TAMPERE UNIVERSITY OF TECHNOLOGY

MUHAMMAD ADEEL WARIS

NO INTRUDERS - SECURING FACE BIOMETRIC SYSTEMS FROM SPOOFING ATTACKS

Master's thesis

Examiner(s):
Professor Moncef Gabbouj
Dr. Iftikhar Ahmad
Examiner and topic approved by the Faculty Council of the Faculty of Computing and Electrical Engineering on 15th January 2014.

ABSTRACT

TAMPERE UNVIERSITY OF TECHNOLGOY
Master's Degree in Information Technology
MUHAMMAD ADEEL WARIS: No intruders - Securing Face Biometric Systems from Spoofing Attacks
Master of Science Thesis, pages 52
Month and year of completion: April 2014 (Examiner and topic were approved in the faculty council meeting on 15th January 2014)
Major: Signal Processing
Examiner(s):
Prof. Moncef Gabbouj, Department of Signal Processing, Tampere University of Technology
Dr. Ifikhar Ahmad, Department of Signal Processing, Tampere University of Technology
Keywords: biometric, face detection, anti-spoofing, texture, image quality assessment, motion detection, classification.

The use of face verification systems as a primary source of authentication has been very common over past few years. Better and more reliable face recognition system are coming into existence. But despite of the advance in face recognition systems, there are still many open breaches left in this domain. One of the practical challenge is to secure face biometric systems from intruder's attacks, where an unauthorized person tries to gain access by showing the counterfeit evidence in front of face biometric system. The face-biometric system having only single 2-D camera is unaware that it is facing an attack by an unauthorized person. The idea here is to propose a solution which can be easily integrated to the existing systems without any additional hardware deployment. This field of detection of imposter attempts is still an open research problem, as more sophisticated and advanced spoofing attempts come into play.

In this thesis, the problem of securing the biometric systems from these unauthorized or spoofing attacks is addressed. Moreover, independent multi-view face detection framework is also proposed in this thesis. We proposed three different counter-measures which can detect these imposter attempts and can be easily integrated into existing systems. The proposed solutions can run parallel with face recognition module. Mainly, these counter-measures are proposed to encounter the digital photo, printed photo and dynamic videos attacks. To exploit the characteristics of these attacks, we used a large set of features in the proposed solutions, namely local binary patterns, gray-level co-occurrence matrix, Gabor wavelet features, space-time auto-correlation of gradients, image quality based features. We further performed extensive evaluations of these approaches on two different datasets. Support Vector Machine (SVM) with the linear kernel and Partial Least Square Regression (PLS) are used as the classifier for classification. The experimental results improve the current state-of-the-art reference techniques under the same attach categories.

PREFACE

This work has been conducted at the Department of Signal Processing of Tampere University of Technology.

I thank my colleagues at the MUVIS research group and the personnel of the Department of Signal Processing for providing such a pleasant and inspiring working atmosphere. In particular, I would like to thank Prof. Moncef Gabbouj for trusting me and providing me this wonderful opportunity to work on this interesting topic. I thank Dr. Iftikhar and Mr. Honglei Zhang for always being willing to help me when confronted with challenges during whole thesis process.

I am very grateful to my friends Naveed Bin Jaffar, Muhammad Faisal, Aitzaz Haider Kazmi, Ranjeeth Shetty, Rahman Akbar, Mubashir Ali and many more for giving me such a memorable time staying abroad. Special thanks to Umar Iqbal, Zohaib Hassan, Majid Ali Khan for spending time with me and collecting the spoofing dataset.

Last but not least, I would like to thank my siblings and parents for their moral support.

Muhammad Adeel Waris
16th April 2014

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
BoF	Bag of Features
DoG	Difference of Gaussian
DCT	Discrete Cosine Transform
ELBP	Extended Local Binary Pattern
EER	Equal Error Rate
FAR	False Acceptance Rate
FPN	Fixed Pattern Noise
FR	Full Reference
FRR	False Rejection Rate
GLCM	Gray Level Co-occurrence Matrix
GME	Global Motion Estimation
HOG	Histogram of Oriented Gradients
HSC	Histogram of Shearlet Coefficients
HT	Homogeneous Texture
HTER	Half Total Error Rate
ID	Identifier
IQA	Image Quality Assessment
LBP	Local Binary Pattern
LBPV	Local Binary Patterns Variance
LBP−TOP	Local Binary Patterns for Three Orthogonal Planes
LibSVM	Library for Support Vector Machine
MB-LBP	Multi-Block Local Binary Patterns
MVFD	Multi-View Face Detection
NR	No Reference
OpenCV	An Open-source Computer Vision library
PIN	Personal Identification Numbers
PLS	Partial Least Square Regression
PRNU	Photo Responsiveness of Non-Uniform light
ROI	Region of Interest
ROC	Receiver Operating Characteristic
RBF	Radial Basis Functions
SIFT	Scale Invariant Feature Transform
STACOG	Spatio-Temporal Auto-Correlation of Gradients
SURF	Speed-Up Robust Features
SVM	Support Vector Machines

1. INTRODUCTION

Humans distinguish one another according to various physiological characteristics of individuals. We recognize others by their face when we meet them, by their voice as we hear them. Traditionally, identity verification (authentication) has been generally based on something that one holds (key, chip card) or one remember (password). Identity verification occurs when the user claimed to be already enrolled in the system and presents an ID card or login name; in this case the verification biometric data received from the user is compared to the user's data already stored in the database. Identification (also known search) occurs when the identity of the user is a priori unknown. In this case the user's biometric data is matched against all the records in the database as the user can be anywhere in the database. The prevailing techniques of user authentication, which require the use of either passwords and user IDs (identifiers), or identification cards and PINs (personal identification numbers), suffer from several limitations. Things like keys or cards, however, can get stolen or misplaced, passwords and PINs can be illegally acquired by direct covert observation. Once an intruder obtains the user ID and the password, the intruder has full access to the user's resources. To achieve more reliable verification or identification we should use something that really characterizes the person.

With the upsurging of large-scale computer networks and increasing number of applications making use of such networks, true authentication predicated on biometrics have received increased attention during the last few years. Systems that deliver the power to authenticate persons accurately, swiftly, reliably, without invading privacy issues, cost effectively, in a user-friendly manner and without requiring radical modifications to the existing infrastructures are desired. Biometric technologies can be utilized for automated identity verification or identification by combining physiological or behavioural features such as, ***fingerprints, iris, hand geometry, signature, face*** and ***voice recognition*** as illustrated in Figure 1-1. These characteristics are measurable and unique. It is almost impossible to lose or forget biometrics, since they are an intrinsic part of each person, and this is an advantage which they hold over keys, passwords or codes. As a result, they are more reliable since biometric information cannot be lost, guessed or forgotten, easily.

The use of biometrics as a primary source of authentication in commercial application has been communal in last few years [3][4][46]. Many multi-national companies has adopted this swift way of providing the authentication to the employee's. Biometric technologies are also being used in police departments and secret agencies all over the globe to identify the criminals based on the forensic evidences like; fingerprints, DNA, and face verifications obtained from the crime scenes video footage. The most commonly used application of biometric systems deployed to date is the electronic passport, which possess two fingerprints in addition to a passport photograph. Moreover, it speeds up border crossing through the use of scanners, which use the principle of recognition by comparison of the face or fingerprints. Many countries have set up biometric infrastructures to control migration flows to and from their territories. The same applies to visa applications and renewals. Moreover, biometrics technologies are commonly used for commercial applications such as electronic data security, computer, mobile phones etc.

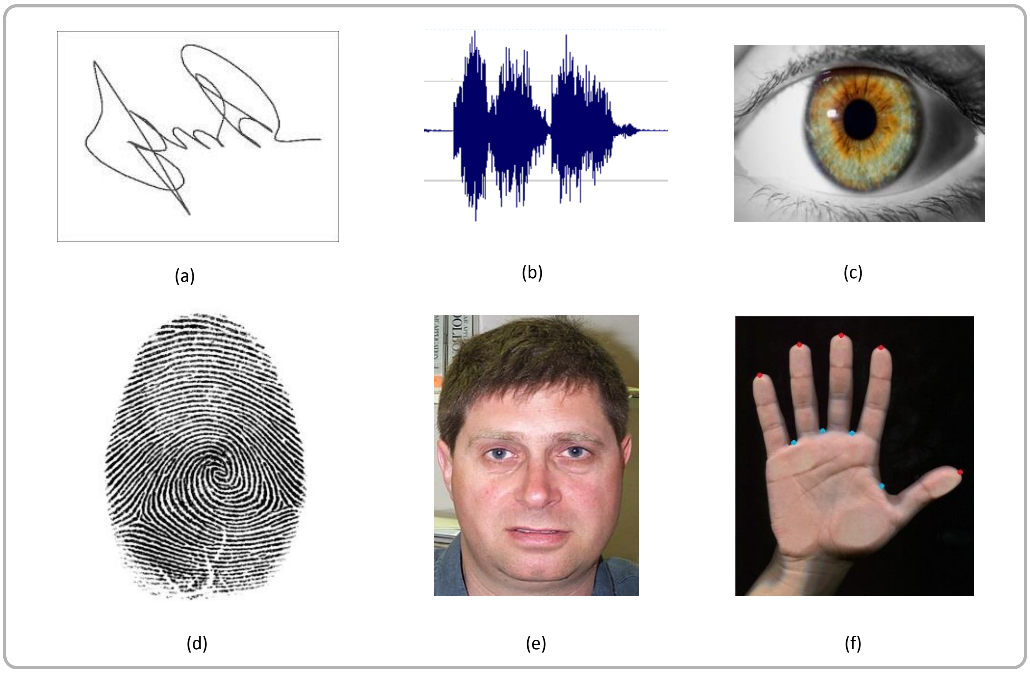


Figure 1-1: Sample biometric traits: (a) signature, (b) voice (c) iris, (d) fingerprint, (e) face, and (f) hand geometry

With growing populations and their increasing mobility, recognition of humans using biological characteristics became a promising solution for identity management. Among many reliable biometric traits, face is the popular one and it owes this reputation mainly due to its accessibility and easiness. But unfortunately, this gift can also be a curse in malicious circumstances, enabling attackers to easily create copies and spoof face recognition systems. Section 1.1 describes the basic concept of spoofing attacks to face biometric systems, the focus area of this thesis.

1.1. Spoofing Attacks to Face Biometric Systems

Spoofing is an attempt to gain authentication through a biometric system by presenting a counterfeit evidence of a valid user [35]. In a spoofing attempt, a person tries to masquerade as another person and thereby, tries to gain access to the system. In this thesis, such events are referred as **Imposter** or **Spoofing attempts** and the rest are considered as **Real attempts**. The aim is to develop non-intrusive methods without extra devices and human involvement. This will ensure the compatibility with the existing face recognition systems. Despite of the advances in biometric authentication systems can still be deceived in one way or the other, e.g., consider the case in which one person instead of showing his/her own face to a biometric system displays a photo of an authorized counterpart either printed on a piece of paper, on a laptop, or even on a cell phone screen. For instance recent mobile phone feature, “*Face Unlock*”, which uses face recognition to unlock a phone, has received criticism for being vulnerable to spoofing attacks.

The face biometric system are based on intensity images and equipped with a generic camera is which cannot distinguish visually between real and invalid attempts. The objective of this thesis is to analyse the multiple videos events where different users are requesting access from biometric systems, the goal is to detect and restrict the imposter attempts without any additional hardware except for a generic web camera. These scenarios can differ with each other in complexity level and can be problematic for face verification systems as better spoofing scenarios come into play. In this thesis, three different countermeasures are presented to secure the face biometric systems from invalid users. The texture-based method explores the texture artifacts and the quality degradation that appear when an image or video is recaptured. The motion-based method magnifies the motion and explores the unnatural movements on the scene in the case of spoofing attacks, while the image quality based method tries to detect the degradation artifacts of the video scene. To compare the performance with state of the art methods, the proposed framework has also been tested on multiple dataset where videos are captured by different equipment’s and different environmental setup such as lighting conditions and different backgrounds. Framework to integrate anti-spoofing countermeasures with existing face verification system in realistic manner is also proposed in this thesis.

1.2. Thesis Outline

The rest of this thesis is organized as follows. Chapter-2 briefly explains types of spoofing attacks, description of feature extraction and classification techniques used in this thesis. It also depicts the literature review about state-of-art anti-spoofing solutions. Based on the theoretical foundations formed in Chapter-2, Chapter-3 describes the implementation details of three different proposed countermeasures for anti-spoofing systems. The real world proposed framework that can be integrated in biometric systems is explained in Chapter-4 followed by the experimental results of the whole framework on two different datasets. Finally, Chapter-5 concludes the whole thesis along with a detailed sketch of possibilities for further enhancements and possible future directions.

2. THEORETICAL BACKGROUND

This chapter develops the foundations for the key concepts used in this thesis. It discusses the theoretical details of ideas used to make face biometrics systems robust to imposter attacks. The chapter starts with the explanation of basic local features and motion analysis used in anti-spoofing algorithms. Despite that, there has been work done in this field none of them were able to explain why the textural features have such distinctive results in capturing the changes in live and non-live videos. The key concept of distinctive noise patterns, induced during the recapturing process is also explained briefly. Later, other key concepts, such as, SVMs (Support Vector Machines) for supervised classification and the literature review about the state-of-the-art anti-spoofing counter measures are described.

2.1. Types of Spoofing Attacks

Although there have been important advances in face recognition systems since last couple of decades, but detection of impostor attempts is still an open research problem. Biometric authentication systems can still be deceived in one way or the other, e.g., Duc et al. in [45] showed how to successfully spoof a laptop verification system using only a printed photograph. Face biometric spoofing can be categorized mainly in three categories;

- Photo attacks, showing printed attacks or a video sequence of pictures of the authorized user.
- Video attacks, displaying a dynamic scene video of the valid user.
- Showing a 3D face model of the valid user to the biometric system

Producing an accurate 3D face model of a valid user might be demanding and need some expertise but other two attack scenarios can be implemented easily due to the fact of growing social media forums such as Facebook, Twitter, LinkedIn, etc. Therefore, this thesis deals with the first two categorize of anti-spoofing, some basic attack scenarios are shown in Figure 2-1.

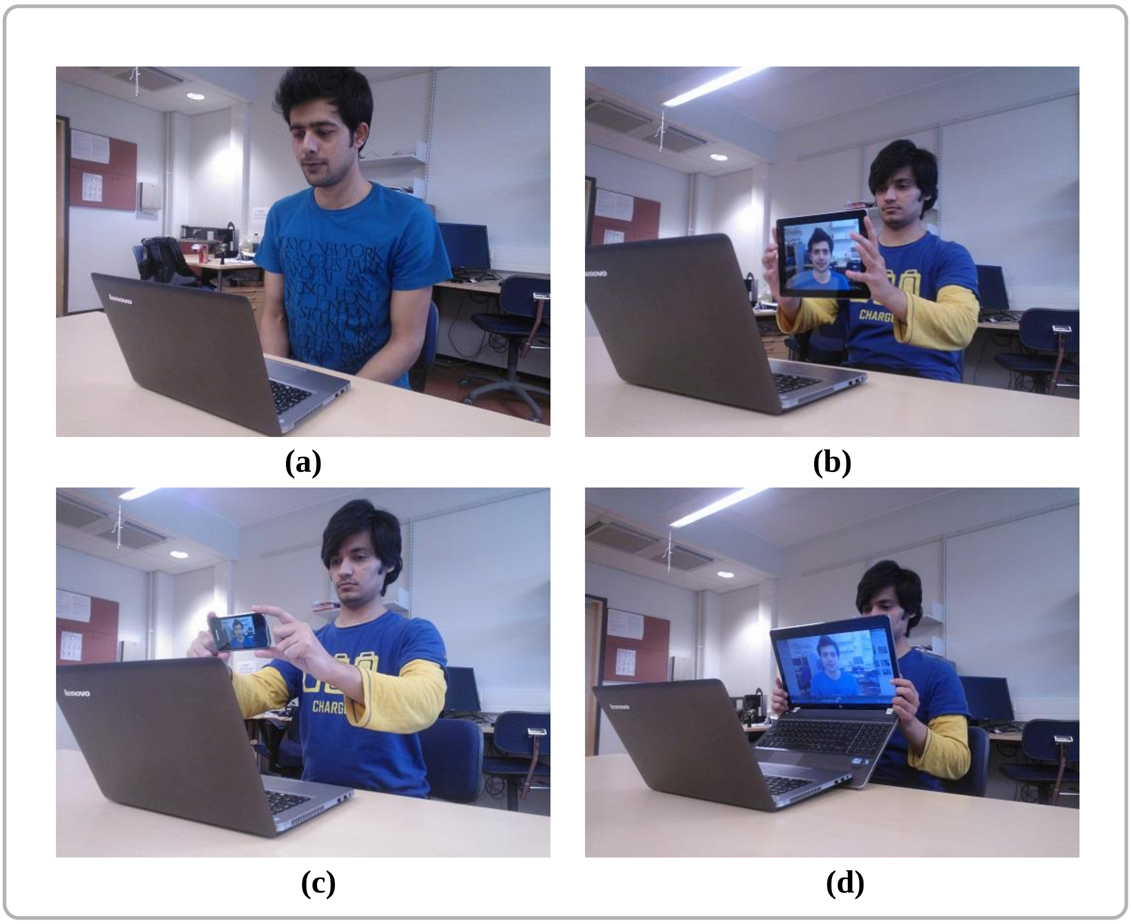


Figure 2-1: Sample biometric attacks: (a) Real, (b) photo attack (c) mobile video attack, (d) dynamic video scene attack

2.2. Feature Extraction

Feature extraction is one of the most intrinsic steps of any classification problem. In general, features extraction techniques are divided into two types; local and global feature extractors. Good feature extraction techniques that give less intra-class variance and large inter-class variance are the most crucial element in the performance of any identification algorithm.

