk} denote the deviation between the estimated $\hat{v}_{ik}$ and observed v_{ik} (see Fig. 5.3), i.e.,

$$e_{ik} = v_{ik} - \hat{v}_{ik} \tag{5.9}$$

So, the mean square error (MSE) for T_{lk} will be

$$E_{lk} = \frac{1}{N}\sum\nolimits_i e_{ik}^2 = \frac{1}{N}\sum\nolimits_i (v_{ik} - \hat{v}_{ik})^2 \tag{5.10}$$

where N is the cardinality of T_{lk}

By substituting eq. 5.8 in eq. 5.10 and replacing with b_{k1} with a_{l1}

$$\begin{aligned} E_{lk} =& \frac{1}{N}\sum\nolimits_i (v_{ik} - (a_{l1}u_{ik} + b_{k0}))^2 \\ =& \frac{1}{N}(\sum\nolimits_i (v_{ik}^2 + a_{l1}^2\sum\nolimits_i u_{ik}^2 + b_{k0}^2 N \\ -2a_{l1}&\sum\nolimits_i v_{ik}u_{ik} + 2a_{l1}b_{k0}\sum\nolimits_i u_{ik} - 2b_{k0}\sum\nolimits_i v_{ik}) \end{aligned} \tag{5.11}$$

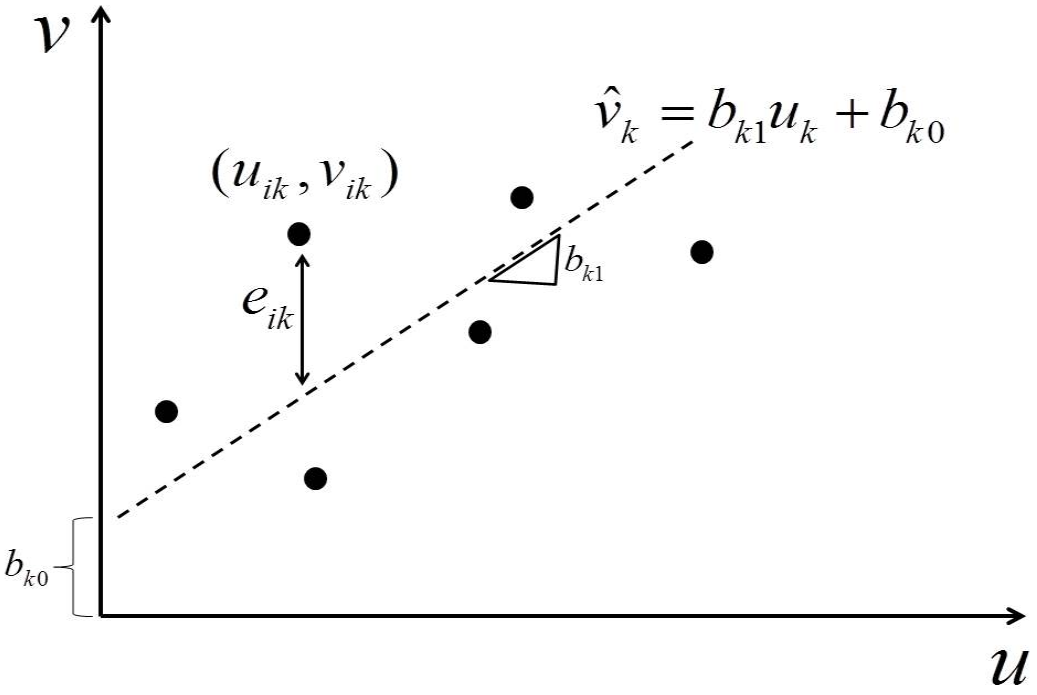

Figure 5.3 Observed locations (•) and expected locus of the landmark positions (–)

Here, b_{k0} should be such chosen that E_{lk} is a minimum. Hence

$$\frac{\partial E_{lk}}{\partial b_{k0}} = 0$$
$$\implies 2b_{k0} + \frac{2a_{l1}}{N}\sum_i u_{ik} - \frac{2}{N}\sum_i v_{ik} = 0$$

$$\implies b_{k0} = \frac{\sum_i v_{ik} - a_{l1}\sum_i u_{ik}}{N} \tag{5.12}$$

Eq. 5.12 can be used to calculate E_{lk}. As there can be multiple face landmarks as well as several distinct linear challenge trajectories, a feature vector F_{colin} can be constructed from the concatenation of these values and used for liveness detection.

$$F_{colin} = [E_{11}, E_{12}, \ldots, E_{1K}, E_{21}, \ldots, E_{lk}, \ldots, E_{LK}] \tag{5.13}$$

More complex and nonlinear stimulus trajectories may also be used and features can be obtained using a similar approach.

Other features can be added to the proposed global feature which may further enhance the distinction between genuine and fake attempts.

$$F = [F_{colin}, F_{coloc}, F_{other}, \ldots]. \tag{5.14}$$

This formulation is similar to the least square regression method [135] but in this case the slope of the best fit line is defined by the challenge and not by the data points.

5.4 Experimental Results

Several sets of experiments were carried out to verify the performance of the proposed features in distinguishing genuine attempts from fake attempts. Initially, the effectiveness of the proposed collinearity features was investigated with a small quantity of data to see if the features were worth investigation in more detail. Then another set of the experiments were carried out to investigate the nature of the challenge. The aim of these experiment was to optimise the challenge.

Preliminary results of this investigation are discussed in Section 5.4.1 of this chapter. For subsequent investigation we collected more data from the test subjects for photo attacks. Also data was collected for photo mask and video replay presentation attacks. The detail of the database is discussed in Chapter 3. In the last set of experiments the collinearity features were further tested with the big database.

5.4.1 Preliminary Experimental Results

The initial experiments were carried out to estimate the potential of the proposed collinearity feature in distinguishing between genuine and fake attempts. Data was collected from 5 subjects in 3 sessions. Each volunteer performed 3 genuine and 3 fake attempts. This small database was composed of 15 genuine and 15 fake photo presentation attack attempts. Each attempt acquired 358 frames of facial images, and the resolution of the images is 35×288 pixels. A setup similar to that shown in Figure 3.6 was used.

For the observations reported here, only the centres of the pupils in the captured frames were used. For the vertically collinear points, the x-coordinate values of

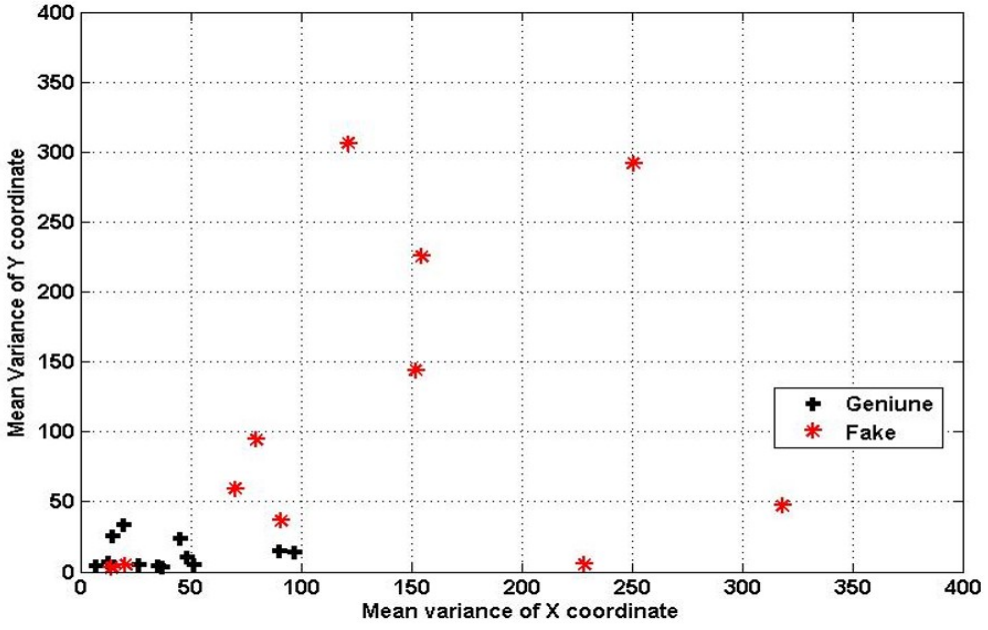

Figure 5.4 Feature distribution with outlier inclusion

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

Figure 5.4 illustrates the distribution of the genuine and impostor attempts in the feature space. As anticipated, genuine attempts showed much smaller variances compared to those of the fake attempts in most of the cases.

If the impostor holds the photo or mask still (i.e., makes no attempt to respond to the challenge) this will produce very small changes in the measured landmarks location. However, this can also happen when a genuine user is non-cooperative

or non responsive. The opposite extreme can also occur, when the genuine user's response to the challenge involves extreme head and eye movements. Some may produce large variances in response to the challenge. The maximum expected gaze deviation from the normal to the screen is approximately 15 degrees for this experimental setup. Such extreme behaviour can be treated as suspect and may be flagged as fake attempts. Figure 5.5 shows the distribution of genuine and fake attempt features when such outliers are excluded. It is evident that in this case the separation between genuine and impostor features has become more prominent through removing the outliers.

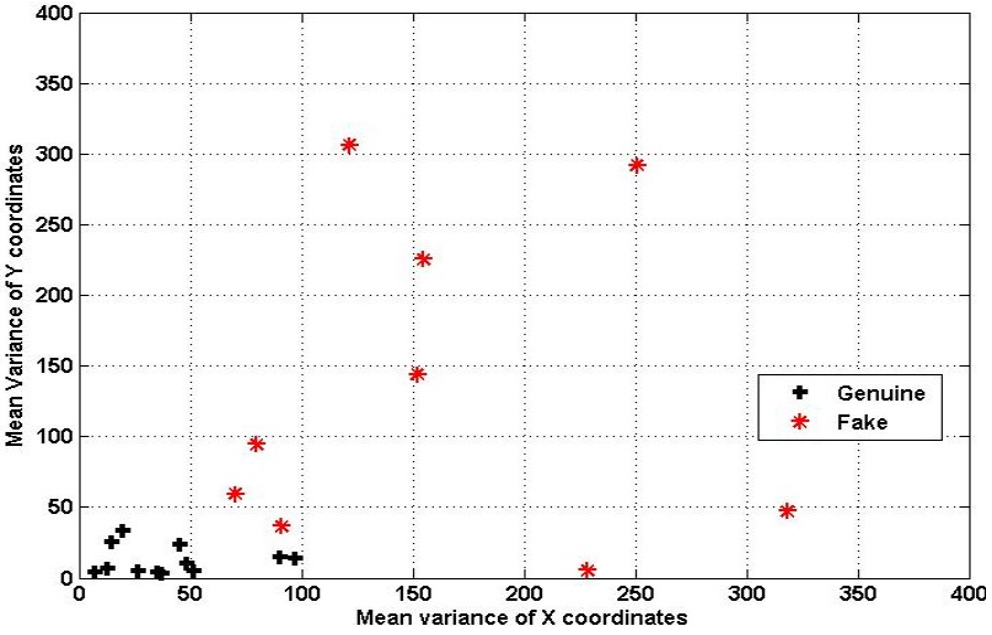

Figure 5.5 Feature distribution without outliers

The criteria used for the identification of outliers are shown in Table 5.1. Rule 1 excludes the cases where the net variance in x and y coordinates are smaller than certain thresholds whereas Rule 2 filters those with very large net variances. The actual thresholds were determined empirically.

The liveness detection scheme proposed here is a two phase process. In the first phase, the scheme applies the outliers' rules and if outliers are detected, identifies

Table 5.1 Outlier filter criteria

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

the attempt as inconclusive and more data would be needed to establish liveness of the user. In the second phase, a linear discriminant classifier [136] using the collinearity features is employed to decide the liveness of the user. Table 5.2 shows the accuracy of the proposed method both with and without the outlier detection phase.

Table 5.2 Performance of the proposed method

	FAR	FNR
Outliers not excluded	13.3%	0.0%
Outliers removed	0.0%	0.0%

The results show that in both configurations (with or without the outlier detection phase) the FNR is 0%. Exclusion of the outliers reduced the FAR to 0% too.

These experiments indicate the potential viability of this approach. Next the experiments were expanded to include more users and attempts and also explore more attack scenarios and more sophisticated spoofing attacks. The impact on performance of reducing, the challenge duration was also explored to cut down the number of vertical and horizontal collinear points set used for feature extraction.

5.4.2 Directional Sensitivity in Gaze Collinearity

Here we explored the sensitivity of the proposed system to different stimulus alignments orientations. The aim was to establish whether there was any such sensitivity and if so to explore how this may be exploited for improving the design of the stimulus.

In this investigation, coordinates of the landmarks were used to analyse and compare the performance of the x and y coordinates of the features. The x, y or both xy coordinates from left and right eyes were passed to independent k-nearest neighbor (k-NN) classifiers [137]. In this implementation, k has been optimized to minimize the leave-one-out error in the training data. The normalized scores (based on the posterior probabilities of class membership) from these classifiers were combined to produce a single score, using various rule-based fusion schemes [71].

A small amount of data was collected to investigate the performance of the proposed scheme. In total 8 subjects participated in the data collection phase. The data was captured in 3 sessions. A total of 26 fake and 26 live attempts were captured. The user presented a high quality colour photo of a target user in front of the camera for a spoofing attempt.

Collinearity feature vectors were extracted from the facial images captured when the stimulus locations appeared along horizontal and vertical lines. In the first phase, only the x coordinates from both eyes (left and right) were used for the classification of fake and live attempts. This was done to investigate the x coordinates of the eyes on their own. Figure 5.6 illustrates where the score from the each eye x coordinates was combined to get single score for decision.

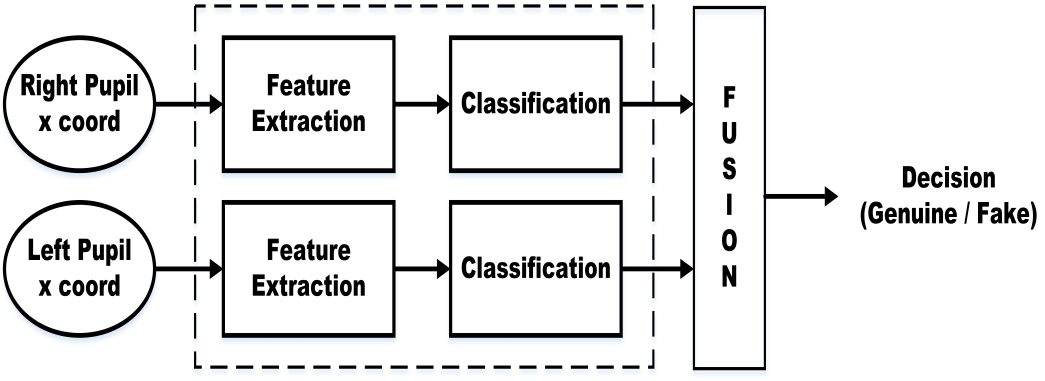

Figure 5.6 Score fusion using x coordinates of the left and right eye

In this experiments the y coordinates from both eyes were used for the classification of fake and live attempts. Figure 5.7 shows the fusion of the y coordinates of both eyes scores.

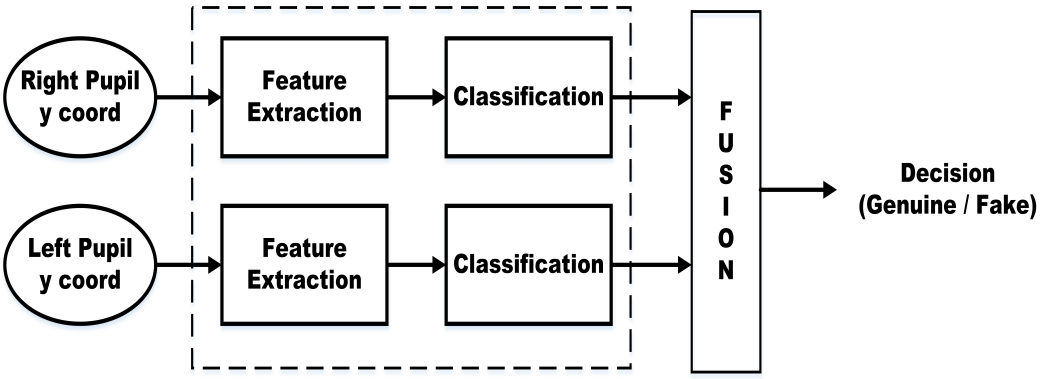

Figure 5.7 Score fusion using y coordinates of the left and right eye

The last set of experiments were the combination of the above two cases. In this phase x coordinates and y coordinates from both eyes were used, using fusion rules as shown in Figure 5.8. In this figure x and y coordinates from the left and right eye were passed to a separate classifier and their scores were fused together using sum, product and majority-vote rules [134].

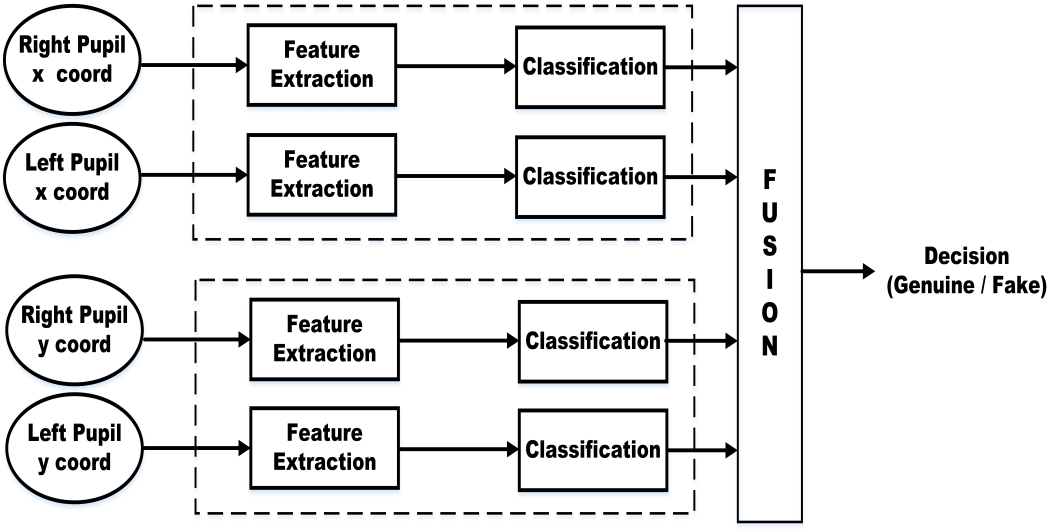

Figure 5.8 Score fusion using x and y coordinates of the left and right eye

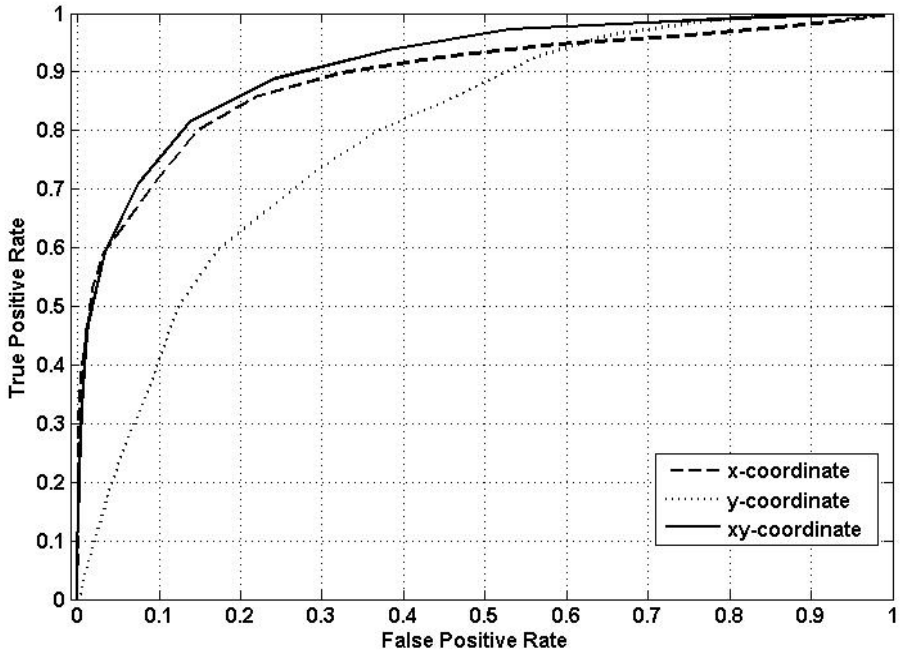

Figure 5.9 ROC curves showing the performances of the three proposed schemes

Figure 5.9 shows the ROC curves, which shows the performance of the feature extracted along x coordinated, y coordinate and combination both xy coordinates form both eyes. The performance of the system was found to be lower when using only the y-coordinate features. Although using x coordinate features alone improved the performance, using both x and y coordinates performed best. Using both x and y coordinates of both eyes, the system performance reached 75% TPR (at 10% FPR). Using only the x coordinate of both eyes the system achieved 72% TPR (at 10% FPR). The scores were combined using the sum rule score fusion.

At the lower FPR (< 0.10), the ROC curve of the x coordinate features is similar to the ROC curve of xy coordinate features. The ROC curves for these features rise rapidly with increasing FAR and show a much better performance than that of the y-coordinate features alone. In conclusion, these results suggest that the x coordinate features are better in comparison to the y coordinate features. The improvement is very small in the performance between using features based

on the x coordinate only, and the x and y coordinates together. The system performance may be improved more efficiently by increasing the number or range of the x coordinate features.

Table 5.3 Comparison of performance of Different Feature

TPR				
	Ruled base fusion	**@FPR = 0.02**	**@FPR = 0.05**	**@FPR = 0.10**
x-coordinate from both eyes	Product	0.53	0.62	0.70
y-coordinate from both eyes	Product	0.07	0.19	0.35
x-y-coordinate from both eyes	Product	0.50	0.63	0.72
x-coordinate from both eyes	Sum	0.52	0.63	0.72
y-coordinate from both eyes	Sum	0.10	0.23	0.41
x-y-coordinate from both eyes	sum	0.54	0.64	0.75
x-coordinate from both eyes	Majority vote	0.21	0.53	0.64
y-coordinate from both eyes	Majority vote	0.14	0.16	0.31
x-y-coordinate from both eyes	Majority vote	0.06	0.52	0.72

Table 5.3 shows the performance of the system using various feature subsets, and fusion schemes. The x coordinate features gave better performance compared to the y-coordinate features. But combining the x and y coordinate improves the performance slightly. The sum score fusion rule gave the most promising result.

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

5.4.3 Extended Experimental Results

Once the sensitivity of the different stimulus alignments was investigated, the design of the challenge was decided. It was designed in such a way that in total,

6 vertically collinear and 5 horizontal collinear point sets were extracted for this data set. The aim of this part of the experiment was to validate the preliminary results with more data and also to investigate other types of sophisticated spoofing attacks.

Error rates were calculated for a range of system operating parameters and are reported here. True Positive Rates at a set of predefined FPR values were obtained and used for comparison. The detection performance of the system for mask attacks was better than that achieved for photo attacks. At 10% FPR, video replay detection TPR is about 99%, mask attack detection TPR is 71% and photo attack detection is about 55%.

In order to establish the tradeoff between the feature dimensionality and liveness detection, the accuracy of the system was determined as the feature dimensions were steadily reduced. A forward feature selection method [131] was used for this purpose.

The feature selection method was run several times, choosing random sets of data for training and testing for each run. The results of these runs were combined to rank the features. It can be seen in Figure 5.10 that the lowest total was observed when the feature dimension was reduced to around 3 for collinearity feature.

Figure 5.10 presents total error rates as a function of the number of features selected to find optimum sets of collinearity features. In this experiment the photo and mask attack modalities were combined and treated as presentation attack against genuine attempts. The combination of these attack modalities allows the establishment of a single optimal feature set that can be used for all of these major spoofing challenges. Video attack data was excluded from this feature ranking exercise as the system already performed very well in detecting video spoofing attacks.

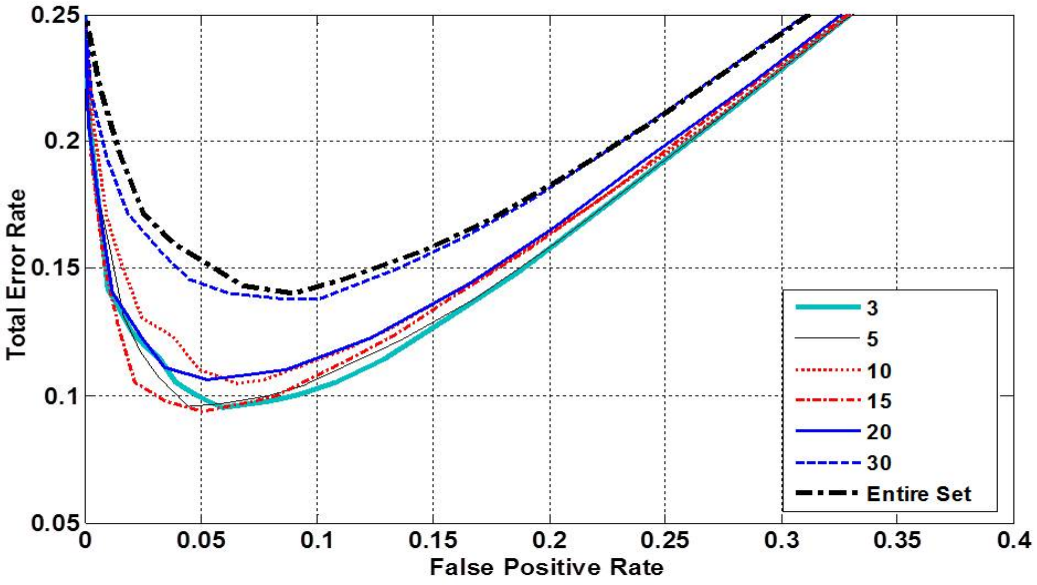

Figure 5.10 Variation in accuracy with FPR for Collinearity Feature

The collinearity feature performance for photo, mask and video spoofing attack using this optimum feature set is illustrated in Figure 5.11.

It is evident that the system displayed a near-perfect performance in detecting video attacks. Here again the replay video was captured from the genuine person in response to the challenge. At the time of impostor attack, the user replayed the video in response to the challenge. The video was not responding to the corresponding stimulus locations as the challenge locations pattern was different from the one at which video was recorded. Hence no collinearity can be seen in the video response. Therefore, video attack resulted in better detection performance. The detection performance of the system for mask attacks was better than that achieved for photo attacks. At 10% FPR, video replay detection TPR is about 100%, mask attack detection TPR is 79% and photo attack detection is about 69%. Table 5.4 presents TPR values at 10% FPR for the three spoofing attack detection scenarios using the collinearity feature. It shows the error rates when

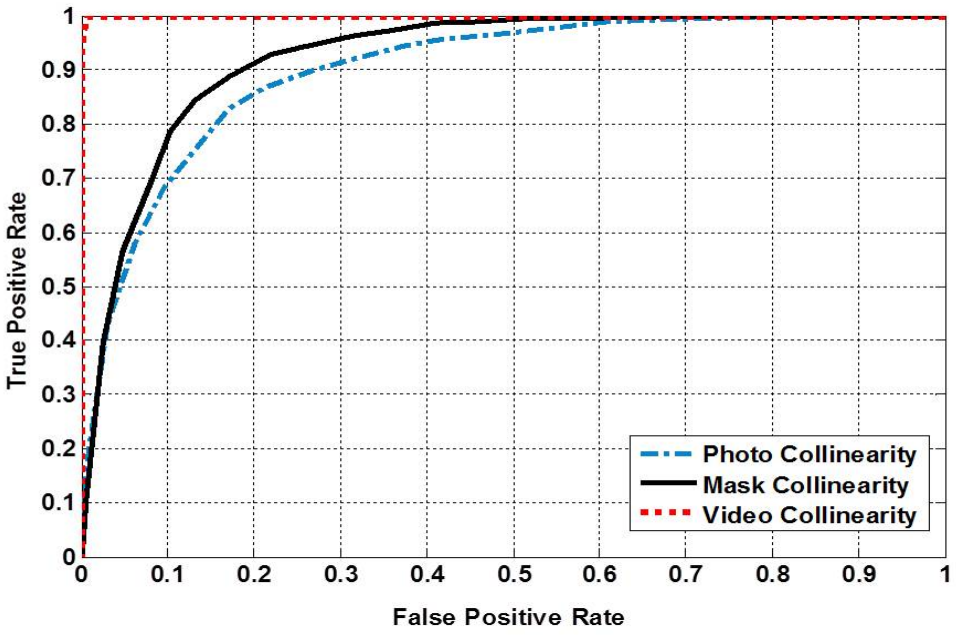

Figure 5.11 ROC curve of the proposed system using optimum feature set schemes

collinearity features are implemented in a multi-classifier configuration. Score fusion has been used in these cases.

Table 5.4 TPR at FPR = 0.10 using optimum feature sets

Feature Sets	Photo	Mask	Vidoe Replay
Collinearity	0.69	0.79	1.00

5.5 Fusion both Collinearity and Colocation

Next the effectiveness of the two proposed features, collinearity and colocation, in combination with each other in detecting liveness were investigated. Several schemes were set up to explore the gain in accuracy achieved by combining features extracted from both the eyes in a multi-classifier configuration incorporating rule-based fusion [71]. Several classification schemes were investigated and the k-NN classifier was found to produce the best performance. These classifiers

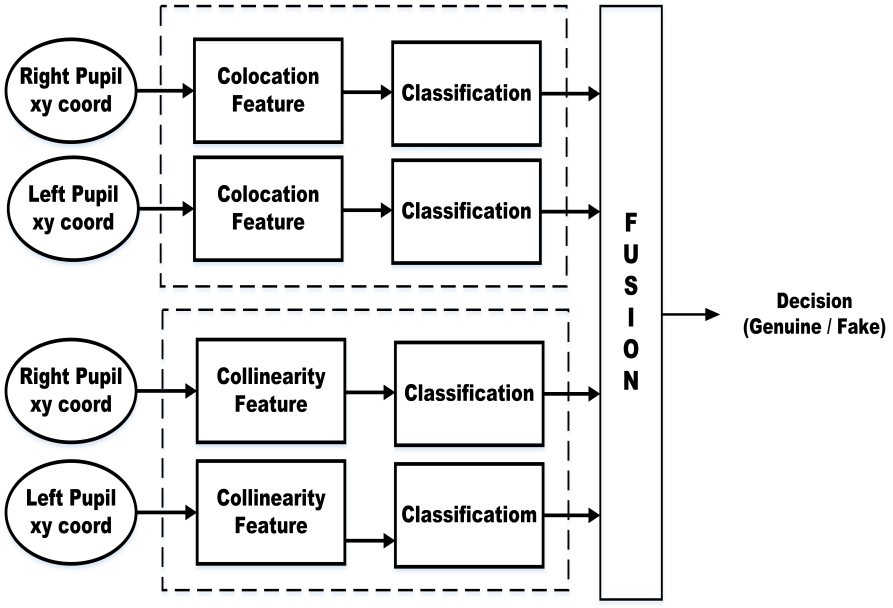

Figure 5.12 Collinearity and colocation fusion

(k-NN) were used to obtain the individual classification scores for each eye and each feature set. The a posteriori probabilities from the separate classifiers were combined at the fusion stage for liveness detection. The scheme is illustrated in Figure 5.12. The k-NN classifier from the prtools package used for this work automatically determined an optimum k-value for each experiment. Each experiment was run 400 times with random sets of data for training and testing, resulting in a dierent optimum k values for each run. These optimal k values were combined to obtain an overall mean k value to be used for operational systems. In the case of collinearity features this mean optimal k value was 7.

Figure 5.13 shows the Receiver Operating Characteristic (ROC) curves for combined collinearity and colocation features using the proposed fusion scheme. It is evident that the system displayed a near-perfect performance in the case of video attacks detection.

The performance of the system for mask attack detection was better than that achieved for photo attacks detection. At 10% FPR, video replay detection TPR is about 99%, mask attack detection TPR is 69% and photo attack detection is

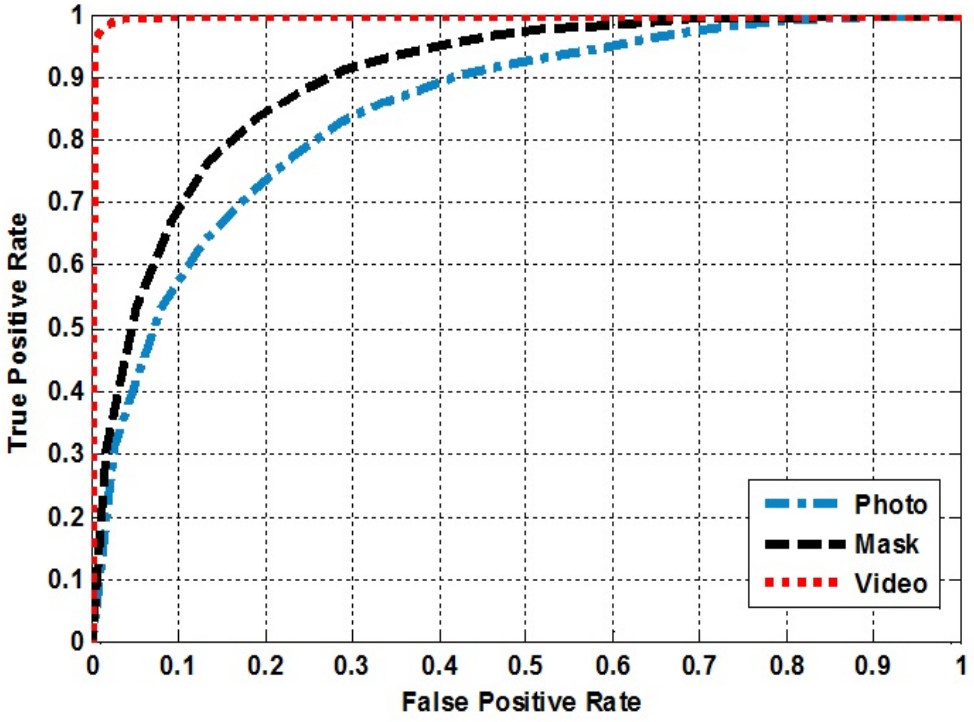

Figure 5.13 ROC curve using entire feature vector

about 58%.