A variety of features have been exploited to capture the temporal variations of video samples. This includes textural feature and motion based feature. Textural features are normally the first choice whenever it comes to differentiate between objects based on the spatial arrangements of colors or intensities in an image. The textural features opted in this thesis includes; 1) Rotation Invariant Uniform Local Binary Patterns ($\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}}$) [28] to extract the local spatial structure of images. 2) Gabor wavelet features to extract the multi-scale, multi-direction spatial frequency characteristics by enriching the intensity variations. 3) Gray Level Co-occurrence Matrix (GLCM) to estimate various properties of spatial layout of an image. The rest of this section briefly explains the aforementioned features.

2.2.1. Local Binary Patterns

Local Binary Patterns (LBP), was first proposed by Ojala et al. [61], have been proved to be robust against illumination variations and effective for capturing the underlying textural information of an image [58],[61]. Since the development of LBP, its many variants have been proposed in the literature such as Extended-LBP [61][70], Improved- LBP [25], MB- LBP [58], Rotation invariant- LBP [61] etc.

The name "Local Binary Pattern" reflects the functionality of the operator, i.e., a local neighbourhood is thresholded at the gray value of the centre pixel into a binary pattern. The basic LBP operates on a 3x3 kernel to encode the local spatial structure of image by comparing pixel intensity of the center pixel with its eight neighbours. The pixels in this block are thresholded by its center pixel value, multiplied by powers of two and then summed to obtain a label for the center pixel. As the neighbourhood consists of 8 pixels, a total of $2^8 = 256$ different labels can be obtained depending on the relative gray values of the center and the pixels in the neighbourhood. An example of an LBP image and histogram are shown in Figure 2-2.

$$\mathrm{LBP}_{P,R} = \sum_{p=0}^{P-1} s\left(g_p - g_c\right)2^p \tag{2.1}$$

$$\mathrm{s(x)} = \begin{cases} 1, & \text{if } \mathrm{x} \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.2}$$

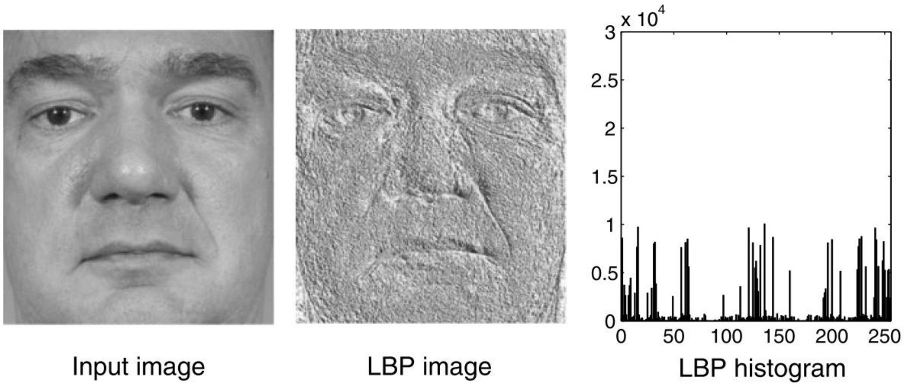


Figure 2-2: Example of LBP histogram

where g_c and g_p denote the gray values of the central pixel and its neighbour, respectively, and p is the index of the neighbour. P is the number of the neighbours, and R is the radius of the circularly neighboring set. Supposing that the coordinate g_c of is (0,0), the coordinate of each neighbouring pixel g_c is then determined according to its index p and parameter (P,R) as$(\mathrm{R}\cos(2\pi p/P))$, $(\mathrm{R}\sin(2\pi p/P))$. The gray values of the neighbors not located at the image grids can be estimated by an interpolation operation. Three circularly symmetric neighbouring sets with different (P, R) are illustrated in Figure 2-3.

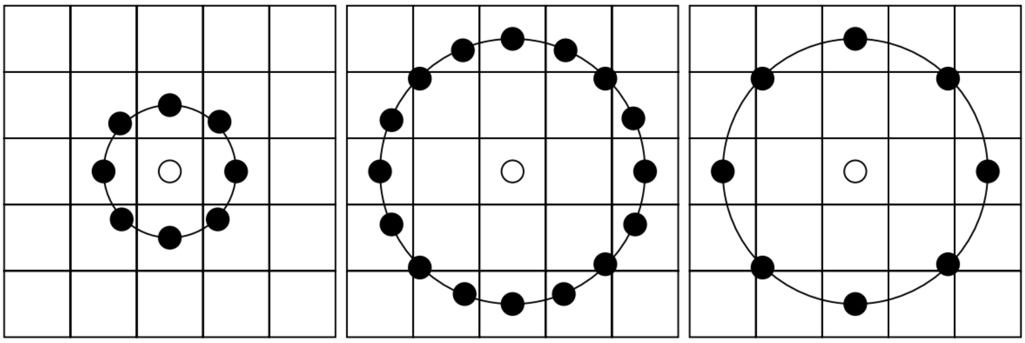


Figure 2-3: The circular (8,1), (16,2) and (8,2) neighborhoods. The pixel values are bilinearly interpolated whenever the sampling point is not in the center of a pixel [61]

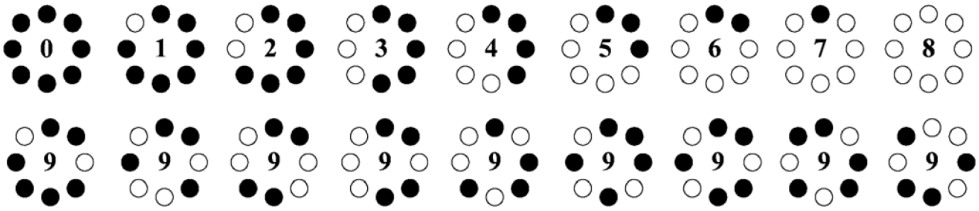


Figure 2-4: LBPs in a circularly symmetric neighboring set of rotation invariant uniform local binary patterns [61]

To obtain the uniform pattern, a uniformity measure is first defined as

$$\begin{aligned}\mathrm{U}(\mathrm{LBP}_{P,R}) &= |s(g_{P-1} - g_c) - s(g_0 - g_c)| \\ &\quad + \sum_{p=1}^{P-1} \left|s(g_p - g_c) - s(g_{p-1} - g_c)\right| \qquad (2.3)\end{aligned}$$

which corresponds to the number of spatial transitions (bitwise 0/1 changes) in the pattern. Based on the uniformity measure, the LBP descriptions of a texture image are defined as follows

$$\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}} = \begin{cases} \sum_{\mathrm{p}=0}^{\mathrm{P}-1} \mathrm{s}(\mathrm{g_p} - \mathrm{g}_c), & \text{if } \mathrm{U}\left(\mathrm{LBP}_{\mathrm{P,R}}\right) \leq 2 \\ \mathrm{P}+1 & \text{otherwise} \end{cases} \qquad (2.4)$$

According to (2.4), LBPs with the U value up to 2 are defined as the $uniform\ patterns$, and its label corresponds to the number of “1” bit in the pattern. Nonuniform patterns are grouped into a category, labelled as $(P+1)$. $\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}}$can be calculated according to (2.4), and superscript “riu2” denotes rotation-invariant uniform patterns with $\mathrm{U} \leq 2$. Hence, $\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}}$ has independent

$P + 2$ output values. For example, $\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}}$with values of (8,1) are shown in Figure 2-4. These uniform patterns represent the microstructures of an image, such as bright spot (0), flat area or dark spot (8), and edges of varying positive and negative curvature (1–7). The pixels in the nonuniform patterns are labelled as 9. After the LBP pattern of each pixel has been identified, a LBP histogram is calculated to represent the texture as follows

$$\mathrm{H(k)} = \sum_{i=0}^{W}\sum_{j=1}^{H} f\left(\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}}(i,j),k\right), k \in [0, K-1] \qquad (2.5)$$

$$\mathrm{f(x,y)} = \begin{cases} 1, & \mathrm{x = y} \\ 0, & \mathrm{otherwise} \end{cases} \qquad (2.6)$$

where K is the number of the patterns equal to $P + 2$ bins. The proportion of the pixels in the nonuniform patterns usually takes a small part in a texture image when accumulated into a histogram. Based on the statistical properties of different patterns, the uniform LBP feature has a strong capability to discriminate textures.

2.2.2. Gray level Co-occurrence matrix

Haralick et al.[52] first introduced the use of co-occurrence probabilities using Gray level co-occurrence matrix (GLCM) for extracting various texture features. Since then GLCM is one of the widely used texture analysis method in image processing. It estimates image properties related to second-order statistics. GLCM describes how often different combinations of gray levels co-occur in an image. Each entry $(\mathrm{i,j})$ in GLCM corresponds to the number of occurrences of the pair of gray levels i and j which are d distance apart in original image. The formal definition of GLCM's is as follows [52].

Suppose an image has $\mathrm{N_x}$ columns, $\mathrm{N_y}$ rows and the gray level appearing at each pixel is quantized to $\mathrm{N_g}$ levels. Let $\mathrm{L_x} = \{1,2,..\mathrm{N_x}\}$ be the columns, $\mathrm{L_y} = \{1,2,..\mathrm{N_y}\}$ be the rows, and $\mathrm{G_x} = \{0,1,2,..\mathrm{N_{g-1}}\}$ be the set of quantized gray levels. The set $\mathrm{L_x} \times \mathrm{L_y}$ is the set of pixels of the image ordered by their row column indices. We used twenty three textural features in our study. Let be $\mathrm{p(i,j)}$ the $\mathrm{(i,j)th}$ entry in a normalized GLCM. The mean and standard deviations for the rows and columns of the matrix are

$$\mu_\mathrm{x} = \sum_\mathrm{i}\sum_\mathrm{j} \mathrm{i.\,p(i,j)}, \mu_\mathrm{y} = \sum_\mathrm{i}\sum_\mathrm{j} \mathrm{j.\,p(i,j)}, \qquad (2.7)$$

$$\sigma_\mathrm{x} = \sum_\mathrm{i}\sum_\mathrm{j} (\mathrm{i} - \mu_\mathrm{x})^2 .\,\mathrm{p(i,j)},\ \sigma_\mathrm{y} = \sum_\mathrm{i}\sum_\mathrm{j} \left(\mathrm{i} - \mu_\mathrm{y}\right)^2 .\,\mathrm{p(i,j)} \qquad (2.8)$$

Some of the basic GLCM features are described below.

Energy:

Energy is also known as “Angular second moment” is the measure of textural uniformity of an image. When gray level distribution has either a constant or a periodic form energy reaches at its highest value. Generally, energy has normalized range therefore; maximum limit of energy is always equal to one. A homogenous image contains very few dominant gray tone transitions, and therefore the p matrix for this image will have fewer entries of larger magnitude resulting in large value for energy feature. If the p matrix contains a large number of small entries, the energy feature will have smaller value. Energy can never be negative.

$$\mathrm{f}_1 = \sum_{\mathrm{i}} \sum_{\mathrm{j}} p(\mathrm{i},\mathrm{j})^2 \tag{2.9}$$

Contrast:

Contrast is a statistic measures the spatial frequency of an image. It is also known as difference moment of GLCM. It measures the amount of local variations present in the image. It is the difference between the highest and the lowest values of a contiguous set of pixels.

$$\mathrm{f}_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(\mathrm{i},\mathrm{j}) \right\} \qquad 2.10)$$

Correlation:

The correlation feature is a measure of gray tone linear dependencies in the image at the specified positions relative to each other.

$$\mathrm{f}_3 = \frac{\sum_{\mathrm{i}} \sum_{\mathrm{j}} (\mathrm{ij}) p(\mathrm{i},\mathrm{j}) - \mu_{\mathrm{x}} \mu_{\mathrm{y}}}{\sigma_{\mathrm{x}} \sigma_{\mathrm{y}}} \tag{2.11}$$

Homogeneity:

It measures image homogeneity as it assumes larger values for smaller gray tone differences in paired elements. A homogeneous scene will contain only a few gray levels, giving a GLCM with only a few but relatively high values. Thus, the sum of squares will be high. GLCM contrast and homogeneity are inversely correlated in terms of equivalent distribution in terms of pixel pairs. It means homogeneity decreases if contrast increases while energy is kept constant. This GLCM statistic is also called as Inverse Difference Moment.

$$\mathrm{f}_4 = \sum_{\mathrm{i}} \sum_{\mathrm{j}} \frac{1}{1 + (\mathrm{i} - \mathrm{j})^2} \mathrm{p}(\mathrm{i},\mathrm{j}) \tag{2.12}$$

Entropy:

Entropy measures the disorder of an image and it achieves its largest value when all elements in P matrix are equal. When the image is not texturally uniform many GLCM elements have very small values, which imply that entropy is very large.

Therefore, entropy is inversely proportional to GLCM energy. Homogeneous scene has high entropy, while inhomogeneous scenes have low first order entropy.

$$\mathrm{f}_5 = \sum_{\mathrm{i}} \sum_{\mathrm{j}} \mathrm{p(i,j)log}\big(\mathrm{p(i,j)}\big) \qquad (2.13)$$

The rest of the textural features are secondary and derived from those listed above.

Autocorrelation:

$$\mathrm{f}_6 = \sum_{\mathrm{i}} \sum_{\mathrm{j}} \mathrm{(i,j)p(i,j)} \qquad (2.14)$$

Dissimilarity:

$$\mathrm{f}_7 = \sum_{\mathrm{i}} \sum_{\mathrm{j}} |\mathrm{i}-\mathrm{j}|.\,\mathrm{p(i,j)} \qquad (2.15)$$

Cluster Shade:

$$\mathrm{f}_8 = \sum_{\mathrm{i}} \sum_{\mathrm{j}} \big(\mathrm{i}+\mathrm{j}-\mu_{\mathrm{x}}-\mu_{\mathrm{y}}\big)^3 \mathrm{p(i,j)} \qquad (2.16)$$

Cluster prominence:

$$\mathrm{f}_9 = \sum_{\mathrm{i}} \sum_{\mathrm{j}} \big(\mathrm{i}+\mathrm{j}-\mu_{\mathrm{x}}-\mu_{\mathrm{y}}\big)^4 \mathrm{p(i,j)} \qquad (2.17)$$

Maximum Probability:

$$\mathrm{f}_{10} = \underset{\mathrm{i,j}}{\mathrm{MAX}}\ \mathrm{p(i,j)} \qquad (2.18)$$

Sum Average:

$$\mathrm{f}_{11} = \sum_{i=2}^{2N_g} i.\,p_{x+y}\,(i) \qquad (2.19)$$

Sum variance:

$$\mathrm{f}_{12} = \sum_{i=2}^{2N_g} (i-f_{13})^2.\,p_{x+y}\,(i) \qquad (2.20)$$

Sum entropy:

$$\mathrm{f}_{13} = \sum_{i=2}^{2N_g} p_{x+y}\,(i)\,\log\{p_{x+y}(i)\} \qquad (2.21)$$

2.2.3. Gabor Wavelet

Use of 2D Gabor wavelet representation in computer vision was pioneered by Daugman in the 1980's [30]. Later on, Gabor features are widely used in various domains to extract information from images [40][31][64]. Gabor wavelet features are exploited in this thesis for textural representation of videos. Gabor wavelet is a set of Gaussian envelope of plane waves, because of its excellent spatial locality and orientation selectivity. The idea is to extract spatial frequencies and local structural characteristics within the local area of the images at multiple directions. This enables us to have certain tolerance on deviations in displacement, deformation, rotation, scaling and illumination. Manjunath et al. in [10] laid the foundations for wide usage of Gabor filters as famous texture descriptor. They also proposed homogeneous texture (HT) descriptor, which was later used as one of the visual texture descriptors in MPEG-7. A two dimensional Gabor function $g(x,y)$and its Fourier transform $G(u,v)$ can be written as:

$$g(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) exp\left(\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi \mathrm{j}\mathrm{W}_x\right)\ and,$$

$$G(u,v) = exp\left(-\frac{1}{2}\left[\frac{(u-W)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right]\right) \qquad (2.22)$$

Where , $\sigma_u = \frac{1}{2\pi\sigma_x}$,$\sigma_v = \frac{1}{2\pi\sigma_y}$ and W is a constant representing the center frequency of the filter bank having the highest frequency. This forms a bandpass filter in the frequency domain. Where center frequency of the filter is directly controlled by the frequency of complex sinusoid. Standard deviation of the Gaussian function controls the bandwidth of this band pass filter. Parameters of Gabor wavelet function controls the Gabor filter bank having a number of bandpass filters with variable center frequencies, bandwidths and orientations.

Given an Image $I(x,y)$, its Gabor wavelet transform is defined as,

$$\mathrm{W}_{mn}(x,y) = \int I(x,y)g_{mn}^*(x-x_1,y-y_1)\,dx_1dy_1 \qquad (2.23)$$

where $\mathrm{W}_{mn}(x,y)$ is the filter response at the spatial location (x,y), $m = 1,2,\ldots M$ is the number of scales and $n = 1,2,\ldots N$ is the number of orientations. Here, g^* specifies the complex conjugate. A fair assumption is proposed by Manjunath et al. [10] that local image regions are spatially homogeneous. There (average) mean and (variance) standard deviation of the magnitude of the filter responses are used to represent the region for classification purposes,

$$\mu_{\mathrm{mn}} = \int\int |\mathrm{W}_{mn}(xy)|dx_1dy_1\ ,$$

$$\sigma_{mn} = \sqrt{\int\int (|\mathrm{W}_{mn}(xy)| - \mu_{\mathrm{mn}})^2 dx_1dy_1} \qquad (2.24)$$

Final feature vector is thus constructed using μ_{mn} and σ_{mn}, as feature components also known as HT descriptor. Where μ_{mn} is the mean and σ_{mn} in equation (2.24) is standard deviation of the magnitude of transform coefficients.

2.2.4. Motion Detection

Usually, motion in video sequence occurs due to motion of the camera e.g., camera panning, zooming or from displacements of individual objects in the scene. Camera movement's results in Global Motion (GM) while the motion of the object in the scene results in local motion. In case of spoofing detection motion estimation can play an essential role in classification. Therefore, numerous motion estimation algorithms have been proposed in the literature [3][18][38][42][56].

Most motion estimation techniques make no distinction between the global and local motion. Global motions (GM) in a video sequence produced by camera displacement are modelled by parametric transforms of two-dimensional (2-D) images . The process of estimating the transform parameters is called Global Motion Estimation (GME). GME is an important tool widely used in computer vision, video processing, and many other fields. Dufaux et al. [18] and Etoh et al.[38] proposed different techniques for GME. Mainly, motion estimation techniques are divided into two categories i.e. feature-based and intensity-based approaches. In feature-based approach, motion is estimated by representing images into corners, edges or more complex structure features defined by the SIFT algorithm [42] and further transformed parameters are estimated. However due to incorrect feature detection, noise and feature matching issues motion estimation may result erroneous [42]. Intensity-based motion estimation techniques are further divided into two groups i.e., block based approaches and frame based approaches. Block based techniques utilize block-based motion vectors (MVs) estimate a global motion field [39]. MVs are calculated from local blocks of two consecutive frames. Frame based approach uses entire frames and the intensities of the frames are subtracted. However these approached also have limitations. In case of block based approaches incorrectly estimated MVs may lead to a distortion of the estimated global parameters. While frame based approaches are accurate but computationally very expensive. Reader interested in overview of Gauss-Newton (GN) gradient-descent technique for motion estimation can refer to [54]. A more efficient version of the GN algorithm called the inverse compositional algorithm (ICA) is proposed in [55] .

Handhold counterfeit evidence shown to biometric systems have GMs problem, which could be classified easily using GME. However, in case of fixed attacks GME will overlap the both classes. Figure 2-5 illustrates the output of frame based global motion estimation on different spoofing videos.

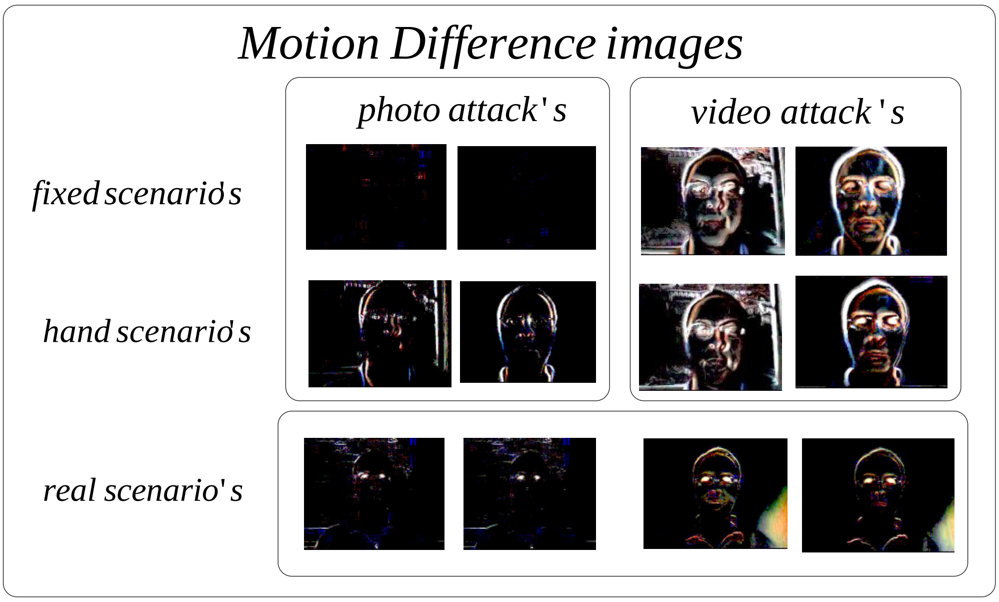


Figure 2-5: Examples of the motion difference video frames of different input samples

2.2.5. Fixed Pattern Noise and Artifacts

The first task performed on any facial biometric system is the data acquisition to authenticate the user. This is performed by a camera that has an imaging sensor with thousands of photosensitive transducers capable of converting light energy into electrical charges. The camera lenses allow light reflected by the objects in the scene to focus on the imaging sensor, transforming light energy into electrical charges, which are converted into digital signals by an A/D converter [8]. During this process of transforming an analog signal into a digital signal, the appearance of noise in the resulting image is inevitable. The analysis of noise in images has been widely explored in the digital document forensic analysis area, more specifically, the problem of identifying the specific camera that acquired a document. In this case, the main goal is to estimate the type and manufacturer of the cameras with just one image. Lukas et al. [33] discuss two types of noise present in images: the fixed pattern noise (FPN) and the noise resulting from the photo-responsiveness of non-uniform light-sensitive cells (PRNU). FPN noise is produced by the presence of dark currents that can be defined by accumulated electrons in the inverse joints of the light-sensitive cell pins of the imaging sensor. Formally, FPN (also called nonuniformity) is the spatial variation in pixel output values under uniform illumination due to device and interconnect parameter variational mismatches across the sensor. It is fixed for a given sensor, but varies from sensor to sensor. On the other hand, PRNU noise is defined by the difference in sensitivity of the light sensitive cells caused by the non-homogeneity of the silicon wafer and other imperfections inserted during the manufacturing process of the sensor [8].

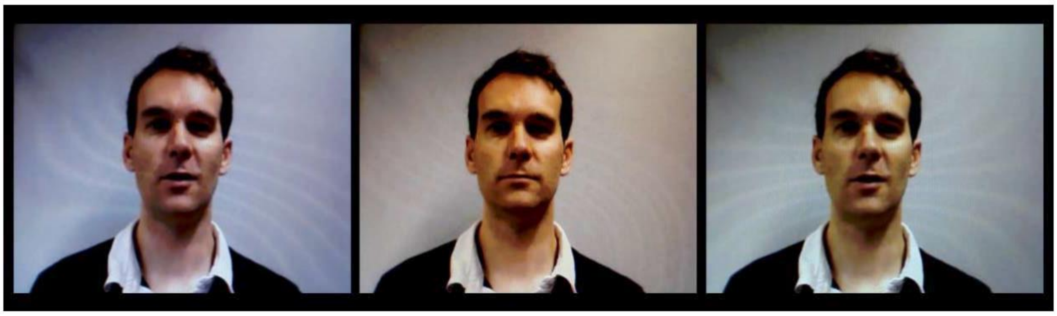


Figure 2-6: Moiring effects in videos shown on three different monitors and captured with a digital camera [7]

Another noticeable fact is the appearance of artifacts generated by means of videos captured from other videos, which do not exist in videos generated from the capture of real scenes. These artifacts are generated mainly during the process of creation and exhibition of the frames on monitor screens, producing undesirable effects such as distortion, flickering, moiring, among others [7]. Figure 2-6 shows the moiring effect in recaptured videos. This is mainly due to different screen frequencies (refresh rates) of three different monitors captured by digital camera [7]. Thus, spoofing attempts submitted to biometric systems (referred to as attack videos) will likely have more noise and artifacts than the biometric samples captured directly from live people (referred to as valid videos).

2.3. Classification Techniques

The goal of classification in general is to select the most appropriate category for an unknown object, given a set of known classes. Since perfect classification has been often impossible, the classification may also be done by specifying the probability for each of the known categories. Classifiers are traditionally divided into two categories: parametric and non-parametric. Both parametric and non-parametric classifiers need some knowledge of the data, be it either training samples or parameters of the assumed feature distributions. They are therefore called supervised techniques. With non-supervised techniques, classes are to be found with no prior knowledge. The classifiers opted in this thesis are mainly based on supervised learning. A supervised learning algorithm analyses the training data and produces an inferred function, which can be used for mapping new unseen examples.

2.3.1. Support Vector Machines

Support vector machines (SVMs) proposed by Boser et al. [9] have been successfully used in many learning problems and it has mostly outperformed other supervised learning algorithms in recent years. When applied to classification, SVMs seek the optimal separating hyper plane between two classes, typically in a higher dimensional space than the original feature vectors. SVMs are often referred as large margin classifiers due to their ability of learning the hyper planes that distinct the nearest training samples (support vectors) with the largest possible margins in higher-dimensional features space. The distance between the support vectors and separating hyper plane is called the margin of the classifier. The aim of SVMs is to decide the parameters of a mapping function that can map all the training samples to some real valued functions which separates them efficiently.

SVMs are inherently designed to solve the binary classification problems. For multiclass problems the commonly used technique is “divide and conquer” in which a single multiclass problem is divided into binary pairs and then a SVM is trained for each pair. Such techniques are known as One-versus-One and One versus-Rest, comparative study about these techniques can be found in [14][17]. SVMs can be linear or non-linear. Linear in the cases when data points are linearly separable, SVMs can also utilize kernel functions, the approach named as kernel tricking, to transform the features into a higher-dimensional space. This trick allows the formulation of nonlinear variants of any algorithm and cast them in terms of dot products. The goal of kernel tricking is to make the features linearly separable. In the case the samples are not linearly separable; cost functions are used to penalize the function for allowing data samples to exist on the wrong side of the hyper plane.

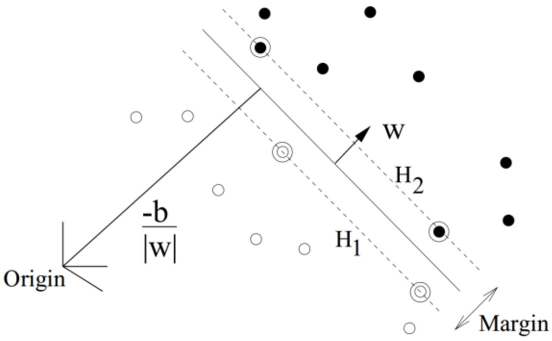


Figure 2-7: Separating hyper plane for the linearly separable classes (Linear SVMs) [12]

Training SVM means to finds a hyper plane that separates the labelled training example with maximum margin. Given, the labelled training samples $\{\mathbf{x_i}, \mathbf{y_i}\}_{\mathbf{i=1 \ldots m}}$ where $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{d}}$ and $\mathbf{y} \in \{\mathbf{-1}, \mathbf{+1}\}$, the goal of SVMs is to learn a decision function $\mathbf{f}(\boldsymbol{x}, \boldsymbol{\alpha})$ that maps any arbitrary input $\boldsymbol{x_b}$, with function parameters α, to a real value closer to its original label. If the training samples are linearly separable, the hyper plane that separates both classes is defined by the points satisfying $x_i . w + b = 0$. Consider all training samples satisfy the following constraints:

$$y_i(w. x_i + b) \geq 1, \forall\, i = 1 \ldots m, \tag{2.25}$$

The margin of such classifier is defined by two separate decision planes, H_1 and H_2 as shown in Figure 2-7, and is equal to $2/||\mathrm{w}||$. While w is the normal to hyper plane and the parameter b is offset of the hyper plane from the origin. Several alternate set of the parameters w and b can be found that correctly classify the training samples for one single problem. However, the classifier with lower margin expects to have a higher expected error and vice versa. Hence, to maximize the margin, the objective can be formed to minimize the Euclidean norm $\|w\|$ with the constraints in equation (2.25). For computational ease and to generalize the same formulation for nonlinear case this problem can also be expressed in form of Lagrange function [12] as follows:

$$L_p = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i y_i (x_i . w + b) + \sum_{i=1}^{m} \alpha_i , \qquad (2.26)$$

where α_i are the Lagrange multipliers. This problem requires minimization of a convex objective function [12]. The problem can be reduced even more by representing the normal vector as $w = \sum_{i=1}^{m} \alpha_i y_i x_i$ and under the constraints $\alpha_i \geq 0$ and $\sum_{i=0}^{m} \alpha_i y_i = 0$. Substituting these constraints and w, in equation (2.26) will provide a new dual form of Langrange function as,

$$L_D = \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j \ . \qquad (2.27)$$

To generalize the function for nonlinear cases the dot product in equation (2.27) can be replaced by the kernel function k as follows:

$$L_D = \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j \, k(x_i, x_j). \qquad (2.28)$$

As mentioned before, the kernel function is used to transform the feature vectors into higher dimensional feature space to make the problem linearly separable. The most famous kernel functions used in SVMs framework are Radial Basis Function (RBF), Polynomial and Sigmoid. However, there are some situations where the RBF kernel is not suitable. In particular, when the number of features is very large, one may just use the linear kernel. Moreover, training RBF kernel requires more time than linear SVM. For better results one must strive for optimal C (penalty parameter) and γ (kernel parameters). One way of finding optimal parameters is “grid-search” using cross-validation. Readers more curious about details may refer to [12].

2.3.2. Partial Least Square Regression

Partial Least Squares (PLS) is a method for modelling relations between sets of observed variables by means of latent variables. It comprises of regression and classification tasks as well as dimension reduction techniques and modelling tools. Partial least squares regression (PLS) is used to describe the relationship between multiple response variables and predictors through the latent variables. PLS regression can analyse data with strongly collinear, noisy, and numerous X-variables, and also simultaneously model several response variables, Y [66]. PLS regression is the most ideal technique to analyse, when the number of observations is much smaller than the number of X-variables in the data set. PLS regression has been paid an increasing attention these days as an importance measure of each explanatory variable or predictor. PLS regression model with two matrices, $X(n \times k,\ predictors)$ and $Y(n \times m,\ responses)$ can be expressed as follows:

$$\mathrm{X} = \mathrm{TP}^{\mathrm{T}} + \mathrm{E} \qquad (2.29)$$

$$\mathrm{Y} = \mathrm{UQ}^{\mathrm{T}} + \mathrm{F} \qquad (2.30)$$

$$u_a \ = b_a t_a + \mathrm{h} \ \ \mathrm{a} = 1, \dots, \mathrm{A} \qquad (2.31)$$

where $\mathrm{T} = (t_1, \dots, t_A)$ and $\mathrm{U} = (u_1, \dots, u_A)$ are latent variable scores of X and Y, respectively, and P and Q are the corresponding loadings, where A is the number oflatent variables. Equations (2.29) (2.30) represent the outer relations of X and Y, (2.31) is the inner relation between two score matrices, and b_a is the regression coefficients of inner relation. The matrices **E** and **F** represent error terms associated with X and Y, respectively, whereas **h** means random error vector in the inner relation. In classic form PLS method, is based on the nonlinear iterative partial least squares (NIPALS) algorithm [21]. The number of latent variables is an important parameter in PLS regression and it can be determined by considering the proportion of variance explained by each latent variable. Usually, it is done by a cross-validation such that the predicted error is minimized.

PLS model can be rewritten to look as a multiple regression model [66]. By using equation (2.33) multiple linear regression coefficients can be estimated from the PLS regression model parameters. Those coefficients describe an increase of a particular Y-variable as a change of a particular X -variable when the other X -variables are fixed. By controlling X -variable with a large coefficient tightly a small variation of related Y-variable can be expected. The beta coefficients $\mathrm{B}_{\mathrm{PLS}}$ can be obtained by considering the equivalent following multiple linear regression models;

$$\mathrm{Y} = \mathrm{X}\,\mathrm{B}_{\mathrm{PLS}} + \varepsilon \qquad (2.32)$$

They can be derived from the PLS regression model since there exists the following relationships between the quantities derived through NIPALS algorithm.

$$Y = UQ^T = TBQ^T \tag{2.33}$$

where B is a matrix whose i-th diagonal element is b_i Since

$$T = XW(P^TW)^{-1} = TBQ^T \tag{2.34}$$

$$Q^T = B^{-1}(T^TT)^{-1} = TY \tag{2.35}$$

Equation (2.33) reduces to

$$Y = XW(P^TW)^{-1}(T^TT)^{-1}T^TY \tag{2.36}$$

Therefore, B_{PLS} can be expressed as follows

$$B_{PLS} = W(P^TW)^{-1}(T^TT)^{-1}T^TY \tag{2.37}$$

The contribution of each X-variable to a response variable can be measured by decomposing the sum of squares (SS) of the response variables. Sum of squares of an n-vector x and n-by-k matrix X is defined by Equation (2.38) and (2.39) respectively,

$$SS(x) = x^tx = \sum_{i=1}^{n} x_i^2 \tag{2.38}$$

$$SS(X) = \sum_{j=1}^{k} SS(x) \tag{2.39}$$

The total sum of squares can be further divided into SS of regression (SSR) and SS of error (SSE),

$$SST = SS(Y) = SSR + SSE \tag{2.40}$$

Here SSR is the minimum of SS of latent, which is shown in Equation (2.41)

$$SSR = \sum_{a=1}^{A} SS(b_a t_a q_a^T) = \sum_{a=1}^{A} b_a^2 SS(t_a) = \sum_{a=1}^{A} SSR_a \tag{2.41}$$

The combination of PLS with SVMs has been studied in [52]. However, in this thesis SVMs and PLS are used separately for classification task.