Table 5.5 TPR at FPR = 0.10 using the entire feature set

Feature Sets	Photo	Mask	Video Replay
Collinearity	0.55	0.71	0.99
Colocation	0.43	0.25	0.83
Collinearity and Colocation	0.58	0.69	0.99

Table 5.5 presents TPR values at 10% FPR for the three spoofing attack detection scenarios. It shows the error rates when collinearity and colocation features are implemented on their own, as well as in multi-classifier configuration. It is evident that in all the schemes implemented, the video replay attack detection outperformed the other two types of attacks. The collinearity features were superior to the colocation features. Using score fusion, the TPR for hand held photo attack detection improved, the photo mask detection performance was slightly decreased, and video replay attack detection remained the same. Product-rule based score fusion has been used in these cases.

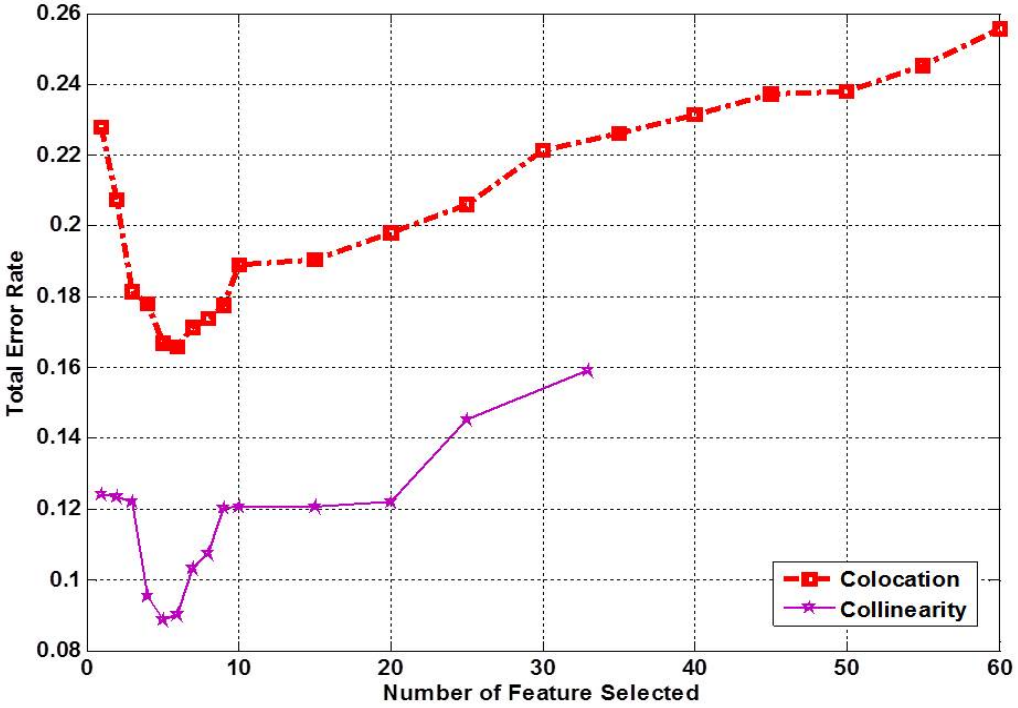

Figure 5.14 Variation in accuracy with feature dimension

In order to establish the tradeoff between the feature dimensionality and liveness detection accuracy of the system was determined as the feature dimensions were steadily reduced. A forward feature selection method [131] was used for this purpose. Figure 5.14 presents total error rates as a function of the number of features selected to find optimum feature sets for collinearity and colocation features.

The collinearity and colocation feature performance for photo, mask and video spoofing attack using this optimum feature set is illustrated in Figure 5.15. Video replay attack detection gives best performance while the photo mask attack detection ranks second in performance followed by hand held photo attack detection using collinearity feature.

At 10% FPR, TPR of 69%, 79% and 100% are achieved for photo, mask and video replay attacks respectively. The colocation feature performance is much weaker compared to the collinearity performance. At 10% FPR about 70%, 38%

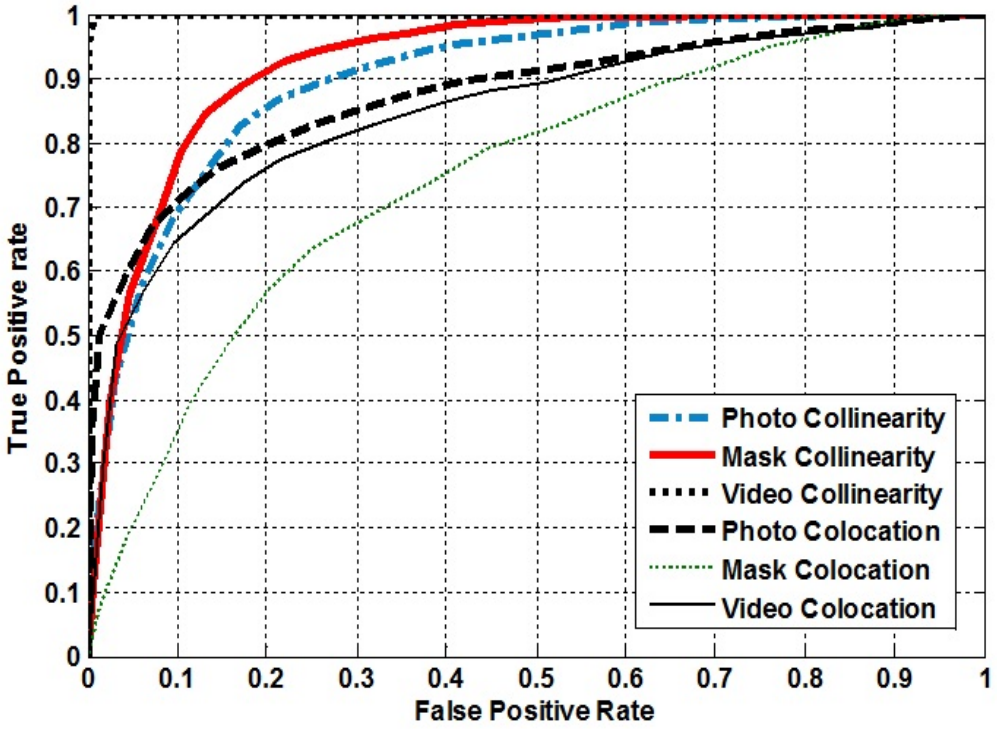

Figure 5.15 ROC curve of the proposed system using optimum feature set

and 65%TPR are achieved for photo, mask and video replay detection respectively.

Figure 5.16 shows the ROC curves for the optimum feature sets for fusion of collinearity and colocation information. The performance of the system was found to be worse when collinearity or colocation features were used separately for most scenarios as can be seen in comparison with Figure 5.15. At 10% FPR, video replay performance is 100% and photo attack TPR increased to about 90%. The mask attack detection performance marginally increased after fusion and is lower compared to the video and photo spoof detection performance.

Table 5.6 summarizes some of the key results from Figure 5.15 and Figure 5.16 and presents results for each feature type separately along with the results for the combined collinearity and colocation features, which gave a better performance. For video replay attack detection, the proposed combined system is error free

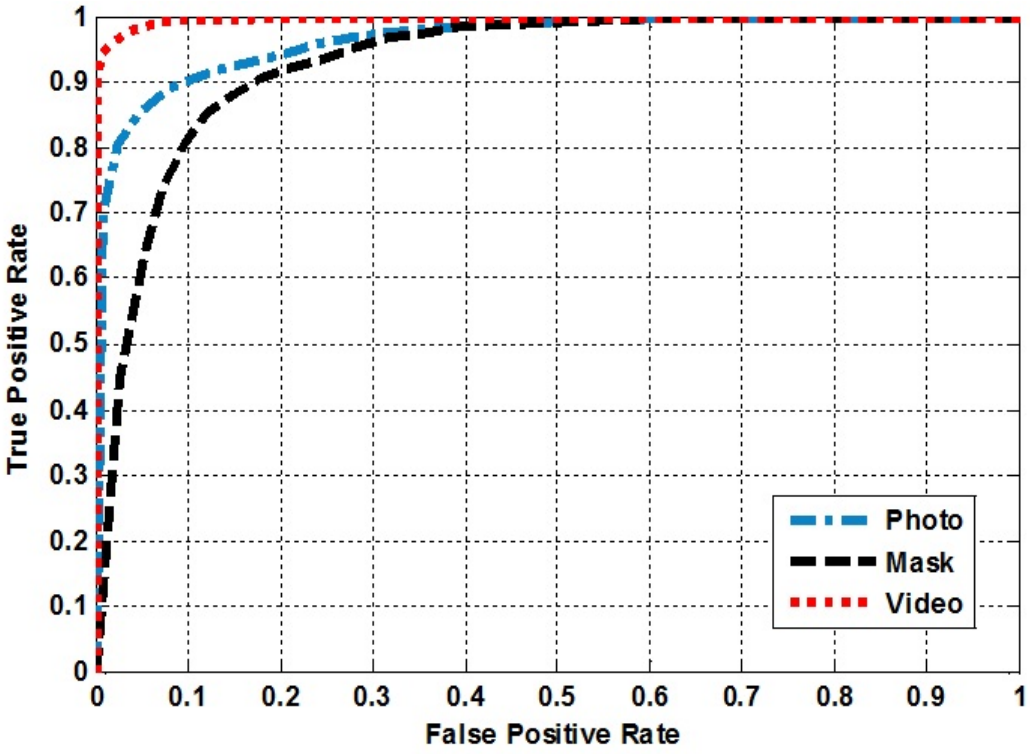

Figure 5.16 Score fusion performance using optimum feature sets

(for this data set). The performances of collinearity, colocation features and their fusion for the entire feature set and optimum feature sets are shown in Table 5.5 and Table 5.6 respectively for comparison. Using the optimum features, the TPRs of combined collinearity and colocation feature increased by 32% and 15% for photo and mask attack detection respectively. For video replay attack the proposed scheme is nearly perfect, and no significant improvements due to fusion were noticed.

Table 5.6 TPR at FPR = 0.10 using optimum feature sets

Feature Sets	Photo	Mask	Vidoe Replay
Collinearity	0.69	0.79	1.00
Colocation	0.70	0.38	0.65
Collinearity and Colocation	0.90	0.81	1.00

In this experiment the photo and mask attack modalities were combined and treated together against genuine attempts. The combination of these attack

modalities allows the establishment of a single optimal feature set that can be used for all of these major spoofing challenges. Video attack data was excluded from this feature ranking exercise as the system already performs very well in detecting video spoofing attacks. The feature selection method was run several times, choosing random sets of data for training and testing for every run. The results of these runs were combined to rank the features. It can be seen that the lowest total error rate was observed when the feature dimension was reduced to around 5 for both feature types.

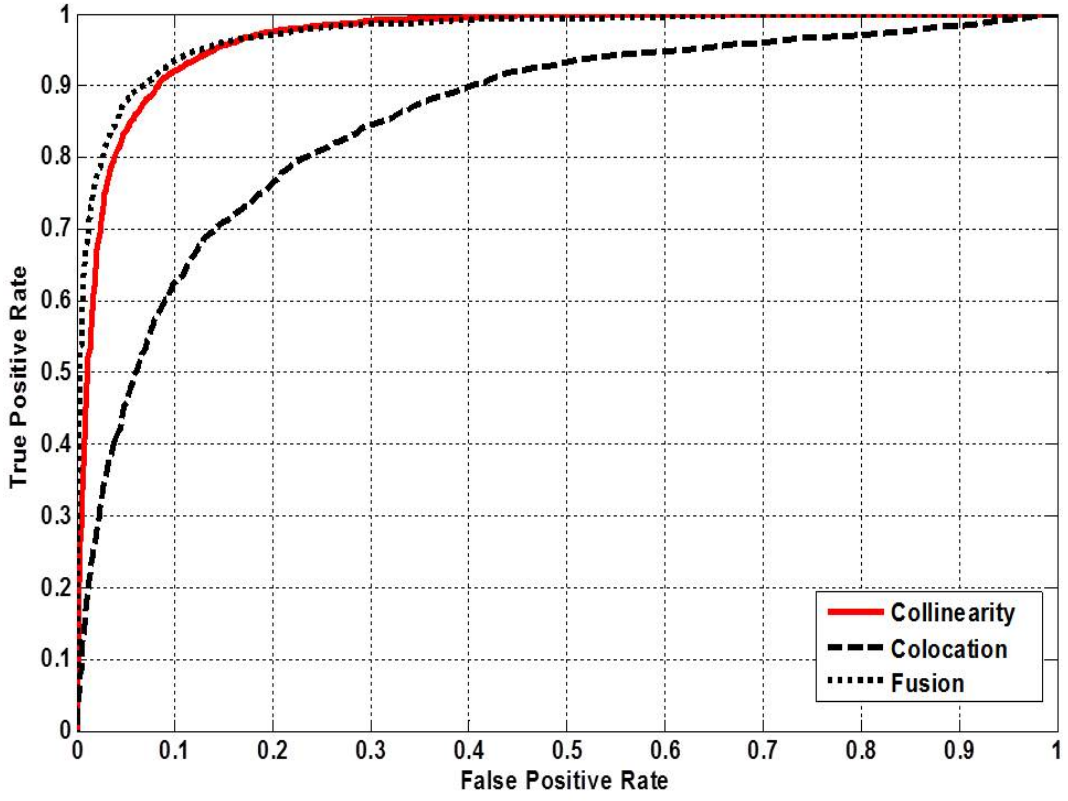

Figure 5.17 Genuine vs fake (photo, mask, video) performance using optimum feature sets

In the following experiments all attacks types were treated as one class (fake) rather than as three separate attack scenarios. Figure 5.17 illustrates the ROC curves for real and fake attempts. Combination of collinearity and colocation

data again gave better performance. The performance of collinearity features is very close to that achieved by the performance of the combined features. At 10% FPR, TPR of about 91%, 63% and 93%, were achieved for collinearity, colocation and their fusion respectively. The colocation feature performance is much weaker compared to the performance of collinearity and fusion based schemes.

5.6 Conclusion

This chapter presents a face liveness detection technique that may be used for a range of biometric applications. The proposed system is a challenge-response approach using a visual stimulus to measure the gaze of the user for the purpose of establishing the presence of photographic, mask and video spoofing attacks.

Collinearity features are proposed and used to provide a measure to discriminate between live and fake attempts. Preliminary experiments were carried out to estimate the potential of the proposed novel features. The first set of experimental results were promising and the challenge was redesigned to collect further data to further support the potential of the novel feature. Before designing the final challenge, we investigated the direction sensitivity of the challenge to enable us to extract more informative features.

We analysed the performance of the system using x and y coordinates of the pupil centre. The features based on x coordinates of the eye centre locations were found to be more effective for liveness detection. Given that the acquisition time will have to be bounded, this implies that the set of challenge points should be chosen to have more vertically collinear sets of points.

The main contribution of this chapter has been,

- Another gaze-based novel feature

- Investigate score fusion of both eyes and score fusion of colocation and collinearity features.

- Investigate three types of presentation attack which were photo, mask and video replay attacks.

In the next chapter another novel feature gaze-based homography features are investigated.

CHAPTER 6

Gaze Homography-based Feature

6.1 Introduction

This chapter presents a third novel gazed-based feature named "gaze homography". Here homography is used to capture the relation between screen coordinates and points on the image sensor. It may be assumed that there is a relatively stable relationship between screen coordinates of stimulus to which gaze is directed and the sensor coordinates of the eye landmarks for a genuine attempt. The absence of such stability can be exploited to detect presentation attacks using artifacts such as printed photographs, photo masks and video replay.

As before, in this approach the user's gaze is directed to random positions on the display and the facial image of the user was captured at each position of the challenge to extract the pupil coordinates of the user. These pairs of coordinates were used to estimate the transformation matrix which relate the screen coordinates to the pupil coordinates. This chapter also covers the fusion

of all the features studied in this research to explore if the performance of the liveness detection system can be further improved.

The organization of this chapter is as follows: A mathematical framework for the use of homography is presented in Section 6.2. All the necessary steps to calculate the homography relationship are also explored in this section. The gaze-based homograph features are derived in Section 6.3. Two systems using these features are proposed in Section 6.4. Section 6.5 gives a summary of the experimental results. Section 6.6 presents the combining of collinearity, colocation and homography features with detailed experimental results using score fusion. Finally Section 6.7 provides the chapter summary and concluding remarks.

6.2 Homography and its Estimation

A function that maps one vector space into another is often achieved by using a transformation matrix. If the vector addition and scalar multiplication are preserved, a mapping is considered to be a linear transformation. There are two types of linear transformations called projective (homography) and affine transformations. Affine transformation is a particular case of the projective transformation. In projective transformations angles, distance, ratios of distances are not preserved [138]. However, the straight lines are preserved projective transformation [138]. While in affine transformation lines map to lines, parallel lines remain parallel lines, ratio of lengths of two parallel segments remains same etc. [139].

A 2D point (x, y) in an image can be represented as a 3D vector $x = (x_1; x_2; x_3)$ where $x = \frac{x_1}{x_3}$ and $y = \frac{x_2}{x_3}$. This is called the homogeneous representation of a point and it lies on the projective plane P^2 [140]. Consider image coordinates (x, y) plane and screen coordinates (u, v) plane are related by homography. To

estimate the homography matrix, the relationship between two corresponding points can be written as in Eq. 6.1.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{6.1}$$

In Eq. 6.1 the transformation matrix is $\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$ where $\begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$ is a rotation matrix. This matrix defines the kind of transformation that will be performed: scaling, rotation, and so on.