2.4. Literature Review

Nowadays we are experiencing an increasing demand for highly secure identification and personal verification technologies. This demand becomes even more ostensible as we become aware of new security breaches and transaction frauds [19]. Anti-spoofing solution for biometric system is very recent research area and there is not much work available in this field, especially because often new intimidations arrive in the form of better, more refined and sophisticated spoofing attacks.

Schwartz et al. [65] categorized current anti-spoofing methods into four groups: data driven characterization, user behavior modeling, user interaction need, and presence of additional devices. Possible solutions to this problem may be engaging additional devices such as deploying an additional 2-D camera, depth camera, thermal sensor, or implementing a human computer interaction interface asking the user to make a particular gesture for authentication. Since, such solutions are intrusive and may not be feasible in the existing systems. So, there is an imminent need to introduce an approach for detecting the spoofing attempts without any additional hardware.

2.4.1. Data Driven Categorization

Considering the group of data-driven characterization methods, some anti-spoofing techniques for facial recognition systems rely on Fourier analysis. Some researchers explored the high frequencies of Fourier spectra in order to collect features to differentiate between live faces and certain types of spoofs, such as printed images.

Other used data-driven approaches include the surface texture of the facial skin from which we can calculate certain measures to characterize optical qualities of the facial skin of live people and compare them to the non-live ones and optical-flow analysis. Assuming the region of analysis as a 2-D plane, Bao et al, obtained a reference field from the actual flow field data on live and non-live images pointing out their differences. Another solution based on optical-flow analysis was presented by Kollreider et al. In their work, the authors described two approaches: one using a data-driven characterization that estimates the face motion based on optical flow analysis over selected frames and a second solution exploring a model-based local Gabor decomposition used in conjunction with SVM experts for face part detection.

Tan et al. [67], proposed a solution based on extracting Difference-of-Gaussian (DoG) and variational retinex features to estimate the Lambertian reflectance properties and distinguish between valid and fake users on NUAA Database [67]. Kollreider et al. [34] used a heuristic classifier based on optical flow analysis that evaluates the trajectories of selected parts of a face region. Anjos et al. in [1] presented a motion-based solution that detects correlations between the person's head movements and the scene context. Pinto et al. in [7] proposed a face classification method based on Gray Level Co-occurrence Matrix (GLCM) feature after extracting noise signatures and calculating the Fourier spectrum on logarithmic scale to create visual rhythms in spoofed videos. Schwartz et al. in [65] presented an anti-spoofing solution based on a set of low level descriptor Histogram of Oriented Gradients (HOG), GLCM and Histograms of Shearlet Coefficients (HSC) using partial least square regression. Kose et al. [44] in proposed an anti-spoofing solution based on textural and contrast measure using Local Binary Patterns Variance (LBPV) with global matching. Chingovska et al. in [28] tested the variants of LBP features on face regions concluded that histogram of Uniform Local Binary Patterns produced the best result. Similar work was proposed by Pereira et al. in [62] against face spoofing attacks using the LBP−TOP (LBP from Three Orthogonal Planes) descriptor combining both space and time information into a single descriptor.

In IJCB 2011 Competition on Counter Measures to 2D Facial Spoofing Attacks [13] , a common trend was set to use multiple anti-spoofing measures combining motion, liveness and texture and the participants were able to achieve impressive results. However, this competition was dealing with only photo and print attacks therefore all best-performing algorithms used also some sort of texture analysis. But in recently organized ICB 2013 2nd competition on countermeasures to 2D facial spoofing attacks [25] a diverse data set was considered including photo, mobile videos, highdef videos and print attacks, best-performing algorithms used texture and motion analysis together to achieve state-of-art results.

2.4.2. User Behaviour Categorization

For the group of approaches counting on the user behaviour in front of the camera, some researchers have focused on motion detection such as eye blinking [12,14] and involuntary movements of parts of the face and head [9, 15]. Koll-reider et al. [10] introduced a technique for motion analysis with applications for spoofing detection using the notion of quantized angle features ("quangles") and machine learning classifiers. Pan et al. [22] proposed a real-time liveness detection approach against photograph spoofing, by conditional modelling of spontaneous eye blinks. The later work by the same authors [24] proposed counter-measure, which include a background context matching that helps avoiding video-spoofing in fixed face biometric systems.

One problem with some of the previously mentioned approaches is that they are still impacted by small head tilts which simulate head movement or by short video sequences displaying an authentic user. If we count on the user behaviour and also require his/her involvement, we can take advantage of multi-modal information (e.g., voice or gesture) and various challenge-response methods such as asking the user to blink the eyes in a given order, or even smile [5, 13].

2.4.3. User Interaction Categorization

Considering human interaction with biometric system can be a good solution for this authentication problem. Possible solution can be made requiring the user to have particular interaction with the system. S. Trewin et al. [59] suggested the possible solution for biometrics system asking the user for voice verification combined with face verification or gesture verification can also be combined with face module to authenticate valid user. Though spoofing attempts can be reduced with these solutions but the less intrusiveness of face validation is reduced, because we want the systems with less user interaction and perfect verification.

2.4.4. Engaging Additional Hardware

Possible solutions to this problem may be engaging additional devices such as deploying an additional 2-D camera, depth camera, thermal sensor along with the face verification system. The simplest way can be engaging the light source along with the camera. The video captured will have face some extra lighting effects, based on the reflective properties we can categorize the spoof and real video. The screen, iPad, or a mobile device because of the glass which reflects light back to the camera. As mentioned earlier, most of the current face recognition systems are based on intensity images and equipped with a generic camera. An anti-spoofing method without additional device is more preferable such that, it could be easily integrated into the existing face recognition systems.

3. IMPLEMENTATION

This chapter discusses the details of the implementation of the proposed counter measures for face biometric systems. It follows the implementation details of multi-view face detection module, pre-processing steps of input samples followed by region of interest selection strategy. Finally, three different counter measures are presented resulting in state-of-art results.

3.1. Face Detection

Face detection is an essential first-step for any face recognition systems. It also has several applications in areas such as content-based image retrieval, crowd surveillance and automated important person detection. Considering the authentication systems face plays most important part hence, various face detection algorithms have been proposed in the literature. Face detection algorithm proposed by Viola and Jones [50] which exploits the boosting algorithm [69] for learning a strong classifier is one of the most famous face detection algorithms. Their system, based on integral image and simple features, promised very high speed and performance comparable to all the previously existing systems. The original algorithm proposed in [50] is based on Haar-like features. The integral image representation and cascade architecture of classifiers is used for a computationally efficient implementation. However, using almost the same algorithm, several other features have been used in the literature such as an extended version of Haar-like by Lienhart et al. [51]. Lienhart et al. proposed the rotated Haar-like features for detecting the inplane rotated faces. Similarly, LBP features based face detector was proposed by S. Liao et al. [58]. However, the already available implementation of the face detector, available in OpenCV [20], is used in this thesis. The OpenCV implementation of the face detector offers two different features; extended Haar-like features [51] and Multi-Block Local Binary patterns (MB-LBP) [58]. MB-LBP features based face detection is chosen in this work due to its computational effectiveness compared to Haar features, and these features allows robustness against illumination changes. A brief overview of MB-LBP features used in implementation of face detectors are explained next.

3.1.1. Multi-block Local Binary Patterns

As local binary patterns is explained in Section 2.2.1, MB-LBP works on the same principle. However, instead of considering each pixel, MB-LBP operates on the rectangular block regions. Average intensity value g_c is compared with average intensities of eight neighbouring rectangles as illustrated in Figure 3-1. The final MB-LBP code is defined as:

$$\text{MB- LBP} = \sum_{i=1}^{8} \text{f}(g_i - g_c)2^n, \tag{3.1}$$

where g_i is the average intensity value of neighbouring pixels ($i = 1, \ldots ,8$) and $f(x)$ is defined as follows:

$$\text{f}(x) = \begin{cases} 1, & if\ x > 0 \\ 0, & if\ x < 0 \end{cases}. \tag{3.2}$$

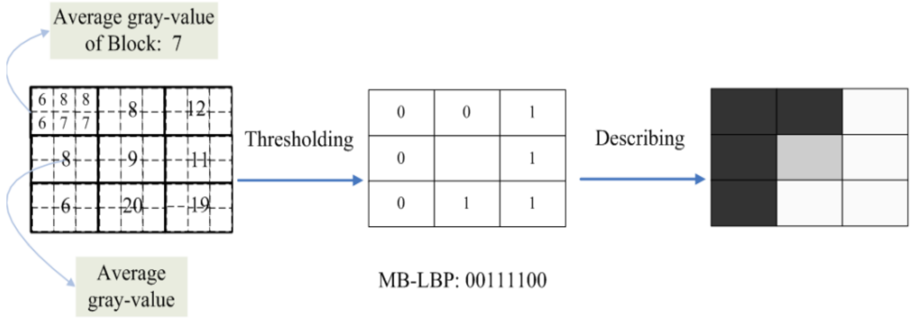


Figure 3-1: Illustration of the Multi-Block LBP [36]

As working of MB-LBP is illustrated in Figure 3-1, in total $2^8 = 256$ unique MB-LBP codes can be obtained. These MB-LBP codes are directly fed as features in face detection process. Given a patch of size 20 × 20 dimensions, 2049 MB-LBP features are computed that are further used for classification of face and non-face regions [36]. However, all these 2049 features are not useful and most of them are redundant hence, boosting approach is used for choosing the most discriminating features for classification. We utilized Adaboost for this purpose however Gentle Adaboost and Real Adaboost [51] can also be used.

Cascade Architecture

The cascade-structured classifier of Viola et al. [50] has been proved very efficient for many object and face detection problems. A sliding window approach is used to detect every possible face from images, which considers every patch of size N × N for classification. Face in real world images can vary in different sizes therefore; this process is repeated at different scales. This results in a huge amount of work load which is not feasible for real-time applications. Even if we consider that image contains only one face it is observable that an unnecessary large amount is spend in evaluation of sub-windows would still result in negatives (non-faces regions). So the algorithm should work in a way that it concentrate on discarding non-faces rapidly and spend more on time on possible face regions. To encounter this problem, Viola and Jones [50] proposed an efficient cascading algorithm. It consists of a cascade of classifiers which significantly decrease the computation time, and also ensures better face detection accuracy. The cascade architecture consists of M stages and at each stage a boosted classifier is trained. The job of each stage is used to determine whether a given sub window is definitely not a face or may be a face. A given sub window is immediately discarded as not a face if it fails in any of the stage. A simple illustration of the cascade architecture can be seen in Figure 3-2. First set of simple classifiers at early stages are used to reject most of the background regions, and more complicated and sophisticated classifiers are utilized in later stages. In this way only strong candidates of face will go in advanced stages and more complex classifiers will be used only for these candidates. For more detailed information on face detection, the readers should refer to [25][36][50][68].

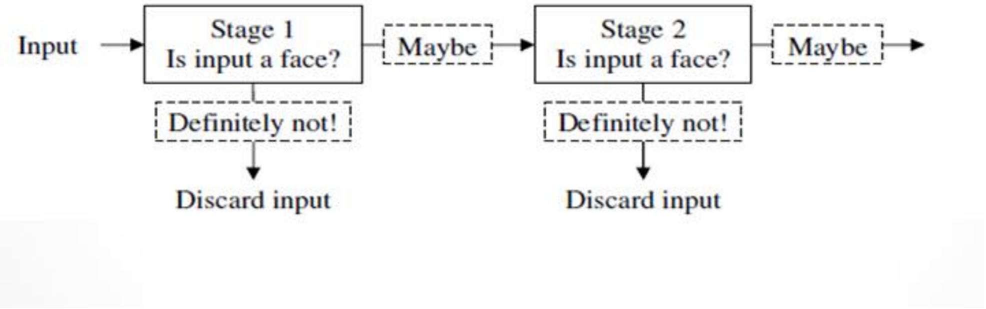


Figure 3-2: Cascade architecture

3.1.2. Multi-view Face Detection

Most of the face verification systems utilize the face matches in various angles and pose of the client for better recognition performance. Hence, a pose invariant face detector robust to illumination changes capable of detecting faces from various pose angles is needed. One possible way to get such detector can be training a single detector with training samples from different yaw angles. However, this may result into generalization problem for the face detector and it will not be able to generalize any face angle. Moreover, training one single detector for all the poses will put too much burden on the classifier which will also result in its non-applicability to real time systems. The rest of this section explains the opted approach in detail. Several existing Multi-view face detection (MVFD) frameworks have been proposed in the literature [26], [41],[71]. However, a simple yet effective approach is adopted in this thesis which utilizes several trained face detectors at dedicated face angles. Several MB-LBP based face detectors are trained for the following yaw-angles:

$$\theta = \{0°, \pm 15°, \pm 30°, \pm 45°, \pm 60°\}$$

For training of face detectors, training images are gathered from various datasets that also provide information about the face pose. These datasets include FERET [49], PIE [63], Prima Head Pose [43], BioId [48], and AR face dataset [6]. Figure 3-3 shows some positive training samples used for training nine face detectors at different yaw angles. To further increase the training set for the positive class, more training samples are generated by applying simple geometrical transformation utility available in OpenCV to the existing images such as changes in scale, position, rotation, translation, etc. [32].

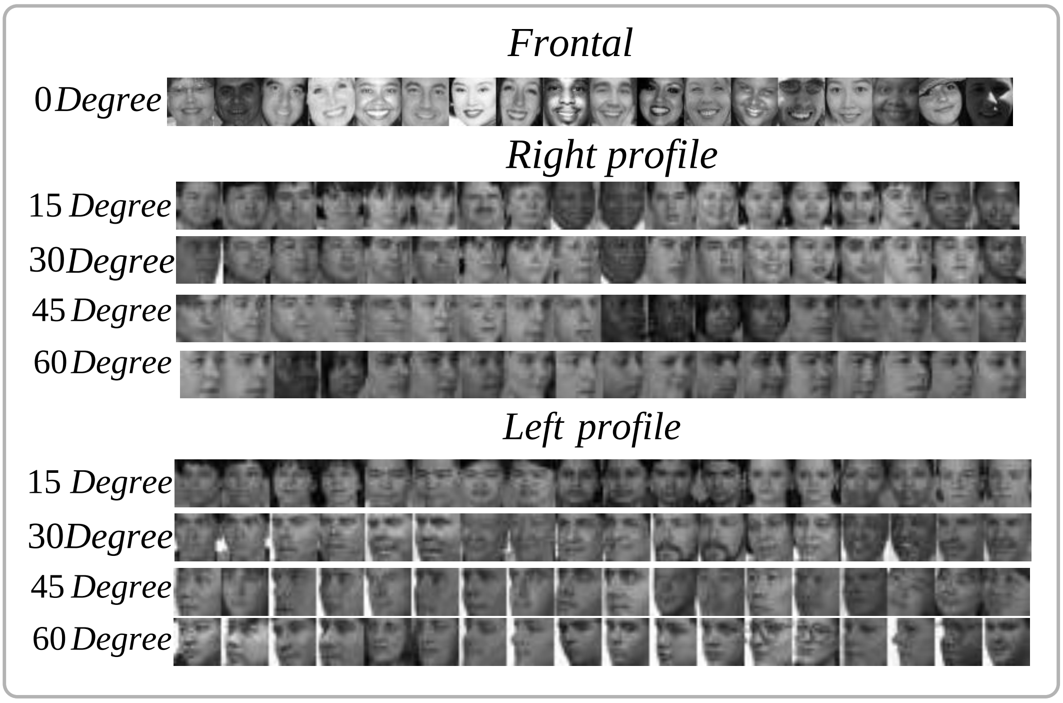


Figure 3-3: Samples used for training of face detectors.

During face detection, the input image is scanned by all view-based detectors and the outputs are merged. The scanning procedure is an exhaustive search, involving a lot of sub-windows. As the classifiers uses sliding window approach and are insensitive to small localization errors, each detector gives several overlapping detections around face regions. Another reason is that these detectors run at different scales to classify a single image-path as face and non-face. We considered only the final detection results from each specific view detector and further grouped the results of all detectors to form final outcome. It is quite likely that the face detected by one specific view face detector also detects the faces at other view angles. For instance, face detector trained at 15 degree will most detect the frontal faces in most of the cases. This enhances the ability of the system to consider that patch as face region. Figure 3-4 shows example results of MVFD detections where, same faces have been detected by different view based detectors. Further, these overlapping detection of all the detector are combined to form the final output of MVFD system. Face detector training process requires a lot of processing time therefore, due to the limitation of positive training samples and time these face detectors were trained with different number of positive samples resulting in different number of cascade stages. The negative samples used for training of detectors were constant 15000. Table 1 shows number of stages trained for each view specific face detector at different number of positive training images.

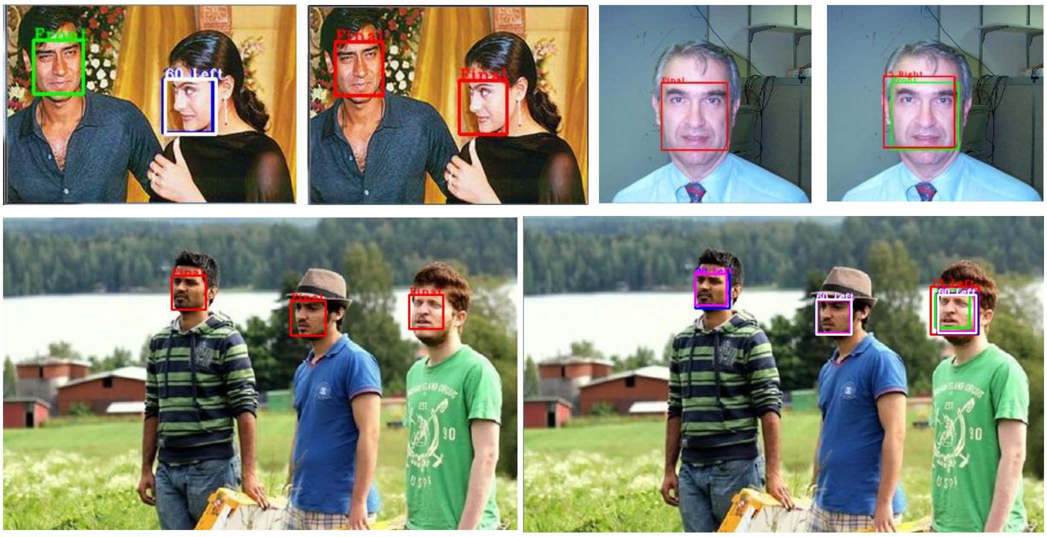


Figure 3-4 Example of the Multi view face detection

Table 1: Specification of trained face detectors

Face Detectors	Cascade Stages	Positive Samples
Frontal	22	8500
Left Profile 15 Degree	17	7000
Right Profile 15 Degree	17	7000
Left Profile 30 Degree	18	7300
Right Profile 30 Degree	18	7300
Left Profile 45 Degree	17	7000
Right Profile 45 Degree	17	7200
Left Profile 60 Degree	19	7800
Right Profile 60 Degree	19	7800

3.2. Pre-Processing of Videos

All the input samples from datasets were re-encoded to Audio Video Interleave (AVI) file format from QuickTime Movie (MOV) file format at bit rate 576 kbps without changing the resolution of the video. Extensive set of experiment was performed to find the optimal solution for anti-spoofing. Detailed descriptions about the datasets used are explained in Section 4.1.

3.3. Region of interest selection

As aforementioned, face detection is pivotal part for any face verification systems. Considering this fact, most of the previously counter measures using on textural analysis [28][62][65] takes analysis only over the face region and thus they are directly dependent on the face detection. Due to the fact that face detection is an erroneous process, such an approach may lead to performance degradations. Besides that we observed that, there are crucial clues around the face region that can contribute to boost the accuracy of the spoof classification. Since, FPN and PRNU [33] induced in the recapturing process are more discriminative on the surrounding region. To validate this key observation we performed a small experiment extracting histogram of $\mathrm{LBP}_{P,R}^{riu2}$ on full image scene and only on the face region resulting from face detector. Experimental setup and classification results is shown in the Figure 3-5 and **Table 2** respectively.

Therefore, all the proposed methods in this thesis analyze the entire image, as it improves classification of real and spoofing images/videos.

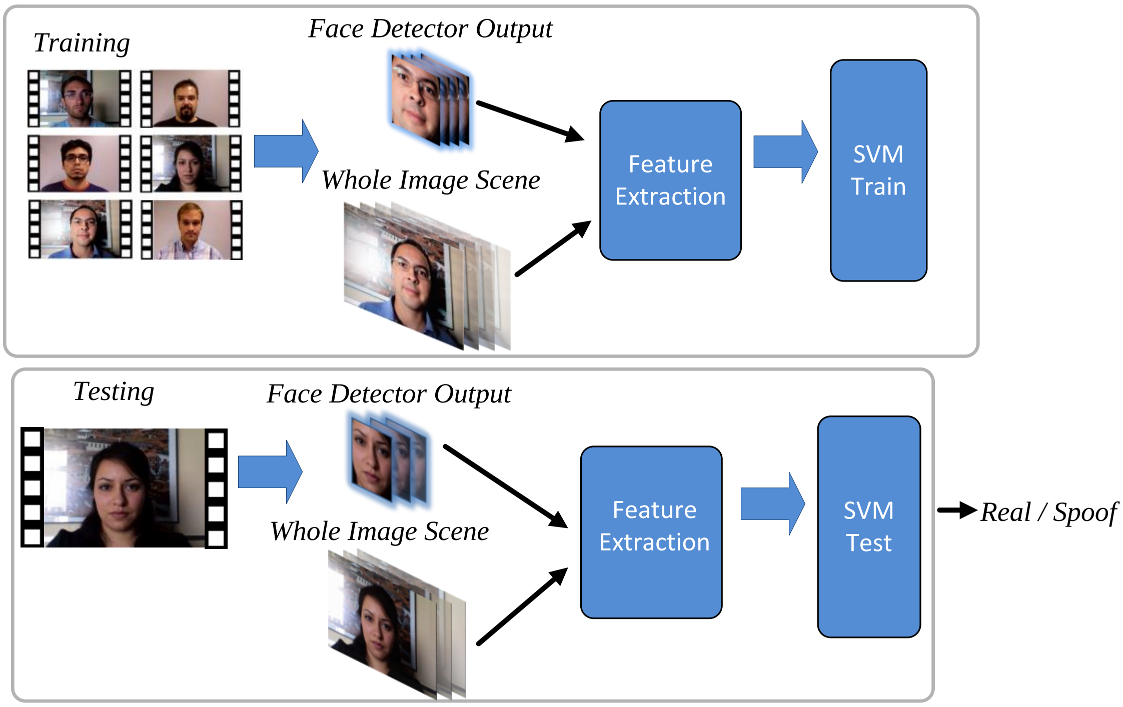


Figure 3-5: Region on Interest selection experimental setup

Table 2: Classification performance of Rotation Invariant Uniform Local Binary Patterns on face region versus entire image.

Feature Extraction	Classification Results SVM
$\mathrm{LBP}_{P,R}^{riu2}$ (Face Region)	85.21%
$\mathrm{LBP}_{P,R}^{riu2}$ (Entire Image)	97.50%

3.4. Counter Measures

Different counter measures were proposed in this thesis. The rest of this section gives the implementation details of texture based approach, motion magnification and space-time auto-correlation of gradients (STACOG) and combined motion and texture based approach.

3.4.1. Texture based Counter Measures

First, approach implemented in this thesis is solely based on textural analysis of the input videos that captures the noise and textural differences resulting in state-of-art classification results. To capture textural characteristics of non-live video sequence based on textural features namely, $\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}}$, Gabor, GLCM and their different variations.

A video, V, can be defined as a 2-D sequence of N frames, each frame as a function $f(x,y)$ of luminous intensities. Rotation-Invariant Uniform LBP feature vector $f_1(V)$ for a video V is computed by averaging LBP histograms of all N frames of the video. According to equation (1), histogram per frame is individually computed as follows:

$$\mathrm{f}_1(\mathrm{V}) = \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{\mathrm{hist}\left(\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}}\left(\mathrm{I_i}(\mathrm{x},\mathrm{y})\right)\right)}{\mathrm{N}} \tag{3.3}$$

Similarly, for the video V its Gabor feature vector $f_2(V)$ is constructed by averaging N Gabor feature vectors computed on all N frames. We extract Gabor wavelet features with 4 scales, $\mathrm{S} = 4$ and 6 orientations, $\mathrm{K} = 6$. The feature vector is then constructed using μ_{mn} and $\sigma_{\mathrm{mn}}$where μ_{mn} in equation (5) is the mean and σ_{mn} in equation (6) is standard deviation of the magnitude of transform coefficients [10].

$$\mathrm{f_{Gab}}\left(\mathrm{I}(x,y)\right) = [\mu_{00}\ \sigma_{00}\mu_{01}\ \sigma_{01}\ \ldots\ \mu_{35}\ \sigma_{35}] \tag{3.4}$$

$$\mathrm{f}_2(\mathrm{V}) = \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{\mathrm{f_{Gab}}\left(\mathrm{I_i}(\mathrm{x},\mathrm{y})\right)}{\mathrm{N}} \tag{3.5}$$

Finally, the GLCM feature $f_3(V)$ for every video is computed by averaging N feature vectors. The feature vector of the GLCM for single image $\mathrm{I}(x,y)$ is formed as follows:

$$\mathrm{f_{GLCM}}\left(\mathrm{I}(x,y)\right) = [\mathrm{r}_1\mathrm{r}_2\mathrm{r}_3 \ldots \mathrm{r}_{23}] \tag{3.6}$$

To take benefit of rich textural information retrieved from above three features obtained using equations (3.3), (3.4) and (3.5), three more features were formed by simple approach of concatenation. Figure 3-6 shows the process of acquiring the features using equations (3.3), (3.4) and (3.5); formation of the feature vectors by concatenation; and finally the classification of real and spoof attacks based on concatenated feature vector. For classification of real and spoofing attempts we used SVMs and PLS regression. Results of all these features with both these classifiers are illustrated in Chapter 4.

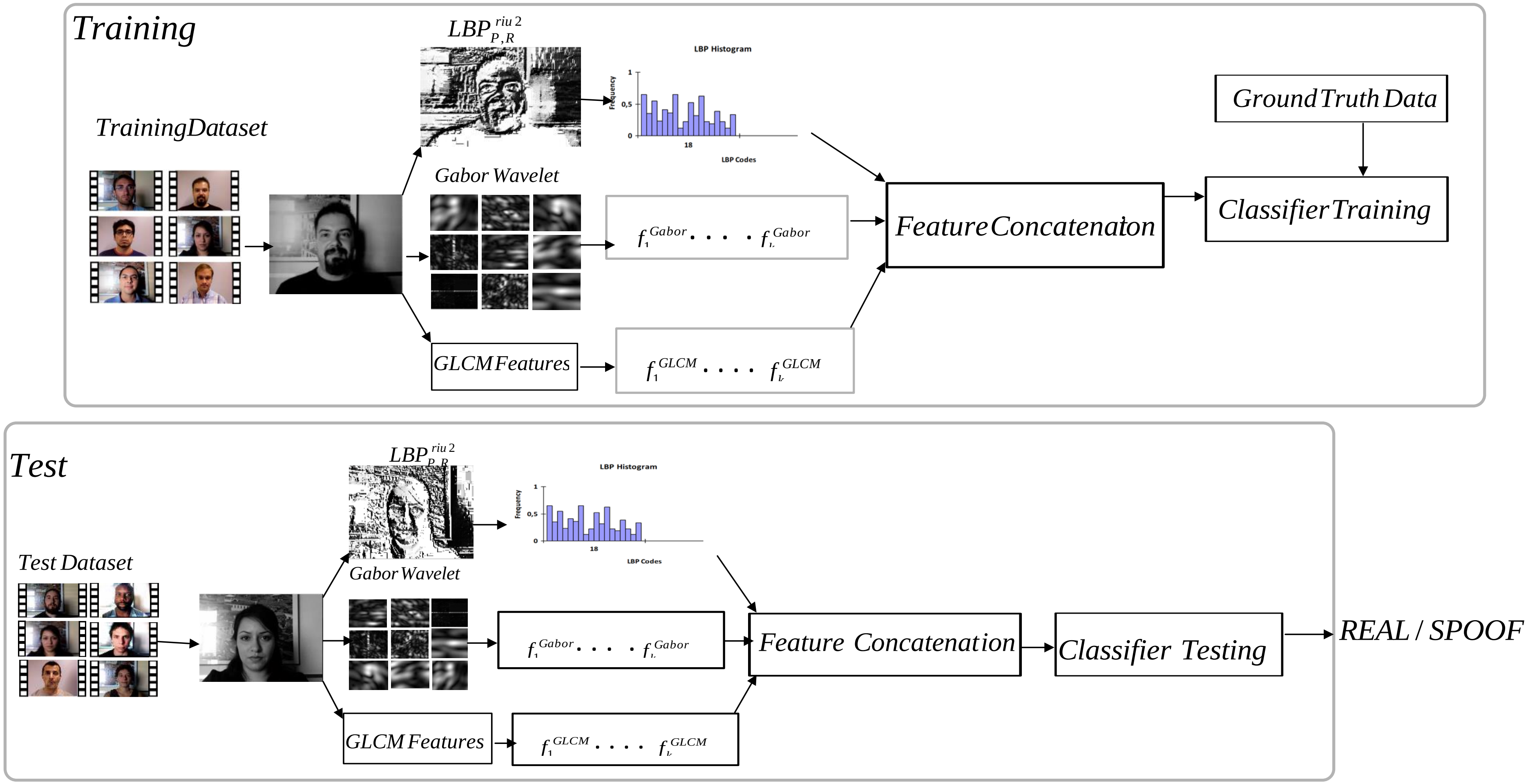


Figure 3-6: Block diagram of texture based Anti-Spoofing experiments

3.4.2. Motion based Counter Measure

In this section the proposed motion based countermeasure based on motion magnification and motion feature extraction is discussed in detail. Motion in both these classes undergoes different flow. Our hypothesis is that these motion variations can be useful in classification. Hence, a better representation of motion features that is robust to intruder attempts is needed as some of the motion patterns can overlap between these samples. Moreover it is our assertion that the performance of spoofing detection techniques can be improved with motion magnification as it might enhance the liveness nature of the face video. This can help in differentiating spoofing attempts under global and local motion variations. Therefore, as a pre-processing step, first motion magnification is applied to all video samples. Later on, feature extraction method utilizing auto-correlations of space–time gradients (STACOG) of three-dimensional motion shape in a video sequence are applied to the each sample. The overall architecture for motion based countermeasure can be seen in Figure 3-7. The details of motion magnification and STACOG features are illustrated below.

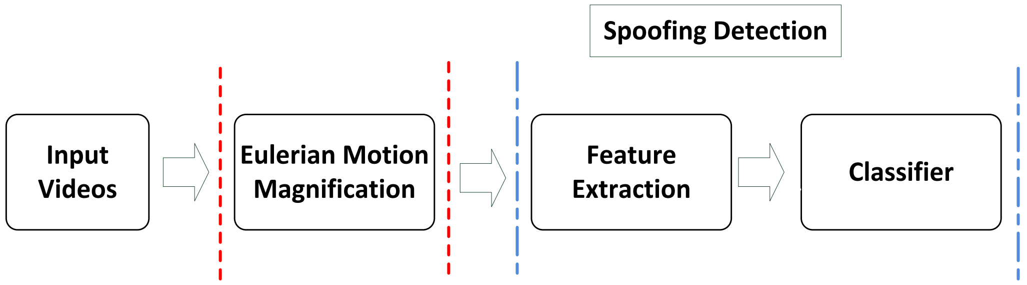


Figure 3-7: Motion based experimental setup

Eulerian Motion Magnification:

As mentioned before, motion magnification can play a vital role in enhancing the liveness nature of real samples, and also can enhance the global motion, noise signature present face spoofed videos. Given a video sequence, for motion magnification we used an Eulerian approach [27]. This Eulerian motion magnification is more than real time, therefore can be easily integrated in existing systems. Previous techniques of motion magnification explicitly tracks pixel’s trajectory over time. Therefore, these approaches are computationally expensive. On the other hand, Eulerian approach directly amplifies temporal intensity changes at a given position without the need for explicit estimation. Using appropriate temporal and spatial filtering, the desired motion is localized and then magnified under Taylor expansion assumption.

First, the video sequence is decomposed into different spatial frequency bands. Because they may exhibit different signal-to-noise ratios, they should be magnified differently. In general case, the full Laplacian pyramid may be computed. Then, the temporal processing is performed on each spatial band. An ideal temporal bandpass filter is applied to each Laplacian band to isolate the desired temporal motion in each band. For eye-lid movement's frequency band is chosen to 0.2-0.5 Hz [24]. The isolated bandpassed signal is then multiplied by an amplification factor α and added to the original signal, as shown in Equation (3.7).

$$\hat{I}(x, y, t) = I(x, y, t) + \alpha B(x, y, t) \tag{3.7}$$

Where $B(x, y, t)$ is the output of a bandpassed filter for video $I(x, y, t)$, at positions x, y, t. Finally, after multiplication with the amplification factor the decomposed Laplacian bands are reconstructed to form the output video. The magnification factor α is appropriately attenuated with respect to a spatial cut-off frequency(λ_c), so as to reduce α for bands of higher frequencies. This minimizes the artifacts in the resultant video. An optimal value of α should be chosen by visual inspection of processed videos from the training set because magnification is dependent on the filter and the magnification factor α. Therefore, magnification factor α was chosen to be 10 throughout entire dataset after visually inspecting the few videos from dataset. The approach enhances facial movements including subtle motion such as blinking, and conjugates eye motion that may otherwise only be visible on close inspection of the video. In case of spoofing videos the distortion, flickering, moiring are enhanced. The block diagram of Eulerian motion magnification is shown in Figure 3-8. As mentioned before, to capture the temporal differences between the real and imposter attempts, motion features based on co-occurrence histogram were computed. Detailed description about motion features are described next.