Similarly $\begin{bmatrix} h_{13} \\ h_{23} \end{bmatrix}$ is the translation vector. It simply moves the points and $\begin{bmatrix} h_{31} & h_{32} \end{bmatrix}$ is the projection vector.

To estimate the homography matrix from Eq. 6.1, the relationship can be re-written as in Eq. 6.2,

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11}u & h_{12}v & h_{13} \\ h_{21}u & h_{22}v & h_{23} \\ h_{31}u & h_{32}v & h_{33} \end{bmatrix} \tag{6.2}$$

Eq. 6.2 can re-written as in Eq. 6.3,

$$\begin{aligned} x &= h_{11}u + h_{12}v + h_{13} \\ y &= h_{21}u + h_{22}v + h_{23} \\ 1 &= h_{31}u + h_{32}v + h_{33} \end{aligned} \tag{6.3}$$

To represent these coordinates in homogenous coordinates,

$$\begin{aligned} x &= \frac{h_{11}u + h_{12}v + h_{13}}{h_{31}u + h_{32}v + h_{33}} \\ y &= \frac{h_{21}u + h_{22}v + h_{23}}{h_{31}u + h_{32}v + h_{33}} \end{aligned} \tag{6.4}$$

Setting $h_{33} = 1$

$$\begin{aligned} x &= \frac{h_{11}u + h_{12}v + h_{13}}{h_{31}u + h_{32}v + 1} \\ y &= \frac{h_{21}u + h_{22}v + h_{23}}{h_{31}u + h_{32}v + 1} \end{aligned} \tag{6.5}$$

Multiplying through by denominator

$$\begin{aligned} (h_{31}u + h_{32}v + 1)x &= h_{11}u + h_{12}v + h_{13} \\ (h_{31}u + h_{32}v + 1)y &= h_{21}u + h_{22}v + h_{23} \end{aligned} \tag{6.6}$$

Rearranging the above Eq. 6.6,

$$\begin{aligned} h_{11}u + h_{12}v + h_{13} - h_{31}ux - h_{32}vx &= x \\ h_{21}u + h_{22}v + h_{23} - h_{32}uy - h_{32}vy &= y \end{aligned} \tag{6.7}$$

Homography has 8 degrees of freedom so there should be a minimum 4 sets point of (x, y) and corresponding (u, v) to estimate the homography matrix.

Subsituting these 4 sets of coordinates in Eq. 6.7, we obtain Eq. 6.8,

$$\begin{bmatrix} u_1 & v_1 & 1 & 0 & 0 & 0 & -u_1x_1 - v_1y_1 \\ 0 & 0 & 0 & u_1 & v_1 & 1 & -u_1x_1 - v_1y_1 \\ u_2 & v_2 & 1 & 0 & 0 & 0 & -u_2x_2 - v_2y_2 \\ 0 & 0 & 0 & u_2 & v_2 & 1 & -u_2x_2 - v_2y_2 \\ u_3 & v_3 & 1 & 0 & 0 & 0 & -u_3x_3 - v_3y_3 \\ 0 & 0 & 0 & u_3 & v_3 & 1 & -u_3x_3 - v_3y_3 \\ u_4 & v_4 & 1 & 0 & 0 & 0 & -u_4x_4 - v_4y_4 \\ 0 & 0 & 0 & u_4 & v_4 & 1 & -u_4x_4 - v_4y_4 \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ x_3 \\ y_3 \\ x_4 \\ y_4 \end{bmatrix} \tag{6.8}$$

In this way the homomoraphy matrix which relates two planes to each can be estimated.

6.3 Feature based on Gaze Homography

A small shape (stimulus) is presented to the subject on the screen whilst they are seated in front of the computer screen and instructed to follow it as it changes its location on the screen with natural head/eye movements. Let C be the set of challenge locations where the stimulus was presented,

$$\begin{gathered} C = \{c_1, c_2, \ldots, c_d, \ldots, c_D\} \\ \text{where, } c_d = (x, y)\,;\ d = 1, \ldots, D \end{gathered} \tag{6.9}$$

The stimulus can be shown at a location multiple times and let P be the sequence of M such presentations.

$$\begin{gathered} P = \{p_1, p_2, \ldots, p_m, \ldots, p_M\} \\ \text{where, } p_m \in C; m = 1, \ldots, M \end{gathered} \tag{6.10}$$

Let R be the set of landmark locations in the captured images for given larmark.

$$\begin{aligned} R &= \{r_{p1}, r_{p2}, \ldots, r_{pi}, \ldots, r_{pM}\} \\ \text{where, } r_{pi} &= \{(u_i, v_i)\} \quad 1 \leq i \leq M, \end{aligned} \tag{6.11}$$

and (u, v) are the pixel positions in the image coordinate system for a given landmark k (e.g. corner of the left eye) and K is the total number of such landmarks.

Let J_l be a subset of P, at least 4 points

$J_l \subseteq P \quad l = 1, \ldots, L$ where L is the number of set points of stimulus locations.

For each J_l there is a corresponding subset of R. Let this be denoted by O_l. where $O_l \subseteq R, l = 1, ..., L$

To estimate homography matrix, denoted by H_l,

$$\begin{aligned} O_l &= H_l \cdot J_l \\ H_l &= O_l \cdot J_l^{-1} \end{aligned} \tag{6.12}$$

H_l can be estimated using a set of challenge and image coordinates.

$H_l \cdot J_l$ will give the estimated points on the sensor $(\tilde{u}, \tilde{v})$, let that be $\tilde{R}$,

$$\tilde{R} = H_l \cdot P \tag{6.13}$$

$$\begin{aligned} \tilde{R} &= \{\tilde{r}_{p1}, \tilde{r}_{p2}, \ldots, \tilde{r}_{pi}, \ldots, \tilde{r}_{pM}\} \\ \text{where, } \tilde{r}_{pi} &= \{(\tilde{u}_i, \tilde{v}_i)\} \quad 1 \leq i \leq M, \end{aligned} \tag{6.14}$$

To find the Euclidean distance between pupil centre coordinates (u, v) and estimated coordinates $(\tilde{u}, \tilde{v})$ is shown below,

$$\begin{aligned} Euc_l &= \sqrt{(v_i - \tilde{v}_i)^2 + (u_i - \tilde{u}_i)^2} \\ F_{homog} &= [Euc_1, Euc_2, \ldots, Euc_l, \ldots, Euc_L] \end{aligned} \tag{6.15}$$

The Euclidean distance between the pupil centre (R) and the estimated points $(\tilde{R})$ was expected to be small for genuine attempts and larger for impostor attempt. This phenomenon was then exploited to distinguish between the genuine and presentation attacks.

Many other features can be extracted from facial landmarks which can further enhance the proposed method. All these can be combined into a global feature vector,

$$F = [F_{colin}, F_{coloc}, F_{homog}, F_{other}, \ldots]. \tag{6.16}$$

6.4 Proposed Systems

Several sets of experiments were carried out to explore the performance of the gaze-based homography features. The typical arrangement of the evaluation framework is shown in Figure 6.1. Gaze-based homography features were extracted using both eyes and then were passed to separate classifiers for training and testing using this framework. Initially two methods which were based on two types of features were investigated. These methods were called System 1 and System 2 in this study.

System 1 is discussed in detail in Section 6.4.1 and System 2 is explored in detail in Section 6.4.2.

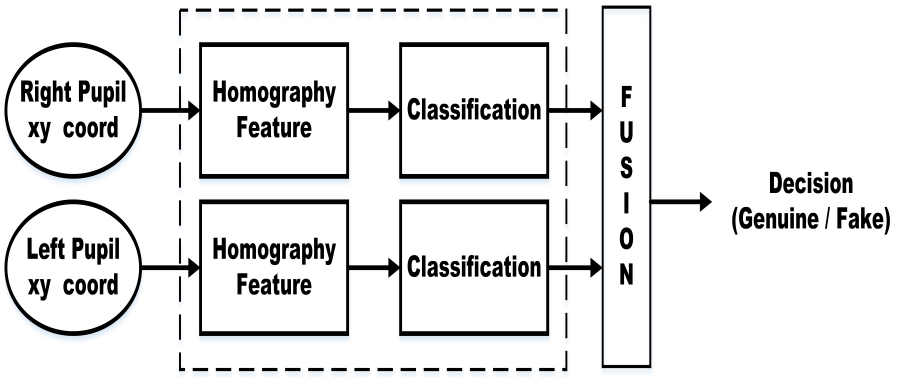

Figure 6.1 Score fusion using feature extracted from left and right eye

6.4.1 System 1

In System 1 several homography matrices were calculated using various combinations of corresponding screen and landmarks (observed) coordinates. Small sets of points were extracted out of the landmarks points along with the corresponding screen coordinates in order to estimate the homography matrix for each set. The stability of these homography matrices was used to discriminate between genuine and fake attempts. The repeatability of the values of these matrices should ideally be exactly the same. Mean and standard deviation for these homography matrices were calculated.

$$\bar{h}_{ij} = \frac{1}{L}\sum_{l=1}^{L} h_{ijl} \tag{6.17}$$

where $1 \leq ij \leq 3$ and h_{ijl} is the element h_{ij} in the homography matrix H_l,

$$\sigma_{ij} = \sqrt{\frac{\sum_{l=1}^{L}(h_{ijl} - \bar{h}_{ij})^2}{L}} \tag{6.18}$$

$$F_{homog} = [\bar{h}_{11}, \sigma_{11}, \bar{h}_{12}, \sigma_{12}, \ldots, \bar{h}_{33}, \sigma_{33}]. \tag{6.19}$$

Vector of these means and standard deviations were used as the feature vector and given to the classifier for training and testing of the proposed method.

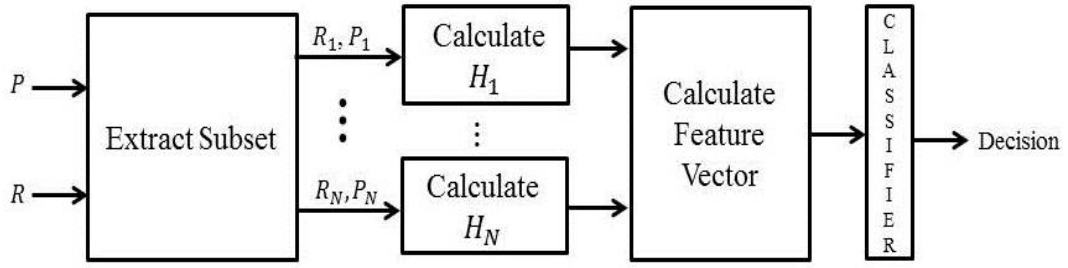

Figure 6.2 System 1, where set of H matrixes are calculated

Figure 6.2 show System 1. P and R are the pupil centre and screen coordinates. Subset of these coordinates were used to estimate the homography matrix.

6.4.2 System 2

System 2 setup is shown in Figure 6.3. In this setup two homography matrices were calculated using two different sets of screen and corresponding pupil coordinates. For the first homography matrix calculation, the odd points (1,2,...,29) screen and pupil coordinates were used. Similarly for calculating the second ho-

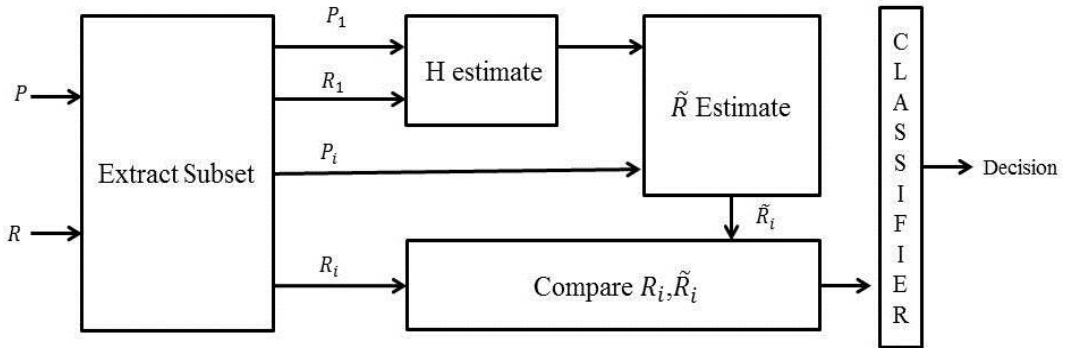

Figure 6.3 System 2

mography matrix, the even (2,4,...30) screen and pupil coordinates were used. The two matrices were then used to estimate the points using the screen coordinates. The Euclidean distances between the second 30 observed and estimated

points were calculated. These Euclidean distances were put in a vector and given to classifier as feature vector.

6.5 Experiment

Several sets of experiments were carried out to evaluate the homography information. In the initial experiments both systems were tested with the same data to find out which is the more effective system. In the final set of experiments the colocation, collinearity and homography features were combined together. The data used for these experiments described in Chapter 3.

In total, 90 pairs of pupil centre coordinates form the captured images were extracted. Various subsets of these landmarks along with the corresponding screen coordinates were used to estimate the homography matrix. This matrix was then used for estimating the gaze of the user to enable us to extract the features which were then used to determine genuine and fake attempts. The spatial coordinates of the landmarks for each session were normalized using the Min-Max normalization technique [134] prior to feature extraction. Min-max algorithm was used in this application due to its simplicity and the absence of outliers in the genuine attempts.

6.5.1 Preliminary Experiments

Preliminary experiments were carried out to explore the accuracy of System 1 and System 2. In the first set of the experiments, the screen coordinates and the observed coordinates were used to estimate the homography matrix.

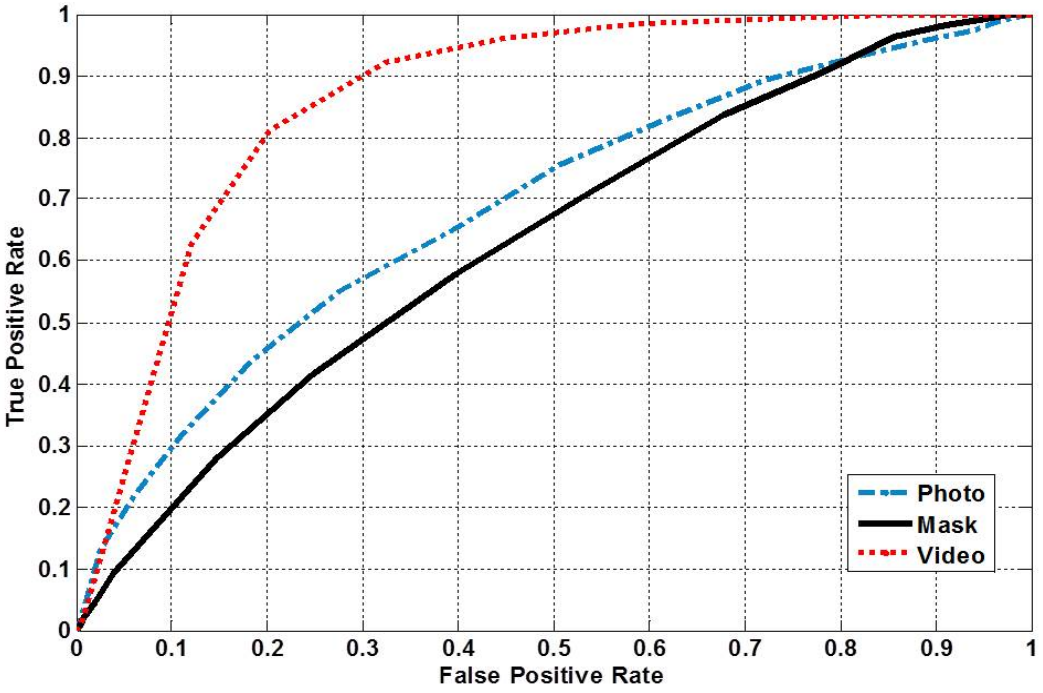

Figure 6.4 ROC curves of System 1 using normalised pupil centre and corresponding screen coordinates for calculating H

Figure 6.4 and Figure 6.5 shows the performance for photo, mask and video replay attack detection using System 1 and System 2 respectively. These experiments were carried out with normalised coordinates. It is clear from the figures, System 2 performed better.

Using System 2, other experiments were carried out with normalizd coordinates to find out the impact of the normalization on the the accuracy of the proposed method. Performance for photo, mask and video replay attack is illustrated in Figure 6.6.

It is clear from these small experiments that System 2 with normalised coordinates performed better. Therefore, the remaining experiments were carried out using System 2 with normalized pupil centre coordinates.

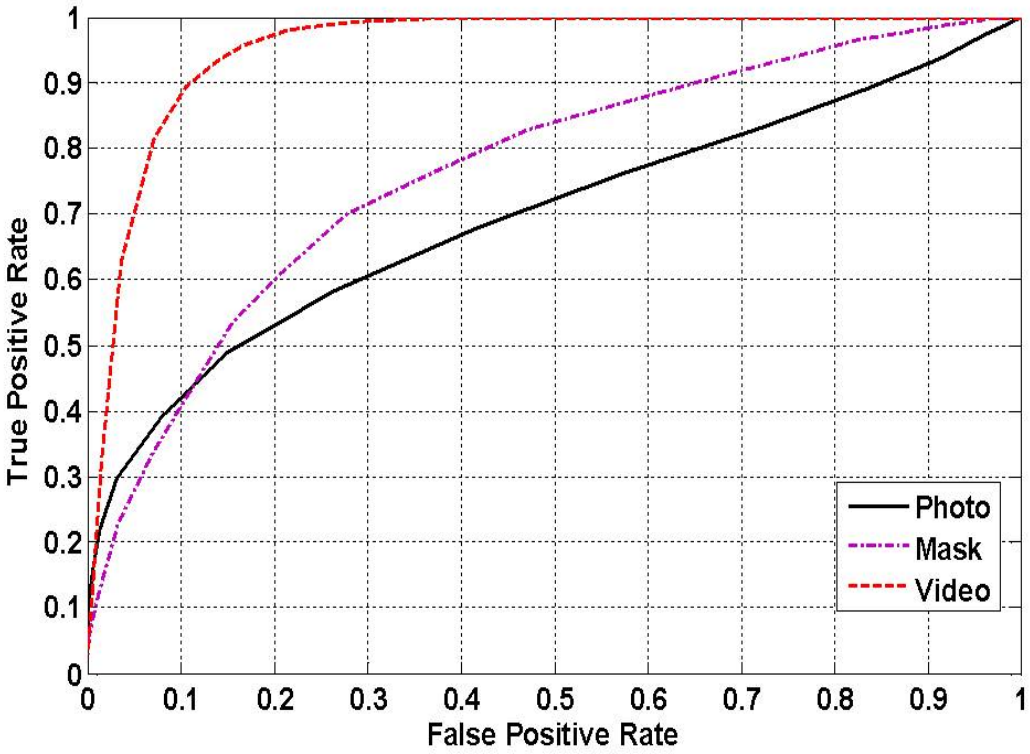

Figure 6.5 ROC curves of System 2 using normalised pupil centre and corresponding screen coordinates for calculating H

6.5.2 Extended Experiments

Once the preliminary experiments were carried out using both System 1 and System 2, it was decided to carry out further detailed experiments using System 2. The aim of this part of the experiments was to further investigate System 2 in more detail to see if the performance of the method could be improved.

Figure 6.5 shows the Receiver Operating Characteristic (ROC) curves for photo, mask and video spoofing attacks using the proposed scheme shown in Figure 6.3. At 10% FPR the TPRs were about 42%, 40% and 90% for photo, mask and video respectively. The system performance for photo and mask attack detection was poor, the mask performance lightly lower than the photo attack. Video

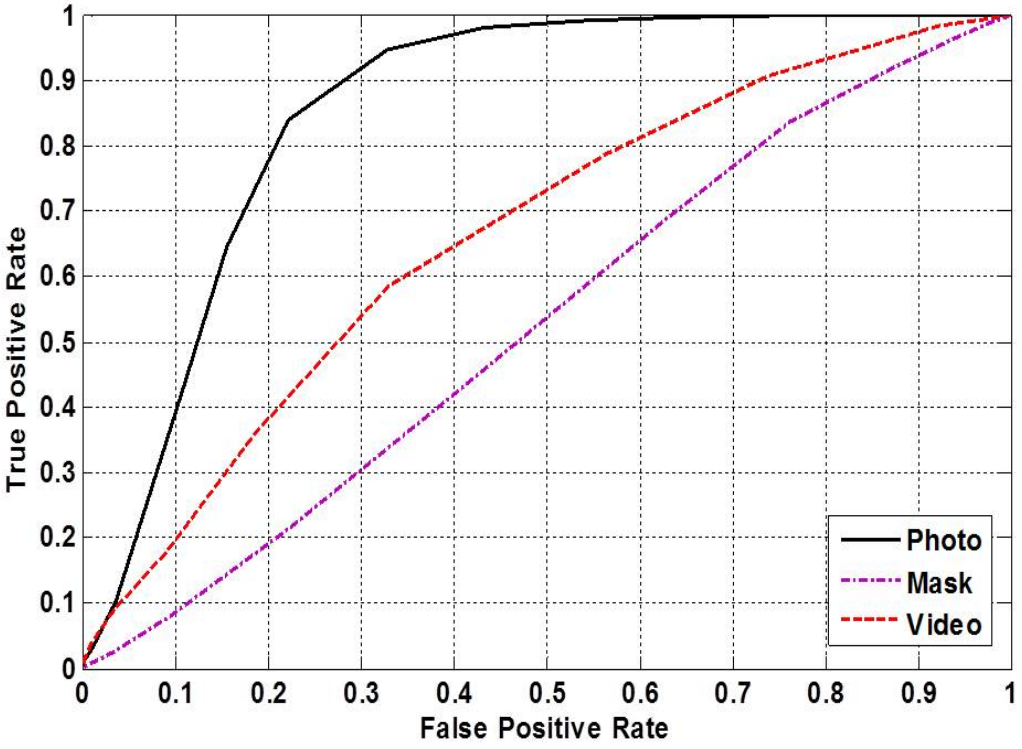

Figure 6.6 ROC curves of System 1 using pupil centre and corresponding screen coordinates for calculating H

performed better compared to the photo and mask attack. This goes back to same reason that video was played to a challenge which was different from the challenge when recording the video. Therefore, the captured relation between the screen coordinates and the pupil centre, did not represent the correct relation between them. Hence the gaze stability did not hold any more which is the sign of impostor attack.

To further improve the accuracy of the proposed method, features were ranked using the forward feature selection method [131]. The feature dimensions were steadily reduced by excluding the least informative features to increase the system performance.

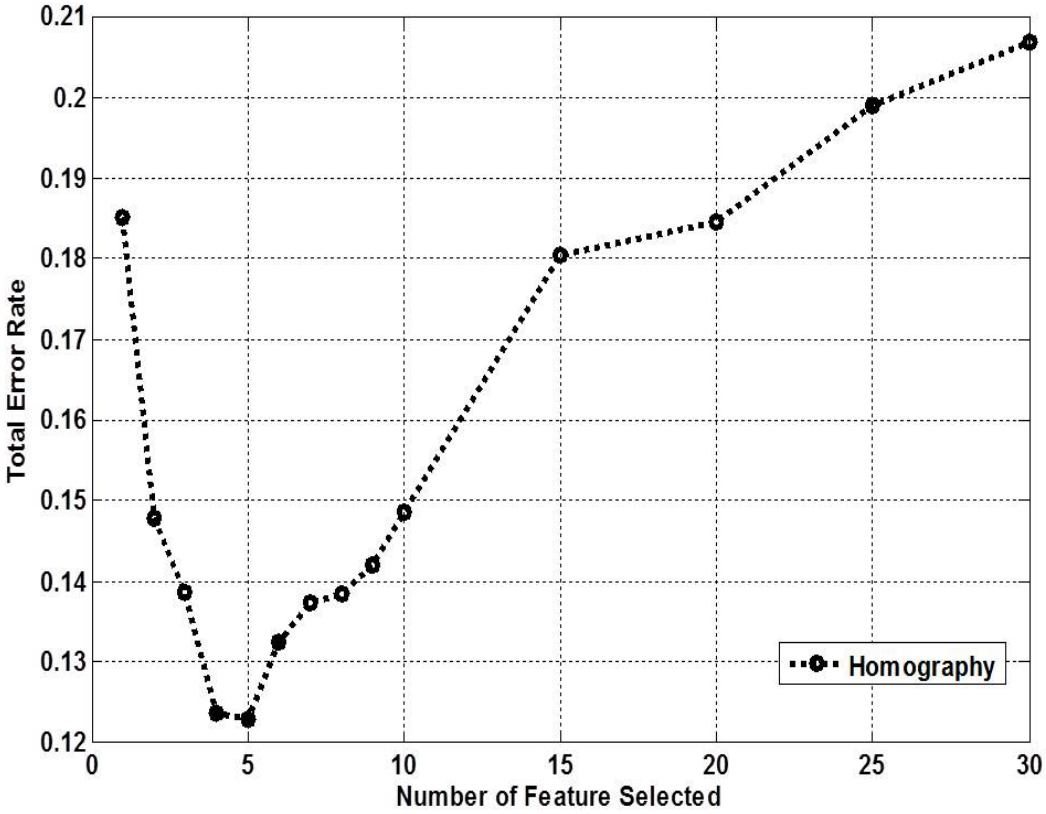

Figure 6.7 Variation in accuracy with homography feature dimension

Figure 6.7 presents total error rates versus the number of features dimensionality selected to find optimum sets for the homography features. In this experiment the photo and mask attack modalities were treated as one presentation attack type against genuine attempts.

The feature selection method was run several times, choosing random sets of data for training and testing for each run. The results of these runs were combined in similar way discussed in Chapter 4. It is shown in Figure 6.7 that the lowest total error rate was observed when the feature dimension was reduced to around 5.

Figure 6.8 shows the ROC curves for the optimum feature set for photo and mask attack detection. Photo and mask attack detection performance improved

using the optimum feature. However, video performance slightly decreased. As video attack performed with high accuracy in all features investigated so far, we did not rank the video feature to find the optimum feature set for video replay attack. Instead we used the feature vector of mask to select the reduced feature set for video replay attacks, due to this exclusion the system performance may be expected to drop for video replay attack. Overall the system displayed a vastly improved performance for photo and mask attacks detection.

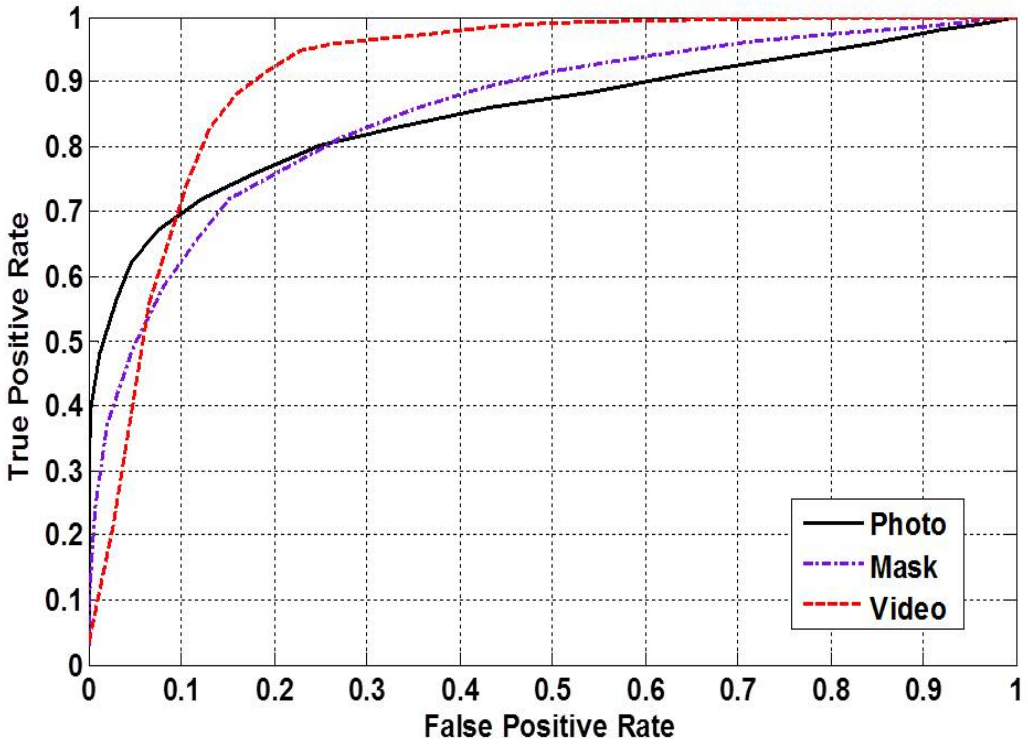

Figure 6.8 ROC curve for phot, mask video using optimum feature vector

Using optimum feature sets, video replay attack 3out-performed while the photo attack detection ranked second and mask ranked third At 10% FPR, video attack detection TPR is 71%, photo attack detection TPR is about 70% and mask attack detection TPR is about 61%.

6.6 Fusion of Colocation, Collinearity and Homography

In these sets of experiments, effectiveness and robustness of the three proposed features, colocation, collinearity and homography information in combination with one another in detecting liveness is explored. The proposed scheme is shown in the Figure 6.9. Features based on gaze stability were extracted from both eyes. These features were explored in detail in this study. Gaze homography is explored in detail in the above sections while colocation and collinearity features are discussed in detail in Chapter 4 and Chapter 5 respectively. In this scheme the scores from collinearity, colocation and homography were combined together to achieve the final score for detecting spoofing attacks.

Several classification schemes were investigated. The k-NN classifier produced the best performance and hence the k-NN classifier was used here. Gaze-based features were extracted from each eye and were passed to separate classifiers (k-NN) to obtain the individual classification scores for each eye. The a-posteriori probabilities from the separate classifiers were combined at the fusion stage using score fusion for liveness detection. The scheme is illustrated in Figure 6.9 which show three gaze-based features were fused using score fusion. The k-NN classifier from the prtools package used for this work automatically determined an optimum k-value for each experiment. Each experiment was run 400 times with random sets of data for training and testing, resulting in a different optimum k values for each run. These optimal k values were combined to obtain an overall mean k value to be used for operational systems. In the case of homography features this mean optimal k value was 7.

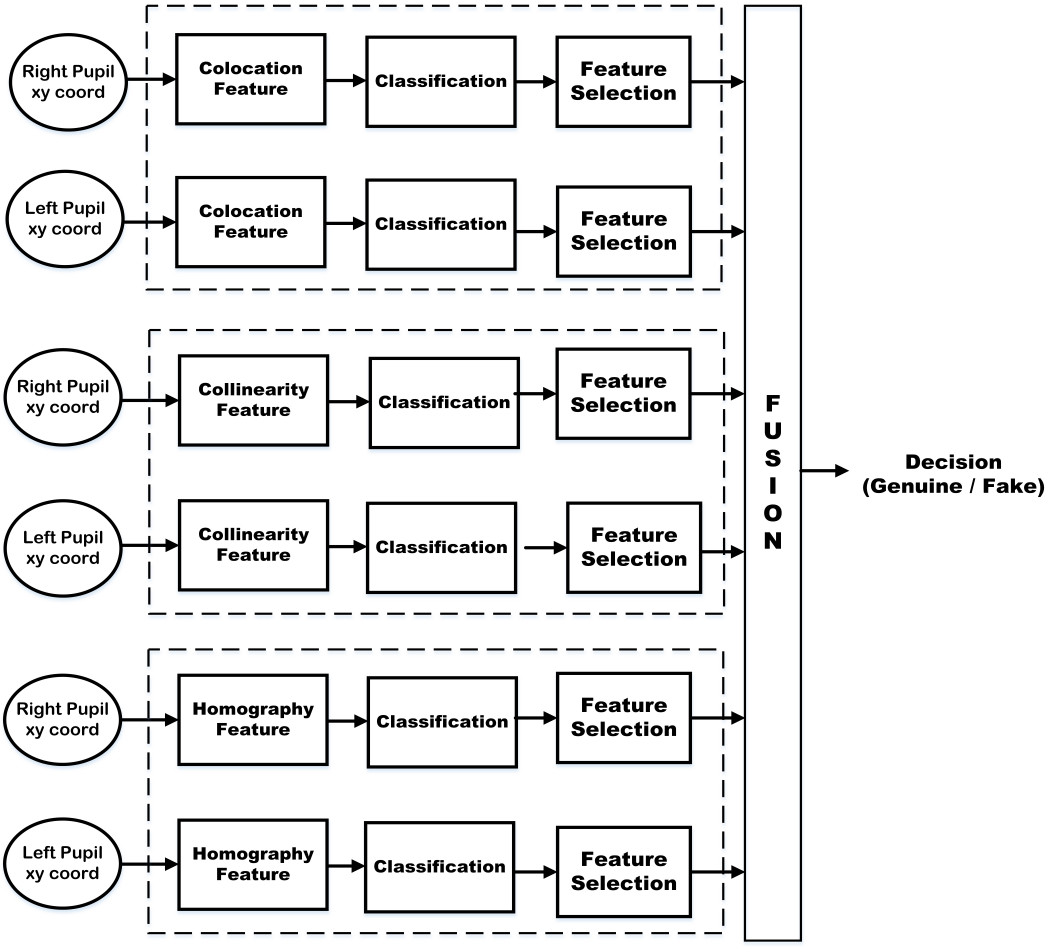

Figure 6.9 Proposed scheme combining collinearity, colocation and homography using score fusion

Figure 6.10 shows the Receiver Operating Characteristic (ROC) curves, using the proposed scheme shown in Figure 6.9 for photo spoof attacks using the entire feature set for collinearity, colocation and homography. It can be seen that the system performance is poor but may be improved further using feature optimization. Photo attack detection TPR was about 75% at 10% FPR.

Similarly Figure 6.11 shows the Receiver Operating Characteristic (ROC) mask attack detection. The performance is close to the photo performance. At 10% FPR, mask attack detection TPR is about 72% which slightly lower than the photo attacks.