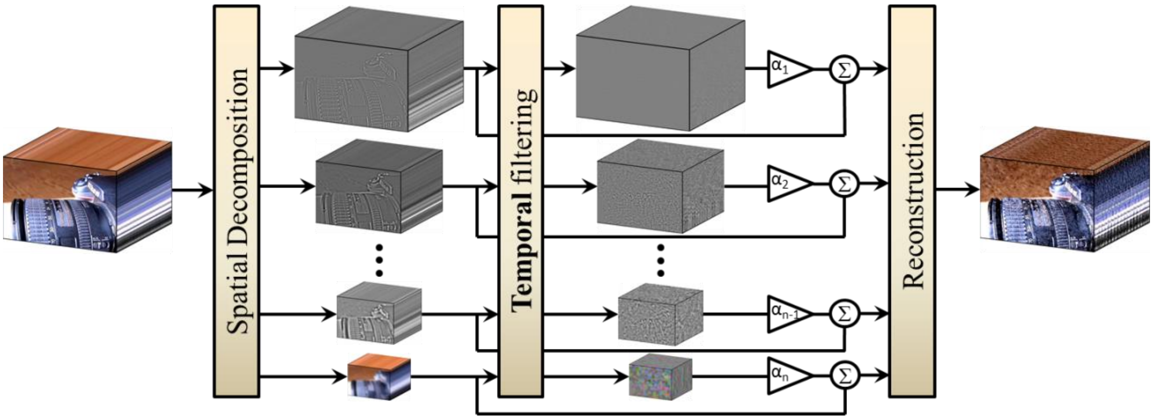


Figure 3-8 Overview of the Eulerian video magnification framework [27]

STACOG:

After the motion of the video sequences is magnified, motion features [51] based on co-occurrence histograms of the space-time 3D gradient orientations are extracted. They are employed for frame based features to densely characterize the motion. These frame-based features are extracted from sub sequences densely sampled along the time axis. Extracted features are computational faster than real time. In addition, STACOG has the property of shift-invariance which is desirable for recognition. The space–time (three-dimensional) gradient vector is calculated by derivatives $\left(\mathrm{I_x}, \mathrm{I_y}, \mathrm{I_t}\right)$ of motion image volumes $\mathrm{I}(\mathrm{x}, \mathrm{y}, \mathrm{t})$ at each space-time point in an image sequence. As shown in Figure 3-9 (Left), the gradient vectors can be geometrically represented by the magnitudes $\mathrm{m} = \sqrt{I_x^2 + I_y^2 + I_t^2}$ and two types of angle: spatial orientation $\theta = \arctan(\mathrm{I_x}, \mathrm{I_y})$ in an image frame and temporal elevation $\theta = \arcsin(\mathrm{I_t}/\mathrm{m})$ along the time axis, where the functions arctan and arcsin output the angles within $[0, 2\pi]$ and $[-\pi/2, \pi/2]$, respectively. The space-time orientation of the gradient defined by these two angles is coded into B orientation bins on a unit sphere by voting weights to the nearest bins as illustrated in Figure 3-9 (Right). Then, the orientation is finally described by a B-dimensional vector h, called space time orientation coding (STOC) vector [60].

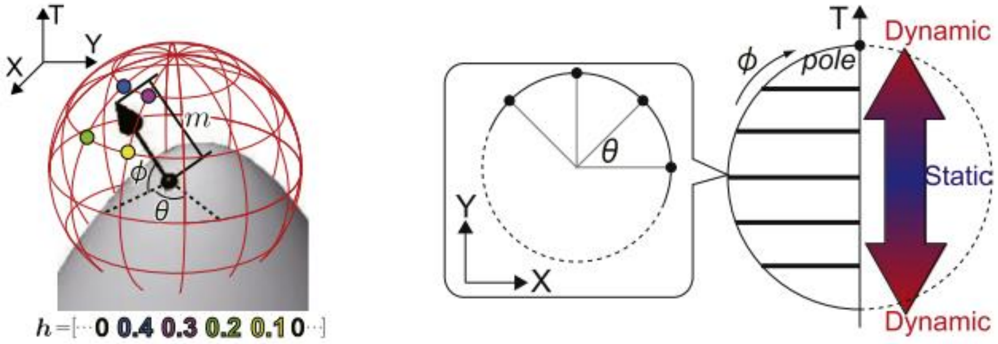


Figure 3-9 Space-time gradient of equi-gray level surface (Left) ; Orientation bins along latitude and longitude (Right)[60]

Since, non-live videos has motion differences, and normally the motion patterns throughout the videos sequences are similar, these features provides us enough categorization power to classify real and attacking attempts using supervised learning methods as described in Section.2.3.

3.4.3. Image Quality based Counter Measures

Final, approach implemented in this thesis is solely based on image quality assessment (IQA) of the input videos that uses scene statistics of locally normalized luminance coefficients to quantify possible losses of "naturalness" in the image resulting in reasonable classification results. These losses in naturalness occur due to distortions induced during recapturing process in case of non-live videos.

Traditionally, IQA is divided into two types, No-reference and Full-reference. Objective blind or No-reference image quality assessment ($\mathrm{NR} - \mathrm{IQA}$) refers to automatic quality assessment of an image using an algorithm such that the only information that the algorithm receives before it makes a prediction on quality is the distorted image whose quality is being assessed. On the other end of the spectrum lie full-reference (FR) algorithms that require as input not only the distorted image, but also a clean reference image with respect to which the quality of the distorted image is assessed.

Normally, a reference image is not available for the verification system the user can have different clothes, expression every time he ask for access. So, the scene is never similar and in case of portable verification system background scene can also changes. Therefore, $\mathrm{FR} - \mathrm{IQA}$ is not the fesiable approach to proceed in case of spoofing detection. We used $\mathrm{NR} - \mathrm{IQA}$ as illustrated in [5]. Given a video V, for each image I we modelled the statistical relationships between neighbouring pixels and structure of pixels were modelled using the empirical distributions of pairwise products of neighbouring MSCN (mean subtracted contrast normalized) coefficients. For a given intensity image $\mathrm{I(i,j)}$, its MSCN coefiicients can be computed using Equation (3.8).

$$\hat{\mathrm{I}}(\mathrm{i,j}) = \frac{\mathrm{I(i,j)} - \mu(\mathrm{i,j})}{\sigma(\mathrm{i,j}) + \mathrm{C}} \tag{3.8}$$

where, C is a constant 1 to prevent demoninator value to become zero. μ and σ are local mean field and local variance field (standard deviation) respectively.

$$\mu(\mathrm{i,j}) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l}\, I_{k,l}(i,j) \tag{3.9}$$

$$\sigma(\mathrm{i,j}) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l}(I_{k,l}(i,j) - \mu(\mathrm{i,j}))^2} \tag{3.10}$$

where w is circular-symmetric Gaussian weighting function. MSCN were modelled along four orientations; main-diagonal $(\mathrm{D_1})$ secondary-diagonal $(\mathrm{D_2})$, horizontal (H) and veritcal (V) as illustrated in Figure 3-10.

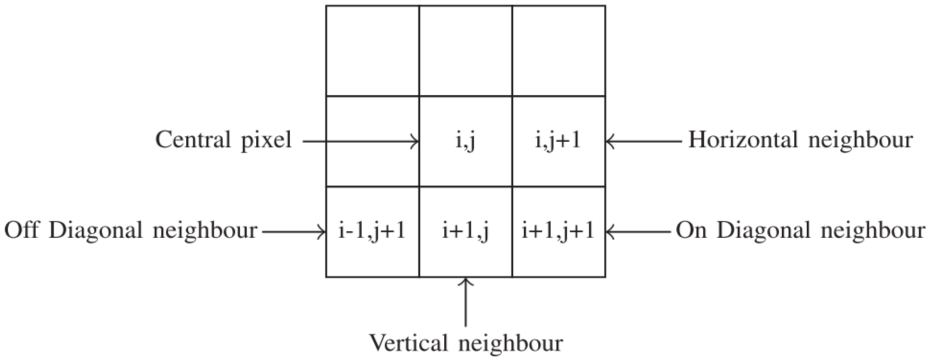
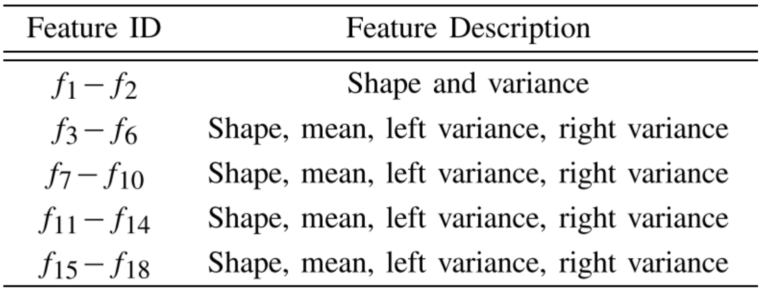


Figure 3-10 various paired products computed in order to quantify neighboring statistical relationship[60]

Thus for each paired product, 16 parameters i.e., 4 parameters/orientation × 4 orientations are computed, yielding the set of features. Table 3 summarizes the $\mathrm{NR} - \mathrm{IQA}$ features utilized,

Table 3: Summary of IQ features extracted at one scale[60]



Feature ID	Feature Description
$f_1 - f_2$	Shape and variance
$f_3 - f_6$	Shape, mean, left variance, right variance
$f_7 - f_{10}$	Shape, mean, left variance, right variance
$f_{11} - f_{14}$	Shape, mean, left variance, right variance
$f_{15} - f_{18}$	Shape, mean, left variance, right variance

Usually, images encounter distortion effects across scales. Therefore, we extracted all features listed above in Table 3 at three different scales - the original image scale, at a down sampled versions by a factor of 2 and 4 respectively. Thus, a total of 54 features, are used to capture distortions and to perform distortion-specific image quality assessment. The system works on a frame-by-frame basis. Final feature vector is formed by averaging all N obtained feature vectors from each video. These features are than fed to the supervised classification process using SVMs and PLS separately. Results of all these proposed approaches are illustrated in Chapter 4.

4. EXPERIMENTAL RESULTS

In this chapter, we evaluate the performance of our proposed anti-spoofing solutions for face verification systems using two different face spoofing datasets. First, evaluation is based on a publically available dataset and second on local dataset with diverse scenarios. Rest of the chapter includes details of two datasets used, followed by quantitative results comparisons of the proposed countermeasures. The chapter concludes with, the analysis of computational complexity of the proposed solution.

4.1. Dataset Details

As mentioned before, there has not been much work done in the field of anti-spoofing for face verification systems, therefore not enough datasets are available except REPLY- ATTACK [48] and NUAA dataset. We have used REPLY- ATTACK which is the state of art benchmark database for evaluation purposes of our proposed techniques. Moreover, we collected the dataset where multiple videos of same client were recorded using different cameras and view angles. Experiments were conducted on both these datasets having different properties and complexities which are described in detail below.

4.1.1. Reply-Attack Dataset

The Replay-Attack face spoofing database [28] consists of short video recorded clips of both real and attack attempts of 50 different clients. This database mainly consists of three different types of attacks: ***printed photographs***, ***digital photographs*** displayed on the screen of a device and ***dynamic video scenes*** replayed on the screen of a device. In dynamic video scenes a client having either head, a body or eye blink motion makes it the most difficult samples for classification. Moreover, in this dataset, attacks are divided into two groups: ***fixed*** (the attack device is fixed to the support so they do not move during the spoof) and ***hand*** (the attacker holds the attack media with his/her hands). Furthermore, the photo and video replay attacks can be of lower quality (taken with an iPhone and displayed on an iPhone screen) and of high quality (displayed on an iPad screen). The lightening conditions during recordings is also categorized in two different type; ***controlled*** (fluorescent lamp is used to illuminate the scene) and ***adverse*** (scene is captured in day light). Canon PowerShot SX150 IS 12.1 megapixels, iPhone 3GS 3.1 megapixel camera were used to capture pictures and videos in highdef and mobile categories respectively. All together, these variants introduce even larger variety in the spoofing attacks present in the database in comparison to other publically available datasets.

The total number of videos in the database is 1200 (360 in the training set, 360 in the development set and 480 in the test set). The resolution of each video is 320 (width) x 240 (height) pixels with a frame rate of 25 frames-per-second and contains 240 frames for each attack videos and 375 frames for each real access video [28]. Real and spoofed videos scenario's for one client in Reply-Attack dataset are illustrated in Figure 4-1.

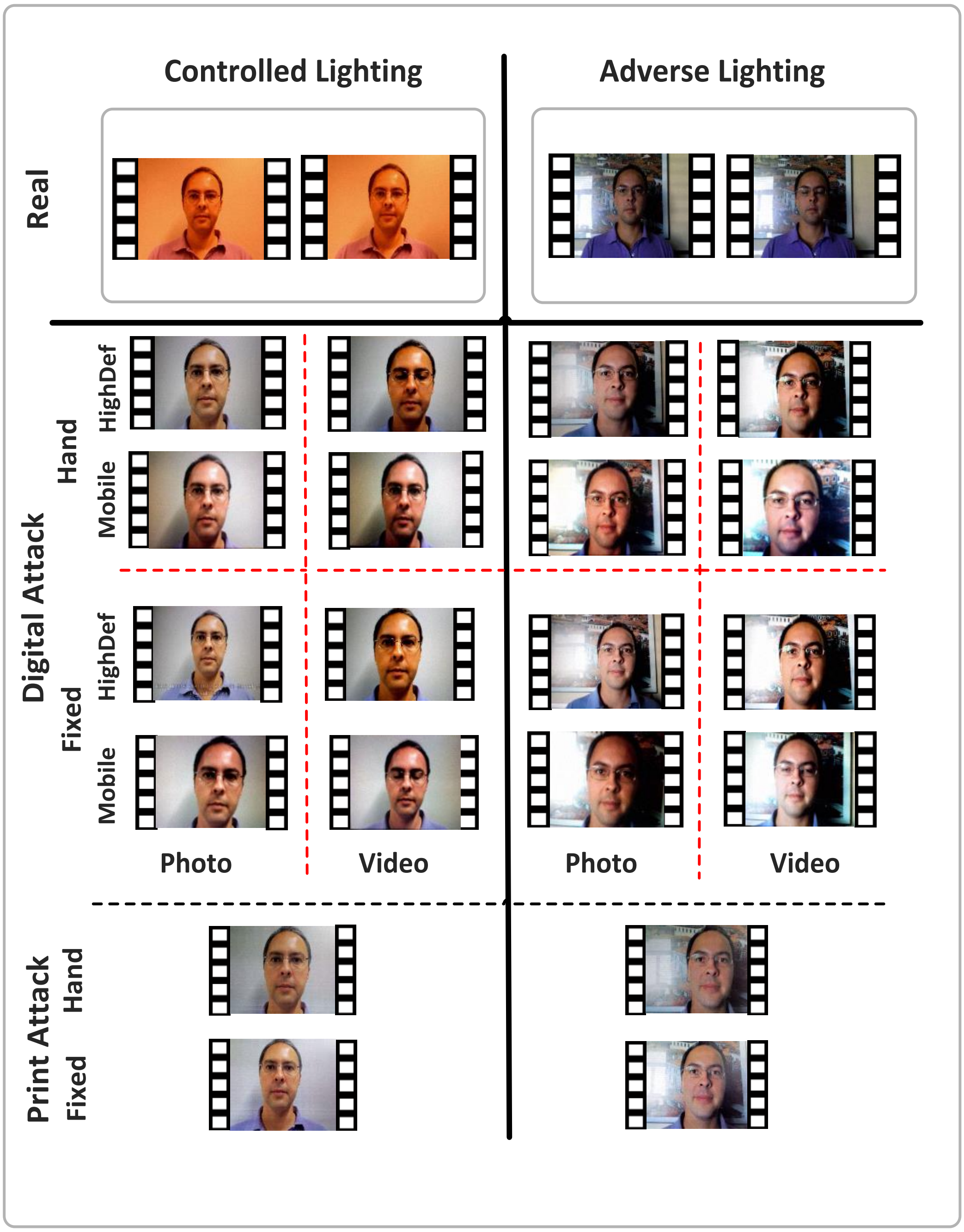


Figure 4-1: Few video frames from the Replay Attack database.

4.1.2. Local Dataset

To validate the effectiveness of the solutions, a small but very diverse dataset was created. The dataset consisted of 12 real videos of 3 different users. These videos were captured using four different cameras (e.g., Nokia Lumia 920 Pure view camera, Apple iPad camera, Canon-D 550 DSLR and HP Probook webcam). To collect more spoofing data for training, these videos were displayed on HP LP-2065 LCD screen and recaptured again using the above mentioned camera sources resulting in 48 imposter attempt videos. The length of the videos varies from 9 seconds up to 12 seconds, with different frame rates depending on the camera.