Video replay attack performed with no errors. At 10% FPR, video replay

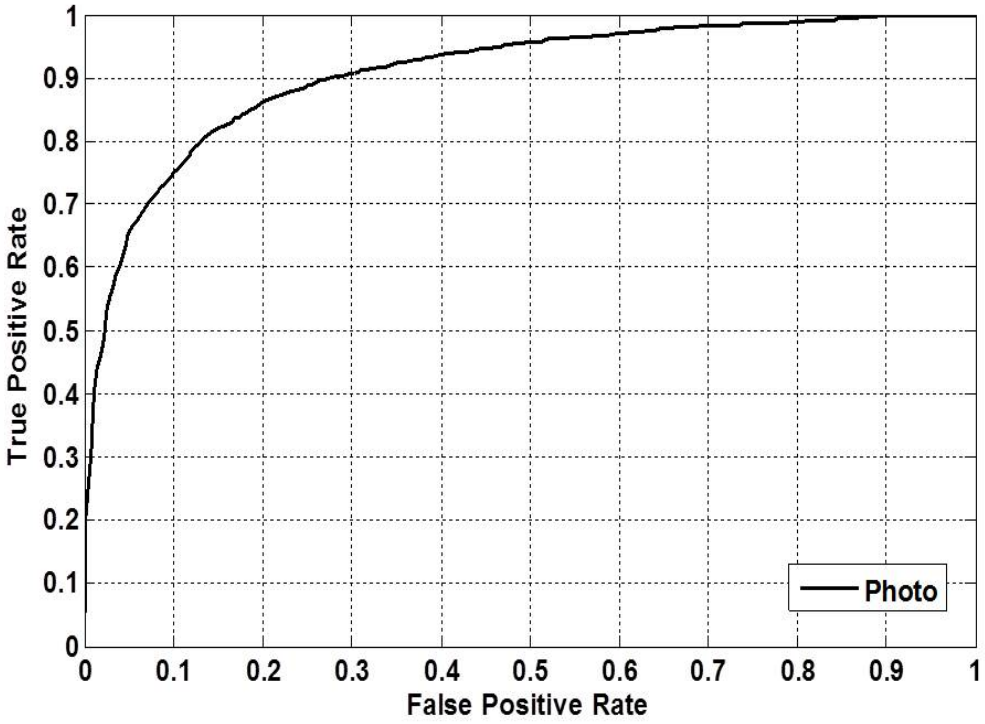

Figure 6.10 ROC curve for photo using proposed fusion scheme

detection TPR is about 100%. Again, the video was captured from the genuine person in response to the challenge. At the time of impostor attack, the user replayed the video in response to the challenge. The challenge presented to the replaying video at the time of simulated impostor attempt, was different from the challenge which was presented to the user at the time of recording the video . Hence gaze stability may not be repeated in any of the extracted features in the video response. That may be the reason why the system performed better for video attack detection.

The performance of the system for mask attacks detection was lower than the photo attacks detection. Masks have small holes in the pupil centre to facilitate the impostor to see through to follow the challenge. In this way the impostor may

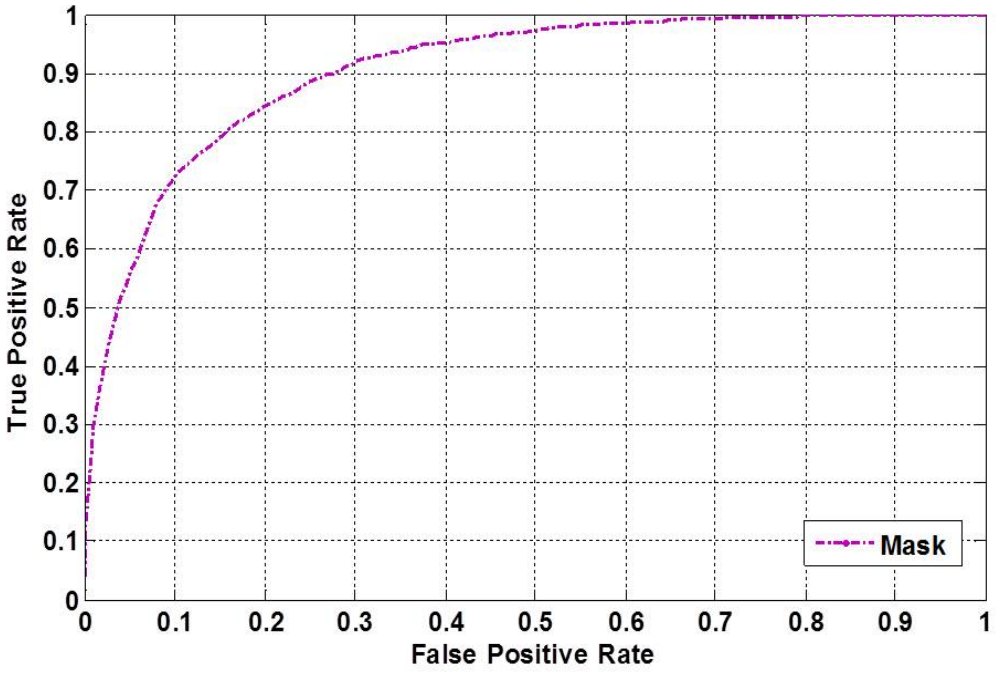

Figure 6.11 ROC curve for mask using proposed fusion scheme

have gaze stability close to genuine attempts. This may have made it difficult for the proposed method to detect the mask attack compared to the photo attack.

In order to establish the tradeoff between the feature dimensionality and liveness detection, the accuracy of the system was improved as the feature dimensions were steadily reduced. A forward feature selection method [131] was used for this purpose.

Two types of feature ranking were carried out. In the first type of ranking photo and mask attack scenarios were ranked separately In the second type of feature ranking, collinearity, colocation and homography features were ranked separately to explore the system performance based on feature type rather then attack scenario.

Figure 6.12 presents total error rates versus the number of features selected to

find optimum sets for the homography features. In this experiment the optimum feature of collinearity, colocation and homography were used for photo and mask attack.

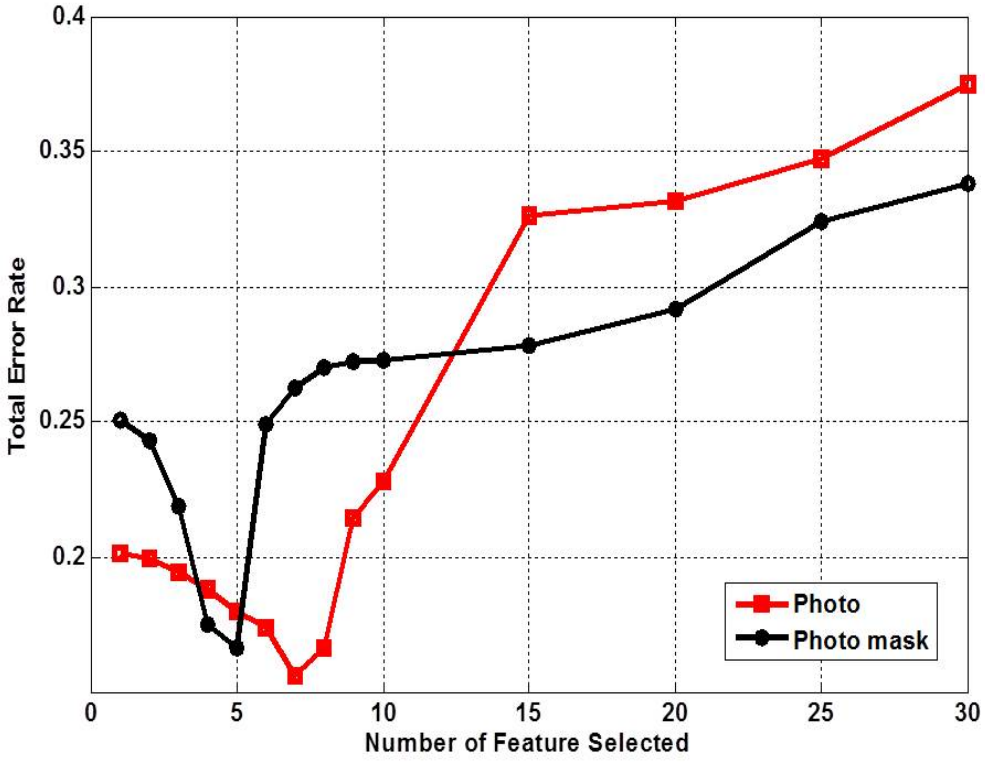

Figure 6.12 Variation in accuracy with feature dimension

The feature selection method was run several times, choosing random sets of data for training and testing for each run. The results of these runs were combined. As shown in Figure 6.12 the lowest total error rate was observed when the feature dimension was reduced to around 5 and 7 for mask and photo attacks respectively.

Figure 6.13 shows the ROC curves for the optimum feature sets for fusion of collinearity, colocation and homography information for photo, mask and video. The performance of the system was found to improve for all attack type as can be seen in Figure 6.13. At 10% FPR, video replay performance is 100% and photo

attack TPR increased to about 98%. The mask attack detection performance also increased after fusion and is lower compared to the video (about 84%).

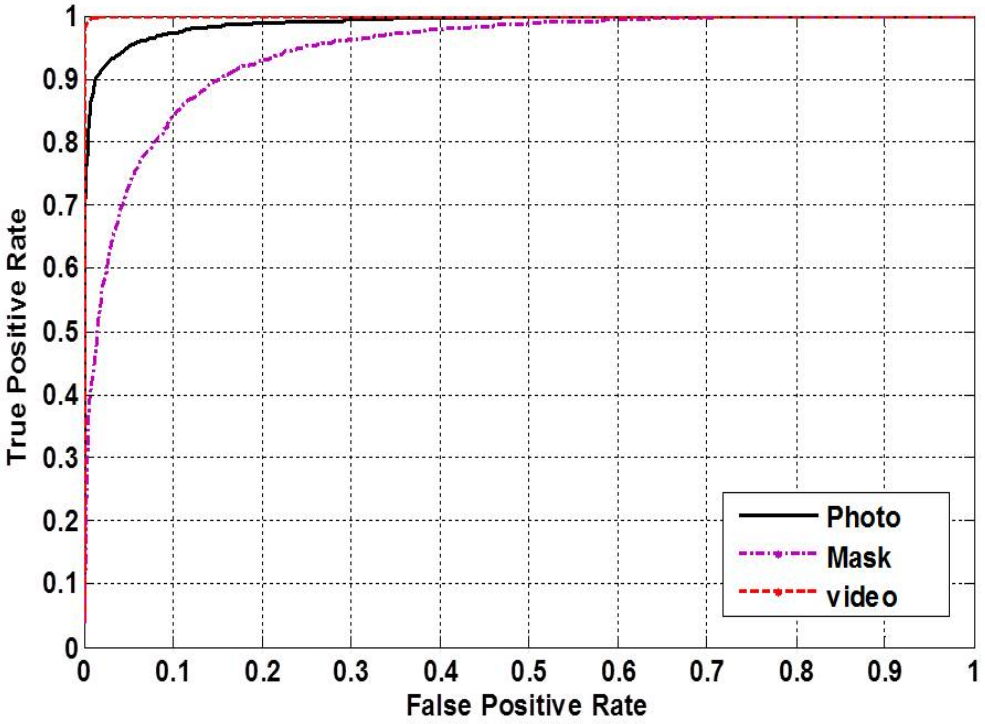

Figure 6.13 ROC curve of the proposed system using optimum feature set schemes

Table 6.1 presents TPR values at 10% FPR for the three spoofing attack detection scenarios. It shows the error rates when collinearity, colocation and homography features are implemented on their own, as well as in multi-classifier configuration. It is clear that in all the schemes implemented, the video replay attack detection outperformed the other two types of attacks. The collinearity features were superior to the colocation and homography features. Using score fusion, the TPR for photo attack detection and the photo mask detection performance improved, and video replay attack detection give perfect performance.

Table 6.1 TPR at FPR = 0.10 using optimum feature sets

Feature Sets	Photo	Mask	Video Replay
Collinearity	0.69	0.79	1.00
Colocation	0.70	0.38	0.65
Homography	0.69	0.60	0.71
Fusion	0.97	0.84	1.00

In this experiment the photo and mask attack modalities were combined and treated together as one spoof attack against genuine attempts. Figure 6.14 illustrates the ROC curves for collinearity, colocation and homography features. The collinearity feature gave the best performance compared to the colocation and homography features. Homography is second in performance followed by colocation. At 10% FPR, TPR of about 73%, 23% and 43%, were achieved for collinearity, colocation, homography. The colocation feature performance is much weaker compared to the performance of collinearity, homography and fusion based schemes.

Table 6.2 summarizes the performance of the proposed method. It presents TPR values for a range of FPRs for the novel gaze-based features and their fusion. It presents the error rates for collinearity, colocation and homography features when they are implemented on their own, as well as in multi-classifier configurations. As can be seen in the Table, the fusion of the collinearity, colocation and homography performed better. The collinearity features were superior to the colocation and homography features. Using score fusion, the TPR for impostor attack detection performance improved compared to when the proposed features were implemented on their own.

In order to establish the tradeoff between the feature dimensionality and liveness detection, the accuracy of the system was improved as the feature dimensions were steadily reduced. A forward feature selection method [131] was used

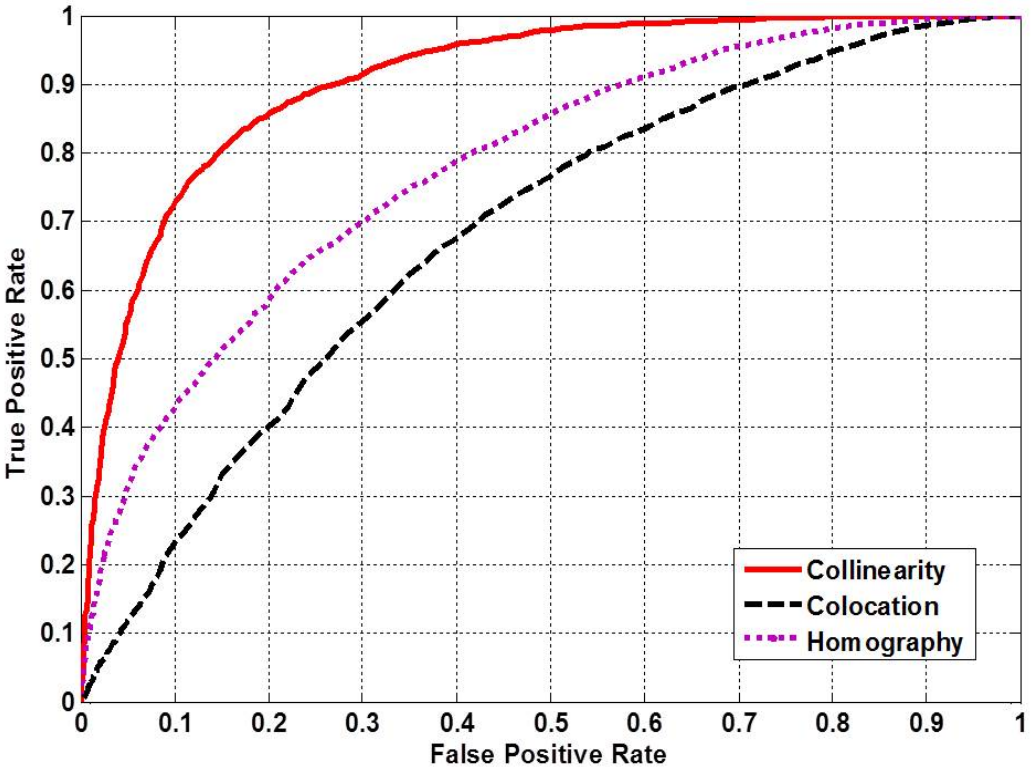

Figure 6.14 Collinearity, colocation and homography feature

Table 6.2 TPR at FPR from 0.01 to 0.05 using entire feature sets

Feature	**FPR =0.01**	**FPR = 0.03**	**FPR = 0.05**	**FPR = 0.07**	**FPR = 0.10**
Collinearity	0.22	0.43	0.56	0.64	0.72
Colocation	0.03	0.08	0.12	0.16	0.23
Homography	0.11	0.24	0.32	0.37	0.43
Fusion	0.37	0.54	0.64	0.72	0.78

for this purpose. In this ranking procedure, photo and mask attack scenarios were combined and treated as a single impostor attack. The collinearity, colocation and homography features were ranked individually to see if the the system performance could be further enhance the proposed method performance.

Figure 6.15 presents total error rates as a function of the number of features selected to find optimum feature sets for combined collinearity, colocation and

homography features for photo and mask attack scenarios. It is clear from the Figure that feature dimension of 5 to 6 performed better.