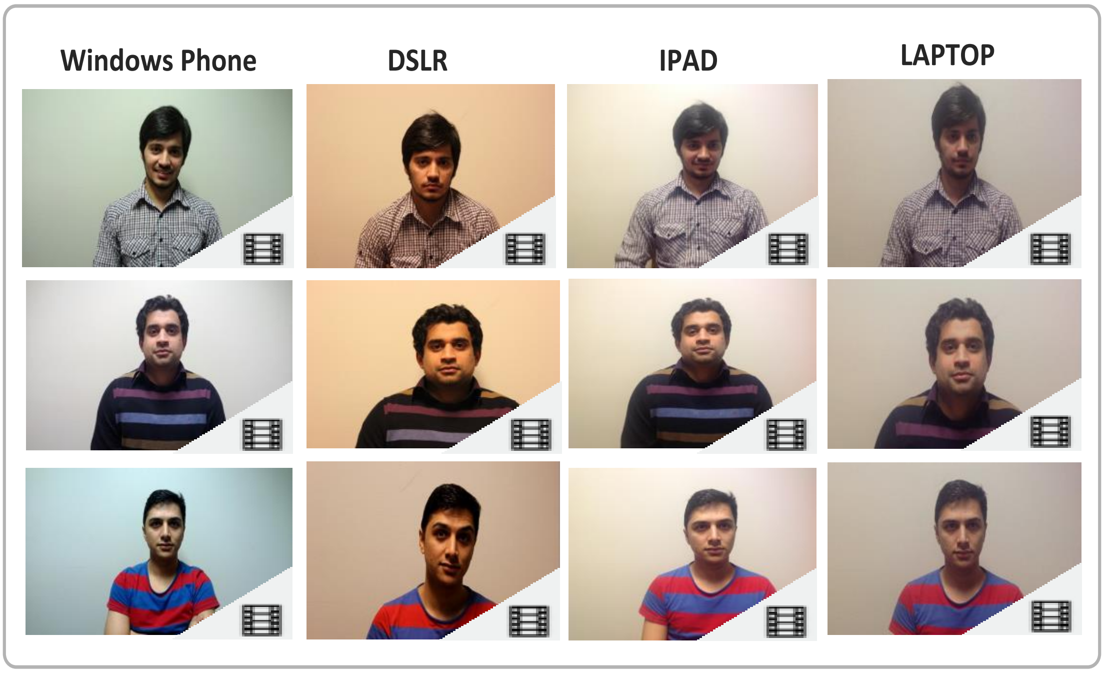


Figure 4-2 Examples of the real video frames of local dataset

4.2. Performance Metric

A spoofing detection system is subjected to two types of errors, either the real access is rejected (false rejection) or an attack is accepted (false acceptance). Its performance is often measured with Half Total Error Rate (HTER), which is half of the sum of False Rejection Rate (FRR), and the False Acceptance Rate (FAR),

$$\text{False Acceptance Rate} = \frac{\text{Number of imposter attempts classified as real attempts}}{\text{Total number of imposter attempts}}, \qquad (4.1)$$

$$\text{False Rejection Rate} = \frac{\text{Number of real attempts classified as imposter attempts}}{\text{Total number of real attempts}}, \qquad (4.2)$$

$$\text{Half Total Error Rate} = \frac{\text{False Rejection Rate} + \text{False Acceptance Rate}}{2}, \qquad (4.3)$$

Since both FAR and FRR depend on a threshold τ, increasing the FAR will usually reduce the FRR and vice-versa. For this reason, results are often presented using the Receiver Operating Characteristic (ROC) curve, which plots the FAR versus the FRR for different values of τ.

4.3. Classification Results

This section discusses the classification accuracies and HTER of all the proposed approaches. The performances of these countermeasures were evaluated on both aforementioned datasets. Different features were tested individually as well as by combining multiple features together as described in Chapter 3.

4.3.1. Results for Reply Attack dataset

To examine the effectiveness of proposed solutions, two types of experiments were performed. The first experiment included the training of two different classifiers using the entire training set, the results are shown in Table 4.

Table 4: Results of different features after training over full dataset

Features (Training over Full dataset)	**Classification Rate (SVM)**	**Classification Rate (PLS)**
$\text{LBP}_{P,R}^{riu2}$	98.54%	99.37%
Gabor	94.16 %	96.25%
GLCM	91.04%	94.04%
$\text{LBP}_{P,R}^{riu2}$ + Gabor	98.12%	100.00%
$\text{LBP}_{P,R}^{riu2}$ + GLCM	97.91%	100.00%
GLCM + Gabor	94.79%	96.04%
Motion magnification + $\text{LBP}_{P,R}^{riu2}$	100.00%	100.00%
Motion magnification + STACOG	93.54 %	95.21%

For further validation of the fact that these proposed counter measures provides state of art results, a more challenging experimental setup was created which used half of the training data of each sub category of attack, real videos while testing dataset was kept as is. Table 5 provides the results of counter measures over half training dataset. Results are evaluated for all methods used in this thesis, using the performance metrics explained in Section 4.2.

Table 5: Results of different features after training over half training dataset

Features **(Training over Half dataset)**	**HTER** **(SVM)**	**HTER** **(PLS)**
$\mathrm{LBP}_{P,R}^{riu2}$	4.50 %	2.25%
Gabor	9.25 %	10.62%
GLCM	26.25%	10.12%
$\mathrm{LBP}_{P,R}^{riu2}$ + Gabor	4.37%	0.00%
$\mathrm{LBP}_{P,R}^{riu2}$ + GLCM	3.25%	0.12%
GLCM + Gabor	9.25%	10.37%
Motion magnification + $\mathrm{LBP}_{P,R}^{riu2}$	0.00%	0.00%
Motion magnification + STACOG	10.85 %	9.12%

We can see how the final results depict the effectiveness of all approaches. The success of using motion magnification as pre-processing step can also be seen by the increase in classification rate and decrease in the HTER for texture based counter measure, from 4.50% to 0.00%. The experimental results also shows that Gabor features along with $\mathrm{LBP}_{P,R}^{riu2}$ are more reliable for capturing the textural differences between both classes. The results for STACOG based counter measures are not significant as others, because, most of the misclassified real videos were having global motion which occurred during the dataset collection. In some cases real videos were treated as fixed printed attack because there was no motion at all from the clients, not even eye blinks.

4.3.2. Results for Local dataset

For the local data set, we again evaluated the results for all the aforementioned counter measures in terms of HTER rates. However the classification rates are not mentioned. Table summarizes the results obtained over the local dataset using all methods.

Table 6: Results of different features over local dataset

Features **(Training over Half dataset)**	**HTER** **(SVM)**	**HTER** **(PLS)**
$\mathrm{LBP}_{P,R}^{riu2}$	4.37%	2.25%
Gabor	13.45%	14.65%

GLCM	20.34%	12.79%
$\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}}$ + Gabor	3.35%	0.00%
$\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}}$ + GLCM	6.34%	2.25%
GLCM + Gabor	12.65%	12.65%
Motion magnification + $\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}}$	0.12%	0.00%
Motion magnification + STACOG	17.43 %	10.37 %

We can see that all the approaches again worked well. These results demonstrate the effectiveness of the proposed algorithms. As this dataset is captured in different illumination, background conditions, and appearance variations.

4.3.3. Results for Anti-Spoofing Competition

In 2nd Competition on Counter Measures to 2D Face Spoofing Attacks [29] eight teams from different countries participated and proposed different counter measures. We (MUVIS) adopted only the texture- based approach by extracting two texture features: $\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}}$ and Gabor in 4 scales and 6 orientations as described in Chapter 3. To capture enough textural differences proposed solution operated on computing the texture features of whole frame region. In both cases, the feature vector on video-level is computed as average of the feature vectors on frame-level. The result of the concatenation of the two feature vectors is fed into Partial Least Square regression classifier.

Table 7: Performance results for the anti-spoofing algorithms proposed in 2nd competition on anti-spoofing 2013(in %)[29]

	Development			Test		
Teams	**FAR**	**FRR**	**HTER**	**FAR**	**FRR**	**HTER**
CASIA	0.00	0.00	0.00	0.00	0.00	0.00
LNMIIT	0.00	0.00	0.00	0.00	0.00	0.00
MUVIS	**0.00**	**0.00**	**0.00**	**0.00**	**2.50**	**1.25**
PRALAB	0.00	0.00	0.00	0.00	2.50	1.25
MaskDown	1.00	0.00	0.50	0.00	5.00	2.50
IGD	5.00	8.33	6.67	17.00	1.25	9.13
ATVS	1.67	0.00	0.83	2.75	21.25	12.00
Unicamp	13.00	6.67	9.83	12.50	18.75	15.62

The algorithms were trained and evaluated considering all the types of attacks in the REPLY- ATTACK database. Results were evaluated based on HTER measured on the anonymized test set using a threshold calculated ‘a priori’ on the development set. The threshold, which was chosen using the Equal Error Rate (EER) criterion, is the value equalizing FAR and FRR. The performance figures are given for both development and anonymized test set as shown in Table 7.

4.4. Proposed Framework

Usually, 2D face biometrics system are designed in a way such that they can distinguish between two classes: valid as positive and invalid users as a negative class. However, due to rapid increase of spoofing attacks biometric systems are now concerned with three main classes: valid users, invalid users and spoofing/imposter attacks. This can be considered as the one enhanced and extended negative class, because we are interested in rejecting both invalid users and spoofing attempts. However, if the spoofing attacks are of good quality, they may overlap in the distribution of valid users. As a consequence, the positive and the enhanced negative class are not that well separated anymore. To cure this problem, systems can be designed in different approaches.

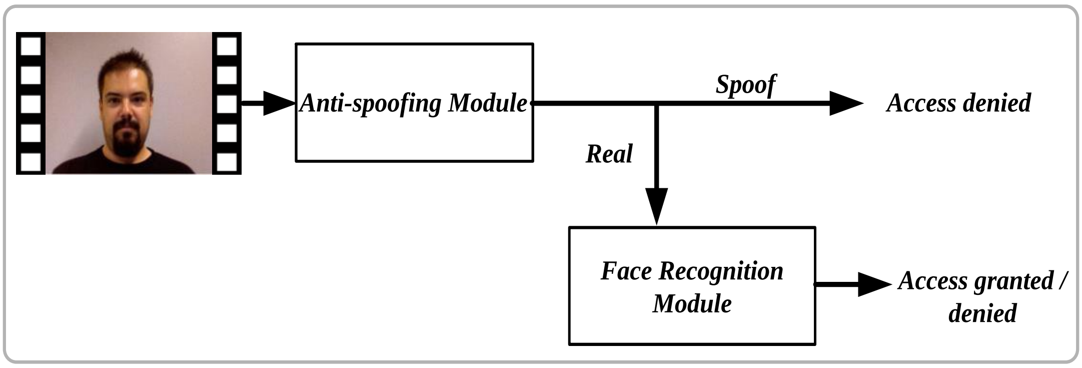


Figure 4-3 Anti-spoofing algorithm as pre-requisite for verification systems

The basic system design can include the anti-spoofing system as a prerequisite step for verification system. This system can rapidly reject imposter attempts, and based on positive score from the anti-spoofing system, the face recognition can start processing for verification. The second approach can be designed in a way that, both these units can works in parallel and the scores of both modules can be fused to form final output of the system. The fusion of these modules can be done on decision-level fusion of recognition system and anti-spoofing system. The decision-level approach works on the judgment taken by the verification system with an additional check performed by an anti-spoofing system because recognition system and anti-spoofing systems are of different nature. In particular, a recognition system has to reject a zero-effort impostor because positive class for one system can be negative for the other and vice-versa as illustrated in Table 8. In decision- level fusion rejection from one system is enough and acceptance is considered if access is accepted by both systems. For fusing the decisions of these two systems AND fusion rule can be applied. Both versions of the proposed frameworks can be seen in Figure 4-3 and Figure 4-4 respectively.

Table 8: Criteria for positive and negative class of a typical verification and anti-spoofing system and the final system of interest

Input	Anti-Spoofing System	Face Recognition System	Final System
Valid User	true	true	Access Granted
Invalid User	true	false	Access Denied
Fake User	false	true	Access Denied

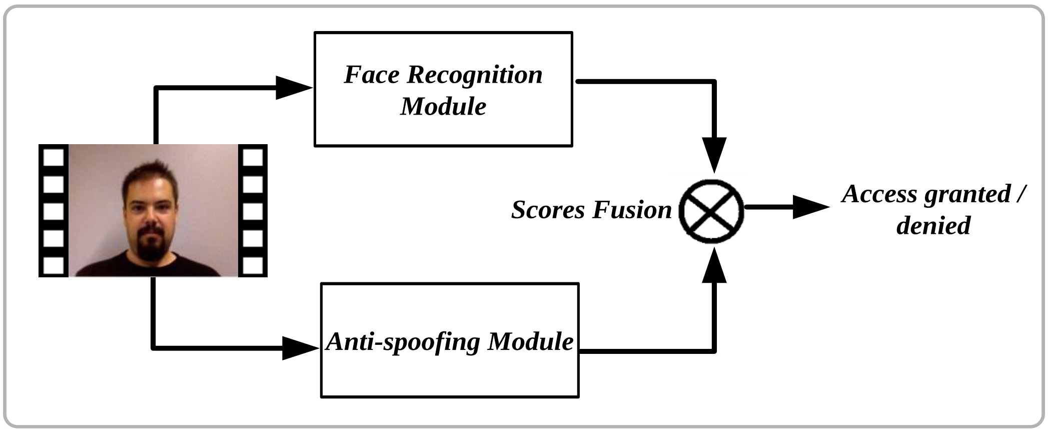


Figure 4-4 Fusion of verification and anti-spoofing algorithm

4.5. Analysis of Computational Complexity

Face verification systems are usually intended to be real time systems. Some of anti-spoofing counterfeits proposed in this thesis are computationally expensive. Therefore, it is really important to have a ceiling analysis of these countermeasures running in parallel or as preprocessing step for face recognition systems to know their computational power. This can help in analyzing their applicability in real time systems. For this purpose, average time taken by these anti-spoofing solutions are analyzed in this section.

A comparative time consumption study of feature extraction algorithms is summarized in the table. Features were extracted for a video of 11 seconds duration using Intel(R) Core i7 CPU@ 3.40 GHz, 32 GB RAM computer.

Gabor features have shown to have beneficial properties in texture classification and feature extraction for many computer vision tasks, but their computational complexity has prevented their use in practice. Experimental results show that Gabor wavelet feature extraction process is most expensive in terms of time complexity as shown in the Figure 4-5. One way to reduce this time complexity would be to have the features extracted only for certain frames over an interval of 5-10 frames in a video. This would certainly speed up the process of feature extraction. However there would be some degradation of accuracy in classification. One could experiment with these and arrive at an optimal tradeoff between speed and accuracy.

Table 9: Table illustrating the time consumed by different modules of the proposed countermeasures

Stages	Time (sec) / video (11 sec duration)
Eulerian Motion Magnification	9.06526
$\mathrm{LBP}_{P,R}^{riu2}$	13.4928
Gabor	53.2396
GLCM	4.94334
STACOG	2.18830
Image Quality assessment	0.91722

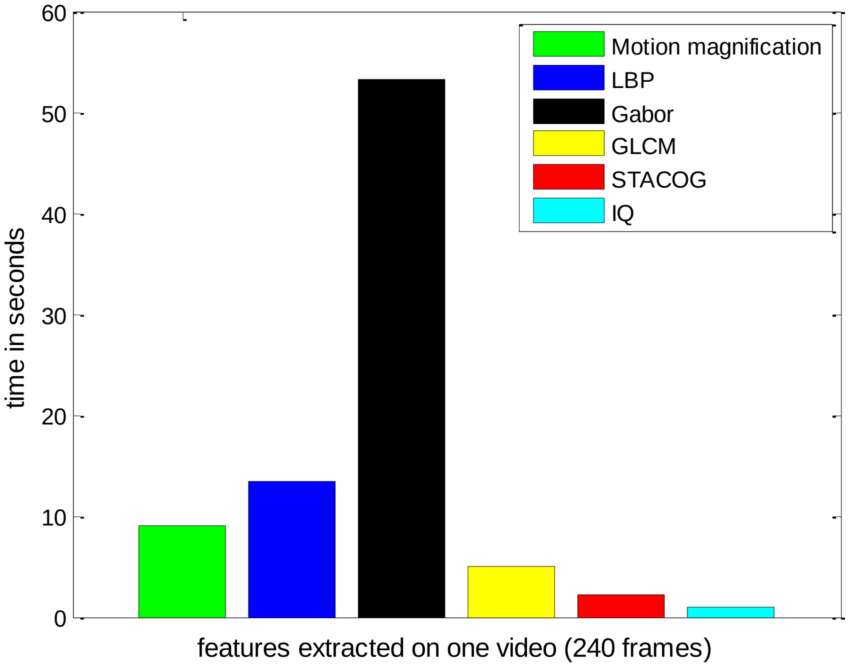


Figure 4-5: Bar chart illustrating the time consumed by different modules of the proposed countermeasures

$\mathrm{LBP}_{P,R}^{riu2}$ and eulerian motion magnification can easily be combined to form a real-time system. Another possible approach to build a speedy system would be to rely on only motion based countermeasure. As both STACOG and Eulerian motion magnification have clearly a very low time complexity and produce good performance in classification, the combination could promise a fast and reliable real-time authentication system.