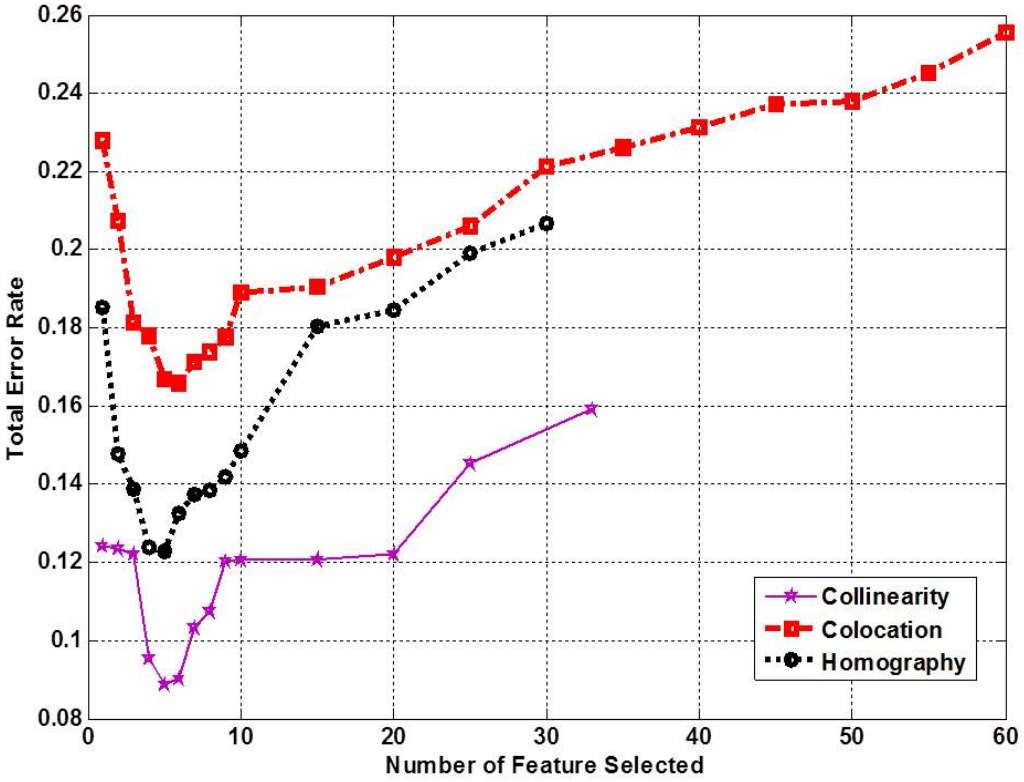

Figure 6.15 f:Variation in accuracy with gaze feature dimension for collinearity, colocation and homography

The collinearity, colocation and homography features were ranked. In this ranking the photo and mask were treated as single spoofing attack. The combination of these attack modalities allows the establishment of a single optimal feature set that can be used for all of these major spoofing challenges. Video attack data was excluded from this feature ranking exercise as the system already performs very well in detecting video spoofing attacks. The feature selection method was run several times, choosing random sets of data for training and testing for every run. The results of these runs were combined to rank the features. In this scheme, each feature is ranked separately and optimum feature sets for each feature type is passed to a separate classifier.

Figure 6.16 illustrates the ROC curves for collinearity, colocation and homography using optimum feature. Collinearity feature gave the best performance compared to colocation and homography features. Homography is second in performance followed by colocation. At 10% FPR, TPR of about 92%, 62% and 78%, were achieved for collinearity, colocation, homography. The colocation feature performance is much weaker compared to the performance of collinearity, homography and fusion-based schemes.

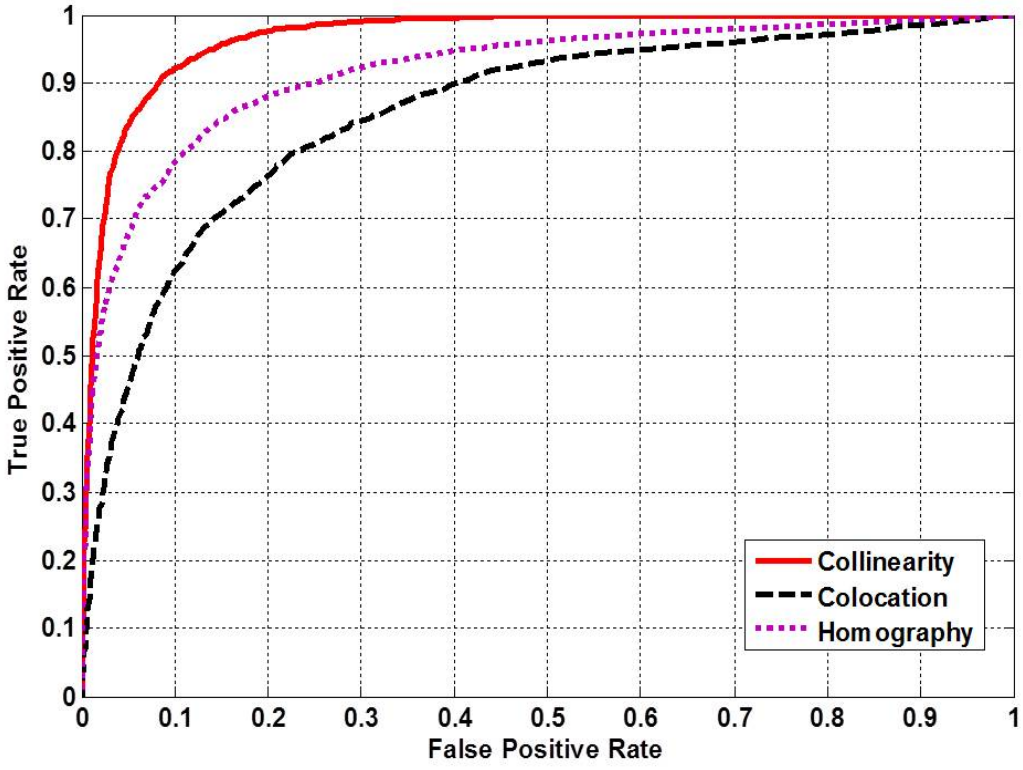

Figure 6.16 Collinearity, colocation and homography feature using an optimum feature sets

In this experiment all spoof attack types were combined as a single spoof attack. Figure 6.17 illustrates the ROC curves for genuine and fake attempts using score fusion. The performance of the proposed system is relatively good to avert impostor attack. At 10% FPR, TPR of about 78% and 98%, were achieved for entire and optimum feature respectively.

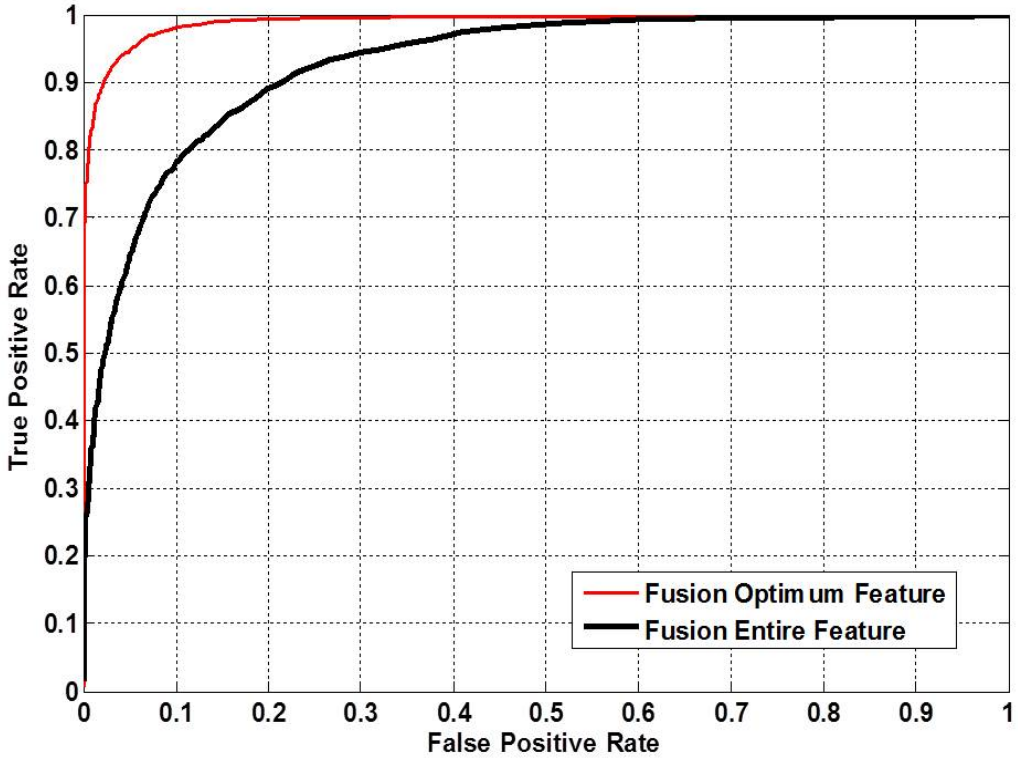

Figure 6.17 ROC curve for impostor detection using fusion

Table 6.3 shows a comparison of our experimental results (GS for Gaze Stability) with the performances reported for similar photo spoofing attacks published in the literature. Although the results are based on different databases they indicate the relative promise of the proposed methods. The performance of our proposed approach can be seen to compare favourably with the other methods considered and lends support to its applicability in detecting spoofing attacks.

6.7 Conclusion

This chapter presents another novel technique for face liveness detection that may be used to protect against presentation attacks. The proposed system is a challenge-response approach using a visual stimulus to measure the gaze of the

Table 6.3 Comparison of performance reports

Method	FPR	FNR
Kollreider *et al.* [12]	0.02	0.19
Tan *et al.* [12]	0.09	0.18
Peixoto *et al.* [141]	0.07	0.07
IGD [142]	0.17	0.01
MaskDown [142]	0.00	0.05
GS Photo Attack	0.03	0.07
GS All Attack	0.07	0.03

user for the purpose of establishing the presence of photographic, mask and video spoofing attacks.

Gaze homography features are extracted from the pupil centre and used to classify between live and fake attempts. Preliminary experiments were carried out to explore different schemes to find the one which gave better performance.

We analysed the performance of the system using the pupil centre and the screen coordinates. The normalised features gave better performance.

The main contribution of this chapter has been,

- Gaze based novel feature
- Investigating various schemes.
- Investigating three types of presentation attack which were photo, mask and video replay attacks.

The next chapter provides the summary, conclusion and recommendations for future work.

CHAPTER 7

Conclusions and Future Work

7.1 Introduction

In this chapter, a summary of the work covered in this thesis is provided. This will be followed by a discussion of the major research findings of this work, and suggestions for future research directions.

7.2 Summary and Conclusion

The work presented here explores the notion of gaze stability and features based on it for the task of detecting presentation attacks on facial biometric systems. Using a visual stimulus to direct the gaze, the system provides an accurate measure to discriminate between genuine and fake attempts. Information from different eyes and using different algorithms have been combined in a score-fusion framework and evaluated for a number of spoofing attack scenarios. Three attack scenarios were investigated and data was collected to evaluate the performance of

the proposed system using different combinations of features and attack modalities. The multi-classifier approach, combining information from separate feature sets using score fusion provided the best results and was seen to compare well with the state of the art showing the potential effectiveness and viability of this approach.

This thesis investigated novel features for face liveness detection based on gaze information. The goal of this thesis has been to investigate novel features for face liveness detection and performed an extensive experimental study of various classification and combination rules to overcome the problem of liveness detection.

A review of the literature in face liveness detection was presented in Chapter 2. A number of techniques have been reported in the literature in recent years addressing the problem of face liveness detection. Previous work was grouped into two main categories: active and passive. These categories are further subdivided into several groups based on the features used.

This thesis includes four contributions in the field of biometric liveness detection.

In Chapter 3 we provided the general framework for the evaluation of the proposed systems including the database and its collection procedure. We provided the detail of software and hardware used in the data collection process. We also provided further details on the evaluation strategy used for this work. The data is stored on a secure sever where other researchers can benefit form it in their study.

Chapter 4 presented the first novel feature explored in this study. An efficient and robust collinearity feature is presented here. A number of preliminary experiments were carried out to designed the challenge and to reduce the duration.

The experiment shows that features extracted along the horizontal direction gave better performance for distinguishing fake attempts. The results indicated that features extracted from combined xy coordinates of the pupil center performed better compared to single x or y coordinates. A number of classifiers and combiners were used to fuse the information from a number of facial landmarks. The combiners both feature-based and score-based fusion were implemented. ROC curves were used to analyze and assess the performance of the proposed collinearity feature.

Chapter 5 presents a novel feature to improve the performance of face liveness detection using a combination of various stimuli locations. Also Chapter 5 provides the results of combining collinearity and colocation features using fusion based rules. The results suggest that a combination of features may be used to improve the performance of the face liveness detection systems. The score fusion based performed better than the single feature. The video replay attack have outperformed. The performance of the photo and mask attack is also improved.

Chapter 6 provides details of a novel gaze homography features. Here we present the performance results for a number of schemes using homography. The K-nearest neighbour classifier was used for experiments. Finally this chapter covers the fusion of all the features that were investigated in the research. The results were promising using gaze homography feature. However the fusion of collinearity, colocation and homography features performed significantly better compared to individual feature.

In Chapter 7 we provided some concluding remarks and future work suggestions for researcher in the field.

7.3 Main Contributions

The main contributions of this thesis are the following:

- A study and review of the literature on liveness detection algorithms.
- Detailed evaluation of the proposed new features.
- Development of fusion schemes for combining different scores of liveness information.
- The development of three novel features based on gaze information.

7.4 Recommendations for Future Work

One area where more improvement can be made to the proposed system is in the feature extraction stage, where more robust landmark detection methods can be used to extract the gaze information more accurately. Two possible areas for future research to extract novel features include fixations and saccade [143] [144], [145]. Similarly one could investigate the the duration of the fixation and saccade of the user in following the challenge. Features may be grouped into two categories as follows:

Feature which can be collected at particular instances during the challenge such as collinearity, colocation, homography, etc, so called static feature.

Features which can be extracted continuously called dynamic feature. Example of the feature can fixation duration, saccade duration, acceleration between fixation and saccade and so on.

Although some satisfying solutions exist in the fixation and saccade [145], there is much room for improvement. One area of research is to investigate the suitability of ordinary web cameras for fixations and saccades rather then using expensive eye trackers that normally used for this purpose.

Future work may also be focused on the design of the stimulus so that the duration of the challenge is minimized while maintaining a high spoof detection rate. The conjecture that human beings can move and position their head/eye more easily, and with more accuracy in the horizontal direction, could also be further investigated with a series of tests using alternative stimuli and screen arrangements.

In this study a small object appears at random locations and these locations should not be too close to one another and each should be visited multiple times, using a 21.5" LCD screen to display the object. However, one could investigate whether a similar system can be developed to work with a smaller screen and a mobile devices.

Then this liveness detection approach can be incorporated in smart phones and tablets and can be used for logging into smart devices. This liveness detection system may further enhance the security of the existing face recognition systems (and other biometric technologies) which are used in the mobile devices.

In this study only 2D mask was investigated. In future 3D mask can also be investigated. More data and subjects can be another part one can contribute.

List of Publications

Articles in conference proceedings

1. Asad Ali, Farzin Deravi and Sanaul Hoque, "*Liveness detection using gaze collinearity*," In Emerging Security Technologies (EST), 2012, IEEE Third International Conference, pp. 62-65, 2012. DOI: 10.1109/EST.2012.12

2. Asad Ali, Farzin Deravi and Sanaul Hoque, "*Spoofing attempt detection using gaze colocation*," in Biometrics Special Interest Group (BIOSIG), 2013 International Conference, pp. 1-12, 2013.

3. Asad Ali, Farzin Deravi and Sanaul Hoque, "*Directional Sensitivity of Gaze-Collinearity Features in Liveness Detection*," In Emerging Security Technologies (EST), 2013, IEEE Fourth International Conference , 2013, pp. pp. 8-11. DOI: 10.1109/EST.2013.7

Article submitted in journal

1. Asad ali, Farzin Deravi and Sanaul Hoque, "*Fusion of Gaze Stability Information for Facial Liveness Detection*," Special issue on "Information Fusion in Biometrics" (Information Fusion, Elsevier), 2015.

Bibliography

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

[2] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, Oct 2007, pp. 1–8.

[3] 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, no. 3-4, pp. 215–225, 2011.

[4] H.-K. Jee, S.-U. Jung, and J.-H. Yoo, "Liveness detection for embedded face recognition system," *International Journal of Biomedical Sciences*, vol. 1, no. 4, pp. 235–238, 2006.

[5] L. Wang, X. Ding, and C. Fang, "Face live detection method based on physiological motion analysis," *Tsinghua Science & Technology*, vol. 14, no. 6, pp. 685–690, 2009.

[6] 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 *Biometrics (ICB), 2012 5th IAPR International Conference on*, March 2012, pp. 67–72.

[7] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

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

[9] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of fourier spectra," in *Defense and Security*.

[10] L. Wu, X. Xu, Y. Cao, Y. Hou, and W. Qi, "Live face detection by combining the fourier statistics and lbp," in *Biometric Recognition*. Springer, 2014, pp. 173–181.

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

[12] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in liveness assessment," *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 3, pp. 548–558, 2007.

[13] A. M. K. Saad, "Anti-spoofing using challenge-response user interaction," 2015.

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

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

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

[17] P. P. Chan and Y. Shu, "Face liveness detection by brightness difference," in *Machine Learning and Cybernetics*. Springer, 2014, pp. 144–150.

[18] C. Roberts, "Biometric attack vectors and defences," *Computers & Security*, vol. 26, no. 1, pp. 14–25, 2007.

[19] S. Prabhakar, J. Kittler, D. Maltoni, L. O'Gorman, and T. Tan, "Introduction to the special issue on biometrics: Progress and directions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 513–516, 2007.

[20] A. Ross and A. Jain, *Multimodal biometrics: An overview*. na, 2004.

[21] A. K. Jain and A. Ross, "Multibiometric systems," *Communications of the ACM*, vol. 47, no. 1, pp. 34–40, 2004.

[22] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 1, pp. 4–20, 2004.

[23] J. A. Markowitz, "Voice biometrics," *Communications of the ACM*, vol. 43, no. 9, pp. 66–73, 2000.

[24] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of biometrics*. Springer Science & Business Media, 2007.

[25] D. A. Reynolds, "An overview of automatic speaker recognition," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)(S. 4072-4075)*, 2002.

[26] H. E. Talhami, A. S. Malegaonkar, R. A. Malegaonkar, and C. D. Summerfield, "Voice authentication and speech recognition system and method," Jan. 15 2015, uS Patent 20,150,019,220.