However one must not forget that texture based methods produce comparatively high performance as shown in Section 4.3. On systems with high computational power texture based features is still the preferred way to proceed.

5. CONCLUSION AND FUTURE WORKS

Finally, thesis is concluded with brief concluding remarks in this chapter. A short discussion on potential possibilities for further improvements of the proposed counter measures and some new interesting ideas are also specified. Some good reference papers are also mentioned for the readers who are interested in continuation of this work.

5.1. Conclusion

In this thesis spoofing challenges to face biometric system were tackled considering different kind of imposter scenarios namely, printed attacks, digital photo attacks and dynamic video attacks. The research carried out in this thesis introduced techniques for forfeiting these imposter attempts. Three different countermeasures were proposed which utilizes several textural variations, motion differences and noise variations to classify the imposter attempts. Two standalone frameworks were proposed that can easily be integrated into the existing face biometric systems without any extra hardware cost. All schemes proposed in the thesis achieved higher performance compared to the conventional reference under the same category e.g. texture based or motion based or image quality based.

Experimental results on two diverse and challenging datasets illustrate the effectiveness of these approaches. First, the role of image regions was analysed by selecting two different regions; only face region and full image scene. The region of interest selection as a whole image scene demonstrates that full image scene provides higher discriminative power as compared to the face region. The classification results were promisingly increased from 85.21 percent to 97.50 percent. The final approach that utilizes Eulerian motion magnification as preprocessing step exhibits a significant decrease in HTER to zero percent for both REPLY- ATTACK and self collected dataset.

Finally, utilizing motion magnification as pre-processing step before any of the proposed countermeasure reported in this thesis confirmed that the results are dramatically boosted. The combination of motion magnification and texture based countermeasure reported in this thesis achieved higher classification accuracy at the cost of time complexity.

5.2. Possible Future Directions

Imposter attempts on biometrics are major security threat and new intimidations arrive in form of better, more refined and sophisticated spoofing attacks. This thesis is just an initial effort toward this great field of making efficient biometric systems robust to intruder attacks. Proposed methods and algorithms in this thesis can be extended considering different tuning parameters than those already used. However, there are many potential ideas that were not implemented due to time and resource limitations. The rest of this section briefly discusses these possibilities and ideas

In this direction, one obvious possibility of improvement is to strive for better features which are not computationally expensive. Use of better algorithms for blind image/video quality assessment could be a possible solution for effectively restricting imposter attempts. Motion compensation algorithms can also be effective in this domain. Moreover, Feature synthesis for increasing discriminative power could be a potentially useful for classification. Other possibilities of improvement are to utilize better machine learning techniques for higher classification accuracy. For this purpose, one can start working with collective network of binary classifier [57].

There are many possibilities of improvement in this field. Training of classifiers with more realistic training samples will also help in better performance. Although the training samples used in this thesis from real world datasets and contains a wide variety, but most of the videos are taken in only limited backgrounds and lighting conditions. Also there are not much datasets available in this domain to extensively evaluate the performance of such systems. As current approaches might be only effective on the limited databases therefore, more databases is required which uses different capturing devices and display technologies. Collecting a new dataset of attacking events at different lighting condition, clients and view angles with their annotation will be a huge contribution to this field.

REFERENCES

[1] A. Anjos and S. Marcel, “Counter-Measures to Photo Attacks in Face Recognition: a public database and a baseline”, International Joint Conference on Biometrics, 2011.

[2] A. Beach, “Real World Video Compression” Peachpit Press, 2008.

[3] A. Jain and B. Klare,“Matching Forensic Sketches and Mug Shots toApprehend Criminals”, Computer, vol. 44, no. 5, pp. 94–96, 2011.

[4] A. K. Jain and A. Ross, “Handbook of Biometrics”, Springer, ch.Introduction to Biometrics, pp. 1–22, 2008.

[5] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-Reference Image Quality Assessment in the Spatial Domain”, IEEE transactions on image proc. vol. 21, no. 12, 2012.

[6] A. Martinez, R. Benavente, “The AR Face Database”, Technical Report, Computer Vision Center, Autonomus University of Barcelona, 1998. http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html (Last accessed: 18-08-2013).

[7] A. Pinto, H. Pedrini, W. R. Schwartz, A. Rocha, “Video-Based Face Spoofing Detection through Visual Rhythm Analysis”, SIBGRAPI, 2012.

[8] A. Rocha, W. Scheirer, T. Boult, and S. K. Goldenstein, “Vision of the Unseen: Current Trends and Challenges in Digital Image and VideoForensics,” ACM Computing Surveys, vol. 26, no. 1, pp. 26–42, 2011.

[9] B. E. Boser, I. M. Guyon, V. N. Vapnik, “A Training Algorithm for Optimal Margin Classifiers”, In Proc. of 5th Annual ACM Workshop on Computational Learning Theory, pp. 144-152, 1992.

[10] B. S. Manjunath and W. Y. Ma, “Texture features for browsing and retrieval of image data”, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI - Special issue on Digital Libraries), vol. 18, no. 8, pp. 837-42, 1996.

[11] C.-C. Chang and C.-J. Lin. LIBSVM: “A library for support vector machines”, ACM Transactions on Intelligent Systems and Technology, 2, 2011.

[12] C. J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition”, Data Mining and Knowledge Discover, 1998.

[13] Chakka, M.M., Anjos, A., Marcel, S., et al.: ‘Competition on counter measures to 2-d facial spoofing attacks’. Proc. IAPR IEEE Int. Joint Conf. on Biometrics (IJCB), Washington, DC, USA, 2011

[14] C.W. Hsu and C.J. Lin, “A Comparison of Methods for MultiClass Support Vector Machines,” IEEE Trans. Neural Networks, pp. 415-425, 2002.

[15] D. A. Clausi, “An analysis of co-occurrence texture statistics as a function of grey level quantization”, Can. J. Remote Sensing, vol. 28, no.1, pp. 45-62, 2002.

[16] De Jong, S. “SIMPLS: An Alternative Approach to Partial Least Squares Regression”, Chemometrics and Intelligent Laboratory Systems. Vol. 18, pp. 251–263, 1993.

[17] E. Allwein, R.E. Schapire, and Y. Singer, “Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers,” Journal ofMachine Learning Research, pp. 113–141, 2000.

[18] F. Dufaux , J. Konrad, “Efficient, robust and fast global motion estimation for video coding,” IEEE Trans. Image Processing, vol. 9, pp.497–500, 2000.

[19] F. L. Podio. Biometrics technologies for highly secure personal authentication. ITL Bulletin, Information Technology Laboratory, NIST, May 200l.

[20] G. Bradski, “The OpenCV Library”, Dr. Dobb’s Journal of Software Tools, 2000.

[21] Geladi, P. and Kowalski, B. R., "Partial least squares regression: a tutorial", Analytica Chimica Acta, Vol.185, pp. 1-17, 1986

[22] G. Pan, Z. Wu, and L. Sun, “Recent Advances in Face Recognition”, ch. Liveness Detection for Face Recognition, InTech, pp. 235–252, 2008.

[23] G. Pan, L. Sun, Z. Wu, and Y. Wang, “Monocular camera-based face liveness detection by combining eye- blink and scene context”, Journal of Telecommunication Systems, 2009.

[24] G. Pan, L. Sun, Z. Wu, S. Lao, “Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcamera”, 11th IEEE ICCV, 2007.

[25] H. Jin, Q. Liu, H. Lu, and X. Tong, “Face detection using improved LBP under Bayesian framework”, In Proc. Third International Conference on Image and Graphics (ICIG), pages 306–309, Hong Kong, China, 2004

[26] H Rowley, S. Baluja, T. Kanade, “Rotation Invariant Neural Network-Based Face Detection”, In Proc of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1998, pp. 38–44

[27] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. T. Freeman. Eulerian video magnification for revealing subtle changes in the world. ACM Transactions on Graphics, 31(4), 2012.

[28] I. Chingovska, A. Anjos and S. Marcel, “On the Effectiveness of Local Binary Patterns in Face Anti-spoofing”, International Conference of the Biometrics Special Interest Group, 2012.

[29] I. Chingovska, J. Yang, Z. Lei, D. Yi, S. Z. Li, O. Kahm, C. Glaser, N. Damer, A. Kuijper, A. Nouak, J. Komulainen, T. Pereira, A. Anjos, S. Gupta, S. Khandelwal, S. Bansal, A. Rai, T. Krishna, D. Goyal, M.-A. Waris, H. Zhang, I. Ahmad, S.Kiranyaz, M. Gabbouj, R. Tronci, M. Pili, N. Sirena, F. Roli, J. Galbally, J. ierrez, A. Pinto, H. Pedrini, W. S. Schwartz, A. Rocha, S. Marcel, “The 2nd Competition on Counter Measures to 2D Face Spoof-

ing Attacks”, The 6th IAPR International Conference on Biometrics (ICB-2013), 4-7 June 2013, Madrid, Spain.

[30] J.G. Daugman, “Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters,” Journal of the Optical Society of America A, vol. 2, pp. 1160-1169, 1985

[31] J. G. Daugman, “Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression,” IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 36,no. 7, pp. 1169–1179, July 1988

[32] J. Kubinek, “Extending training dataset for face detector learning”, Central European Seminar on Computer Graphics, 2009.

[33] J. Lukas, J. Fridrich, and M. Goljan, “Digital Camera Identification from Sensor Pattern Noise”, IEEE Trans. on Information Forensics and Security, vol. 1, no. 2, pp. 205–214, 2006.

[34] K. Kollreider, H. Fronthaler, and J. Bigun, “Non-intrusive liveness detection by face images”, Image and Vision Computing, vol. 27, no. 3, pp. 233–244, 2009.

[35] K. Nixon, V. Aimale, and R.Rowe, “Spoof detection schemes”, Handbook on Biometrics, pages 403-423, Springer US, 2008.

[36] L. Soh and C. Tsatsoulis, “Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices”, IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 2, 1999.

[37] L. Zhang, R. Chu, S. Xiang, S. Liao, S. Li, “Face Detection Based on Multi-Block LBP Representation”, In Proc. International Conference on Biometrics (ICB), 2007, pp. 11-18.

[38] M. Etoh , T. Ankei, “Parametrized block correlation—2D parametric motion estimation for global motion compensation and video mosaicing” , IEICE TR PRMU97, July 1997.

[39] M. Haller, A. Krutz, T. Sikora, “Robust global motion estimation using motion vectors of variable size blocks and automatic motion model selection” , (ICIP) 2010, pp. 737-740.

[40] M. J. Lyons, J. Budynek, A. Plante, and S. Akamatsu, “Classifying facial attributes using 2-d gabor wavelet representation and discriminant analysis”, in Proc. Fourth IEEE Int. Conf.on Automatic Face and Gesture Recognition. pp. 202–207 ,2000

[41] M Jones, P. Viola, “Fast multi-view face detection”, In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2003

[42] M. N. Haque, M. Biswas and M. R. Pickering,“ A Computationally Efficient Global Motion Estimation Using a Multi-Pass Image Interpolation Algorithm,” in Proc. Picture Coding Symposium (PCS), Krakow, Poland, 7-9 May 2012.

[43] N. Gourier, D. Hall, and J. L. Crowley, “Estimating Face Orientation from Robust Detection of Salient Facial Features”, In Proceedings of Pointing 2004, ICPR, International Workshop on Visual Observation of Deictic Gestures, Cambridge, UK, 2004.

[44] N. Kose, J. Dugelay “Classification of captured and recaptured images to detect photograph spoofing”, International Conference on Informatics, Electronics & Vision (ICIEV), 2012.

[45] N. M. Duc and B. Q. Minh, “Your face is not your password”, Black Hat Conference, 2009.

[46] N. Zamani, M. Darus, S. Abdullah, and M. Nordin, “Multiple-frames Superresolution for Closed Circuit Television Forensics”, International Conference on Pattern Analysis and Intelligent Robotics, vol. 1, pp.36–40, 2011.

[47] N. Kose, J. Dugelay “Classification of captured and recaptured images to detect photograph spoofing”, International Conference on Informatics, Electronics & Vision (ICIEV), 2012.

[48] O. Jesorsky, K. J. Kirchberg, R. W. Frischholz, “Robust Face Detection using the Hausdorff Distance”, In 3rd International Conference on Audio and Video-Based Biometric Person Authentication, 2001.

[49] P. Phillips, H. Moon, S. Rizvi, P. Rauss, “The FERET Evaluation Methodology for Face Recognition Algorithms”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 22(10), 2000.

[50] P. Viola, M. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2001, pp. 511–518

[51] R. Lienhart, A. Kuranov, V. Pisarevsky, “Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection”, In Proceedings of the 25th DAGM-Symposium, Magdeburg, Germany, 2003, pp. 297–304.

[52] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features of Image Classification”, IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-3, no. 6, 1973.

[53] R. Rosipal, L. Trejo, and B. Matthews. “Kernel PLS-SVC for Linear and Non linear classification”, In Proc. of the Twentieth International Conference on Machine Learning, pages 640-647, Washington, 2003.

[54] S. Baker, I. Matthews, “Lucas-Kanade 20 years on: A unifying framework”, International Journal of Computer Vision, vol. 56, pp. 221-225, Feb. 2004.

[55] S. Baker , I. Matthews, “Equivalence and efficiency of image alignment algorithms”, in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), December, 2001, pp. 1090-1097.

[56] S. F. Wu and J. Kittler, “A differential method for simultaneously estimation of rotation, change of scale and translation,” Signal Process.:Image Commun., vol. 2, no. 1, pp. 69–80, 1990

[57] S. Kiranyaz, T. Mäkinen and M. Gabbouj, "Dynamic and Scalable Audio Classification by Collective Network of Binary Classifiers Framework: An Evolutionary Approach," Neural Networks, 34 (2012), doi:10.1016/j.neunet.2012.07.003, pp. 80-95.

[58] S. Liao, X. Zhu, Z. Lei, L. Zhang, S. Z. Li, “Learning Multi-scale Block Local Binary Patterns for Face Recognition”, In Proc. of International Conference on Biometrics (ICB), 2007, pp. 828-837

[59] S. Trewin, C. Swart, L. Koved, J.Matino, K.Singh, S. Ben-David, “Biometric Authentication on a Mobile Device: A Study of User Effort, Error and Task Disruption”, Proceedings of the 28th Annual Computer Security Applications Conference (ACSAC), pp. 159-168, 2012.

[60] T. Kobayashi, N. Otsu, Motion recognition using local auto-correlation of spacetime gradients, Pattern Recognition Letters 33 (9) (2012).

[61] T. Ojala, M. Pietikäinen, T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, 2002.

[62] T. Pereira, A. Anjos, J. Martino and S. Marcel, “LBP − T OP based countermeasure against face spoofing attacks”, Asian Conference on Computer Vision, 2012.

[63] T. Sim, S. Baker, M. Bsat, “The CMU Pose, Illumination, and Expression Database”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 25(12), 2003

[64] T.S. Lee, “Image representation using 2D Gabor wavelets,” IEEE Trans. on Pattern Analysis and Machine Intelligence ,vol. 18, no. 10, pp. 959–971, 1996

[65] W. R. Schwartz, A. Rocha, H. Pedrini, “Face Spoofing Detection through Partial Least Squares and Low-Level Descriptors”, International Joint Conference on Biometrics (IJCB), 2011.

[66] Wold, S.Sjostrom, M. and Eriksson, L,"PLS-regression: a basic tool of chemometrics", 2001.

[67] X. Tan, Y. Li, J. Liu, and L. Jiang, “Face Liveness Detection from a Single Image with Sparse Low Rank Bilinear Discriminative Model”, European Conference on Computer Vision, pp. 504–517, 2010.

[68] X. Huang, S.Z. Li, and Y.Wang, “Shape Localization Based on Statistical Method using Extended Local Binary Pattern”, In Proc. Third International Conference on Image and Graphics (ICIG), pages 184–187, Hong Kong, China, 2004

[69] Y. Freund, R.E. Schapire, “Experiments with a New Boosting Algorithm”, In Proc. of the IEEE International Conference on Machine Learning (ICML), pp. 148–156, Bari, Italy, 1996.

[70] Y. -H. Lei, Y. -Y. Chen, B. –C. Chen, L. Lida, W. H. Hsu, “Where Is Who: Large scale Photo Retrieval by Facial Attributes and Canvas Layout”, In

Proc. of the 35th International ACM SIGR Conference on Research and Development in Information Retrieval, Portland, Oregon, USA, 2012.

[71] Y. Li, S. Gong, J. Sherrah, H. Liddell, “Support Vector Machine Based Multi-view Face Detection and Recognition”, Image and Vision Computing, Vol. 22, pp. 413–427, 2004

[72] Y. Rodriguez, “Face Detection and Verification using Local Binary Patterns”, Ph.D. thesis, EPFL, 2006.