[27] T. K. Perrachione, S. N. Del Tufo, and J. D. Gabrieli, "Human voice recognition depends on language ability," *Science*, vol. 333, no. 6042, pp. 595–595, 2011.

[28] A. K. Jain, F. D. Griess, and S. D. Connell, "On-line signature verification," *Pattern recognition*, vol. 35, no. 12, pp. 2963–2972, 2002.

[29] 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.

[30] F. Alonso-Fernandez, M. C. Fairhurst, J. Fierrez, and J. Ortega-Garcia, "Automatic measures for predicting performance in off-line signature," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 1. IEEE, 2007, pp. I–369.

[31] M. Faundez-Zanuy and J. M. Pascual-Gaspar, "Efficient on-line signature recognition based on multi-section vector quantization," *Pattern Analysis and Applications*, vol. 14, no. 1, pp. 37–45, 2011.

[32] F. Deravi and S. P. Guness, "Gaze trajectory as a biometric modality." in *BIOSIGNALS*, 2011, pp. 335–341.

[33] A. Maeder, C. Fookes, and S. Sridharan, "Gaze based user authentication for personal computer applications," in *Intelligent Multimedia, Video and*

Speech Processing, 2004. Proceedings of 2004 International Symposium on. IEEE, 2004, pp. 727–730.

[34] R. Bednarik, T. Kinnunen, A. Mihaila, and P. Fränti, "Eye-movements as a biometric," in *Image analysis*. Springer, 2005, pp. 780–789.

[35] V. Cantoni, C. Galdi, M. Nappi, M. Porta, and D. Riccio, "Gant: Gaze analysis technique for human identification," *Pattern Recognition*, vol. 48, no. 4, pp. 1027–1038, 2015.

[36] C. Galdi, M. Nappi, D. Riccio, V. Cantoni, and M. Porta, "A new gaze analysis based soft-biometric," in *Pattern Recognition*. Springer, 2013, pp. 136–144.

[37] D. Gafurov, "A survey of biometric gait recognition: Approaches, security and challenges," in *Annual Norwegian Computer Science Conference*. Citeseer, 2007, pp. 19–21.

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

[39] N. Boulgouris, D. Hatzinakos, and K. Plataniotis, "Gait recognition: a challenging signal processing technology for biometric identification," *Signal Processing Magazine, IEEE*, vol. 22, no. 6, pp. 78–90, Nov 2005.

[40] D. S. Matovski, M. S. Nixon, S. Mahmoodi, and J. N. Carter, "The effect of time on gait recognition performance," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 543–552, 2012.

[41] J. Daugman, *Biometric decision landscapes*. University of Cambridge, Computer Laboratory, 2000, no. 482.

[42] J. Daugman and C. Downing, "Epigenetic randomness, complexity and singularity of human iris patterns," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 268, no. 1477, pp. 1737–1740, 2001.

[43] R. P. Wildes, "Iris recognition: an emerging biometric technology," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1348–1363, 1997.

[44] C. Sanchez-Avila, R. Sanchez-Reillo, and D. de Martin-Roche, "Iris-based biometric recognition using dyadic wavelet transform," *IEEE Aerospace and Electronic Systems Magazine*, vol. 17, no. 10, pp. 3–6, 2002.

[45] K. W. Bowyer, K. P. Hollingsworth, and P. J. Flynn, "A survey of iris biometrics research: 2008–2010," in *Handbook of iris recognition.* Springer, 2013, pp. 15–54.

[46] Ö. Polat and T. Yıldırım, "Hand geometry identification without feature extraction by general regression neural network," *Expert systems with Applications*, vol. 34, no. 2, pp. 845–849, 2008.

[47] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzalez-Marcos, "Biometric identification through hand geometry measurements," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1168–1171, 2000.

[48] A. Kumar, D. C. Wong, H. C. Shen, and A. K. Jain, "Personal verification using palmprint and hand geometry biometric," in *Audio-and Video-Based Biometric Person Authentication.* Springer, 2003, pp. 668–678.

[49] A. Kumar and D. Zhang, "Hand-geometry recognition using entropy-based discretization," *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 2, pp. 181–187, June 2007.

[50] H. Li, K.-A. Toh, and L. Li, *Advanced topics in biometrics*. World Scientific, 2012.

[51] A. A. Ross, J. Shah, and A. K. Jain, "Toward reconstructing fingerprints from minutiae points," in *Defense and Security*. International Society for Optics and Photonics, 2005, pp. 68–80.

[52] S. Prabhakar, "Fingerprint classification and matching using a filterbank," Ph.D. dissertation, Michigan State University, 2001.

[53] A. Jain, L. Hong, and R. Bolle, "On-line fingerprint verification," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 4, pp. 302–314, 1997.

[54] A. K. Jain, L. Hong, S. Pankanti, and R. Bolle, "An identity-authentication system using fingerprints," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1365–1388, 1997.

[55] S. Shekhar, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Joint sparse representation for robust multimodal biometrics recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 1, pp. 113–126, 2014.

[56] O. Boumbarov, S. Sokolov, and G. Gluhchev, "Combined face recognition using wavelet packets and radial basis function neural network," in *Proceedings of the 2007 international conference on Computer systems and technologies*. ACM, 2007, p. 98.

[57] J.-G. Wang, H. Kong, E. Sung, W.-Y. Yau, and E. K. Teoh, "Fusion of appearance image and passive stereo depth map for face recognition based on the bilateral 2dlda," *Journal on Image and Video Processing*, vol. 2007, no. 2, pp. 6–6, 2007.

[58] C.-C. Liu, D.-Q. Dai, and H. Yan, "Local discriminant wavelet packet coordinates for face recognition," *The Journal of Machine Learning Research*, vol. 8, pp. 1165–1195, 2007.

[59] A. Schwaninger, S. Schumacher, H. Bülthoff, and C. Wallraven, "Using 3d computer graphics for perception: The role of local and global information in face processing," in *Proceedings of the 4th symposium on applied perception in graphics and visualization*. ACM, 2007, pp. 19–26.

[60] G. Rhodes, L. Jeffery, L. Taylor, W. G. Hayward, and L. Ewing, "Individual differences in adaptive coding of face identity are linked to individual differences in face recognition ability." 2014.

[61] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 9, pp. 1160–1165, 2003.

[62] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain, "Performance evaluation of fingerprint verification systems," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 1, pp. 3–18, 2006.

[63] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: a tool for information security," *Information Forensics and Security, IEEE Transactions on*, vol. 1, no. 2, pp. 125–143, 2006.

[64] L. O'Gorman, "Comparing passwords, tokens, and biometrics for user authentication," *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2021–2040, 2003.

[65] J. Galbally, J. Fierrez, and J. Ortega-Garcia, "Vulnerabilities in biometric systems: attacks and recent advances in liveness detection," *Database*, vol. 1, no. 3, p. 4, 2007.

[66] A. K. Jain and A. Kumar, "Biometrics of next generation: An overview," *Second Generation Biometrics*, 2010.

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

[68] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, "Impact of artificial gummy fingers on fingerprint systems," in *Electronic Imaging 2002*. International Society for Optics and Photonics, 2002, pp. 275–289.

[69] V. Ruiz-Albacete, P. Tome-Gonzalez, F. Alonso-Fernandez, J. Galbally, J. Fierrez, and J. Ortega-Garcia, "Direct attacks using fake images in iris verification," in *Biometrics and identity management*. Springer, 2008, pp. 181–190.

[70] "Biometrics new portal," http://www.biometricnewsportal.com/, [Online; accessed 10-07-2014].

[71] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern recognition letters*, vol. 24, no. 13, pp. 2115–2125, 2003.

[72] Z. Dol, R. A. Salam, and Z. Zainol, "Face feature extraction using bayesian network," in *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia*. ACM, 2006, pp. 261–264.

[73] F. Cavallini, Alessio (Nice, "Liveness detection," Patent 20 150 003 692, January, 2015. [Online]. Available: http://www.freepatentsonline.com/y2015/0003692.html

[74] R. Raghavendra, K. Raja, and C. Busch, "Presentation attack detection for face recognition using light field camera," *Image Processing, IEEE Transactions on*, vol. 24, no. 3, pp. 1060–1075, March 2015.

[75] J. Pelz, M. Hayhoe, and R. Loeber, "The coordination of eye, head, and hand movements in a natural task," *Experimental Brain Research*, vol. 139, no. 3, pp. 266–277, 2001.

[76] F. C. Volkmann, "Human visual suppression," *Vision research*, vol. 26, no. 9, pp. 1401–1416, 1986.

[77] M. F. Land, "Eye movements and the control of actions in everyday life," *Progress in retinal and eye research*, vol. 25, no. 3, pp. 296–324, 2006.

[78] S. Chakraborty and D. Das, "An overview of face liveness detection," *arXiv preprint arXiv:1405.2227*, 2014.

[79] J. Galbally, S. Marcel, and J. Fierrez, "Biometric anti-spoofing methods: A survey in face recognition."

[80] I. Chingovska and A. Anjos, "On the use of client identity information for face anti-spoofing," *Information Forensics and Security, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[81] Y. Ijiri, M. Sakuragi, and S. Lao, "Security management for mobile devices by face recognition," in *Mobile Data Management, 2006. MDM 2006. 7th International Conference on*, May 2006, pp. 49–49.

[82] M. G. Doane, "Interaction of eyelids and tears in corneal wetting and the dynamics of the normal human eyeblink," *American journal of ophthalmology*, vol. 89, no. 4, pp. 507–516, 1980.

[83] C. N. Karson, "Spontaneous eye-blink rates and dopaminergic systems," *Brain*, vol. 106, no. 3, pp. 643–653, 1983.

[84] ——, "Spontaneous eye-blink rates and dopaminergic systems," *Brain*, vol. 106, no. 3, pp. 643–653, 1983.

[85] K. Tsubota, "Tear dynamics and dry eye," *Progress in retinal and eye research*, vol. 17, no. 4, pp. 565–596, 1998.

[86] A. R. Bentivoglio, S. B. Bressman, E. Cassetta, D. Carretta, P. Tonali, and A. Albanese, "Analysis of blink rate patterns in normal subjects," *Movement Disorders*, vol. 12, no. 6, pp. 1028–1034, 1997.

[87] A. De Santis and D. Iacoviello, "Robust real time eye tracking for computer interface for disabled people," *Computer methods and programs in biomedicine*, vol. 96, no. 1, pp. 1–11, 2009.

[88] A. Krolak and P. Strumillo, "Vision-based eye blink monitoring system for human-computer interfacing," in *Human System Interactions, 2008 Conference on*. IEEE, 2008, pp. 994–998.

[89] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–511.

[90] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[91] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.

[92] F. Sha and F. Pereira, "Shallow parsing with conditional random fields," in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, 2003, pp. 134–141.

[93] A. McCallum, "Efficiently inducing features of conditional random fields," in *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers Inc., 2002, pp. 403–410.

[94] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[95] 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*. Springer, 2010, pp. 504–517.

[96] J. Komulainen, A. Hadid, and M. Pietikäinen, "Face spoofing detection using dynamic texture," in *Computer Vision-ACCV 2012 Workshops*. Springer, 2013, pp. 146–157.

[97] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 915–928, 2007.

[98] Z. Zhiwei *et al.*, "A face antispoofing database with diverse attacks," in *Proceedings of the 5th IAPR International Conference on Biometrics (ICB12), New Delhi, India*, vol. 1, 2012.

[99] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: a public database and a baseline," in *Biometrics (IJCB), 2011 International Joint Conference on*. IEEE, 2011, pp. 1–7.

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

[101] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, "Computationally efficient face spoofing detection with motion magnification," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. IEEE, 2013, pp. 105–110.

[102] M. M. Chakka, A. Anjos, S. Marcel, R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori *et al.*, "Competition on counter measures to 2-d facial spoofing attacks," in *Biometrics (IJCB), 2011 International Joint Conference on*. IEEE, 2011, pp. 1–6.

[103] J. Komulainen, A. Hadid, and M. Pietikäinen, "Face spoofing detection using dynamic texture," in *Computer Vision-ACCV 2012 Workshops*. Springer, 2013, pp. 146–157.

[104] J. Maatta, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *Biometrics (IJCB), 2011 international joint conference on*. IEEE, 2011, pp. 1–7.

[105] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using texture and local shape analysis," *IET biometrics*, vol. 1, no. 1, pp. 3–10, 2012.

[106] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "Lbp- top based countermeasure against face spoofing attacks," in *Computer Vision-ACCV 2012 Workshops*. Springer, 2013, pp. 121–132.

[107] D. Das and S. Chakraborty, "Face liveness detection based on frequency and micro-texture analysis," in *Advances in Engineering and Technology Research (ICAETR), 2014 International Conference on*, Aug 2014, pp. 1–4.

[108] N. Eveno and L. Besacier, "A speaker independent" liveness" test for audio-visual biometrics." in *INTERSPEECH*, 2005, pp. 3081–3084.

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

[110] S. Kant, "Fake face detection based on skin elasticity," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 5, pp. 1048–1051.

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

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

[113] ——, "Non-intrusive liveness detection by face images," *Image and Vision Computing*, vol. 27, no. 3, pp. 233–244, 2009.

[114] R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, and F. Roli, "Fusion of multiple clues for photo-attack detection in

face recognition systems," in *Biometrics (IJCB), 2011 International Joint Conference on*. IEEE, 2011, pp. 1–6.

[115] A. Pinto, W. Schwartz, H. Pedrini, and A. Rocha, "Using visual rhythms for detecting video-based facial spoof attacks," *Information Forensics and Security, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[116] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *Information Forensics and Security, IEEE Transactions on*, vol. 1, no. 2, pp. 205–214, June 2006.

[117] A. Rocha, W. Scheirer, T. Boult, and S. Goldenstein, "Vision of the unseen: Current trends and challenges in digital image and video forensics," *ACM Computing Surveys (CSUR)*, vol. 43, no. 4, p. 26, 2011.

[118] M. Chung, J. Lee, H. Kim, S. Song, and W. Kim, "Automatic video segmentation based on spatio-temporal features," *Korea Telecom Journal*, vol. 4, no. 1, pp. 4–14, 1999.

[119] G. Chetty and M. Wagner, "Audio-video biometric system with liveness checks."

[120] J. Bai, T.-T. Ng, X. Gao, and Y.-Q. Shi, "Is physics-based liveness detection truly possible with a single image?" in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, May 2010, pp. 3425–3428.

[121] G. Chetty and M. Wagner, "Multi-level liveness verification for face-voice biometric authentication," in *Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the*. IEEE, 2006, pp. 1–6.

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

[123] S. Kim, S. Yu, K. Kim, Y. Ban, and S. Lee, "Face liveness detection using variable focusing," in *Biometrics (ICB), 2013 International Conference on*, June 2013, pp. 1–6.

[124] S. K. Nayar and Y. Nakagawa, "Shape from focus," *Pattern analysis and machine intelligence, IEEE Transactions on*, vol. 16, no. 8, pp. 824–831, 1994.

[125] L. Yang, "Face liveness detection by focusing on frontal faces and image backgrounds," in *Wavelet Analysis and Pattern Recognition (ICWAPR), 2014 International Conference on*. IEEE, 2014, pp. 93–97.

[126] S. Kim, Y. Ban, and S. Lee, "Face liveness detection using defocus," *Sensors*, vol. 15, no. 1, pp. 1537–1563, 2015.

[127] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Computer Vision–ECCV 2008*. Springer, 2008, pp. 504–513.

[128] T. Fawcett, "Roc graphs: Notes and practical considerations for researchers," *Machine learning*, vol. 31, pp. 1–38, 2004.

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

[130] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial intelligence*, vol. 97, no. 1, pp. 245–271, 1997.

[131] A. Ali, F. Deravi, and S. Hoque, "Liveness detection using gaze collinearity," in *Emerging Security Technologies (EST), 2012 Third International Conference on*, Sept 2012, pp. 62–65.

[132] ——, "Directional sensitivity of gaze-collinearity features in liveness detection," in *Emerging Security Technologies (EST), 2013 Fourth International Conference on*, Sept 2013, pp. 8–11.

[133] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[134] N. R. Draper, H. Smith, and E. Pownell, *Applied regression analysis*. Wiley New York, 1966, vol. 3.

[135] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," in *Advances in neural information processing systems*, 2004, pp. 1569–1576.

[136] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, "On combining classifiers," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 3, pp. 226–239, 1998.

[137] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[138] "Affine Transformations," http://www.maa.org/sites/default/files/pdf/pubs/books/meg/meg_ch12.pdf, [Online; accessed 10-07-2013].

[139] E. Dubrofsky, "Homography estimation," Ph.D. dissertation, UNIVERSITY OF BRITISH COLUMBIA (Vancouver, 2009.

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

[141] I. Chingovska, J. Yang, Z. Lei, D. Yi, S. Z. Li, O. Kahm, C. Glaser, N. Darner, A. Kuijper, A. Nouak *et al.*, "The 2nd competition on counter measures to 2d face spoofing attacks." in *ICB*, 2013, pp. 1–6.

[142] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 2000, pp. 71–78.

[143] D. P. Munoz and B. D. Corneil, "Evidence for interactions between target selection and visual fixation for saccade generation in humans," *Experimental Brain Research*, vol. 103, no. 1, pp. 168–173, 1995.

[144] O. Komogortsev and A. Karpov, "Liveness detection via oculomotor plant characteristics: Attack of mechanical replicas," in *Biometrics (ICB), 2013 International Conference on*, June 2013, pp. 1–8.