Addis Ababa University

Addis Ababa Institute of Technology
School of Electrical and Computer Engineering

Liveness Detection based Anti-spoofing method in Face Recognition

By: Eyob Haile Nerea

A thesis submitted to the School of Electrical and Computer Engineering in partial fulfillment of the requirements for the Degree of Master of Science in Computer Engineering

December, 2020

Addis Ababa, Ethiopia

Liveness Detection based Anti-spoofing method in Face Recognition

By: Eyob Haile Nerea

_____	_____	_____
Thesis Advisor	Signature	Date
_____	_____	_____
Chairman of Department	Signature	Date
_____	_____	_____
Internal Examiner	Signature	Date
_____	_____	_____
External Examiner	Signature	Date

Submitted in Partial Fulfillment of the Requirements

For Masters of Science in Computer Engineering

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

December, 2020

Declaration

This thesis presents my original research work. Wherever other people's contributions are involved, every effort is made to make this clear, with appropriate citation from the source. I hereby declare that this thesis was not submitted for a degree at another university.

_____

Eyob Haile Nerea

Acknowledgement

First of all, thanks to Almighty God, for his innumerable blessings through my life and this search.

I want to express my profound and sincere gratitude to my research advisor, Mr. Menore Tekeba for his invaluable support, guidance, constructive insights and feedback that allowed me to gain valuable research experience.

I am deeply grateful to Kal Joba and Monamour for their love, prayers, concern, encouragement and preparation for my future. They served as my source of strength beside God

Abstract

Today, face liveness detection and recognition are an active research area due to the fact that it's unique biometric verification, scientific challenges and numerous practical applications. However, the effort of anti-spoofing to biometric identify are limited in accuracy as a result of the fact that its vulnerability to spoofing attack and inappropriate identification techniques. There are several sophisticated attacks in the video stream. Consequently, it becomes difficult to identify the real face from fake face. This research aims to address this challenge and develop an effective framework that can be used to deal with the face liveness detection and recognition.

This research consists of two phases. The first phase is face liveness detection. The techniques which are involved in this phase are collecting data set, extracting key eye frames, preprocessing, feature extraction, feature selection and classification. Phase two is face recognition. It is cascading with the result of the face liveness detection. The techniques included in this phase are extracting key face frames, preprocessing, feature selection and classification. Then the face frame is classified according to their respective identification classes with the help of classifiers. Two types of classifiers with three types of descriptors are used to assist us in providing rational and fair comparisons between the state-of-the-art.

The proposed approach is evaluated using institute of the automation Chinese academy of sciences (CASIA) anti-spoofing data set. Classification performance reports, 97.8% precision, 98.0% recall and 98.2% f-score were obtained using DBN classifier for liveness detection phase. Compared to the state-of-the-art approaches (DCP, LTP, and SFE method with SVM) an average improvement in f-score of 2.0% was achieved. Face recognition phase also assessed, where, 94.3% precision, 95.1 % recall and f-score 94.7% was achieved using DBN classifier. The overall system provides accuracy of 96.4%. The result showed that the proposed approach has enhanced the security performance of face liveness detection and recognition for a biometric security technology.

Keywords: Liveness detection, Recognition, Deep belief network, Local ternary pattern, Dual cross patterns.

Contents

List of figures

List of tables

List of algorithms

Abbreviation

CASIA: Institute of automation Chinese academy of sciences

DBN: Deep belief network

DCP: Dual Cross Patterns

DOG: Difference of Gaussian

FD: Face Detection

FR: Face Recognition

GKFD: Generalized Kernel Fisher Discriminant Analysis

HOG: Histogram of Gradient

KDA: Kernel Discriminant Analysis

KFD: Kernel Fisher Discriminant

KPCA: kernel principal component

LBP: Local Binary Patterns

LDA: Fisher linear discriminant

LDFR: face Liveness detection and recognition

LPQ: Local Phase Quantization

LQP: Local Quantization phase

LTP: Local Ternary Patterns

PCA: Principal Component Analysis

PLDA: Probabilistic Linear Discriminate Analysis

PLDA: Probability linear discriminant analysis

RQ: Research Question

SFE: Shearlet feature extraction

SVM: Support Vector Machines

Chapter 1

1 Introduction

1.1 Background

With the LDFR automated system, the identification of authorized persons in the event of an attack becomes extremely important in recent technology. A unique personal identification system, such as username and password, is vital identification technique. However, it may be easily attacked and guessed through various advanced tools of attackers. As a result, more sophisticated application is recommended for users like biometric authentication systems [1].

Biometric identification is a multidisciplinary field that aimed at measuring and mapping specific biological behavior. It's a robust technique for different types of attacks than other verification technologies and, overcome the limits in the accuracy of LDFR systems. Nowadays, biometric identification systems are becoming very active fields in face recognition, fingerprinting, iris recognition due to the fact that it has challenging tasks and broad area of practical application [2].

The physiological signs of life in biometric identifiers include eye blinking, lip and head movements, these facial expressions used as clues to prevent several identity attacks [3]. Due to the fact that there is an advanced in technology tools, several attacks have attempted to access unauthorized personal data and thus, become a more difficult challenge threat to biometric authentication [4]. As a result, robust biometric identification and verification are a highly reliable method of authentication techniques through. First, it's easily equipped with a generic camera from this, it has a low cost and high usability. Second, is easy integrate it with existing recognition systems [5].

Face photograph is the most common way to simulate the face liveness detection and recognition system, as the individual's face image is easily accessible to the public. For example, using advanced tools downloaded the photo images from social media without any privilege. The photo attack is one of the simplest and easiest way to spoofing. The attack can rotate, shift, and fold the valid user's photo in front of the camera as a live person to break the authentication system. It's a tough task to detect either the input image frame is from a live person or an attack [4]. Thus, the face liveness detection and recognition are an obstacle, especially for multi-variation of spoofing attacks [4]. The few numbers of ant-spoofing data sets that are more applicable to the LDFR

training and testing data set has high impact on the output result of system accuracy. The complex structure of LDFR identifies system required best design in addition, robust feature descriptors. Since biometric identification is a rich morphological description of form, which shares all of the challenges mentioned above.

The steps toward the development of face liveness detection and recognition can grouped into three major categories. Preprocessing approach, rule-based (handcraft approach) and machine learning (statistical approach). Preprocessing approaches are applied to remove the noise signals from image frame. Rules-based approaches are based on handcrafted that are used to store discriminate and relevant information in the extracted image frame. Machine learning approaches use a collection of training image frames to construct a model based on the training data set. Machine learning (ML) approaches include supervised, unsupervised, and semi-supervised. The supervised approach uses labeled training data set to build LDFR model. Unsupervised approaches do not need structural information to the data set. Semi-supervised approaches use both labeled and unlabeled data set for training purpose.

Face liveness detection and recognition are done by a few researchers. The crucial role of this research is developing a robust LDFR identification system better than from the previous identify approaches.

1.2 Statement of the Problem

Due to the fact that critical role in several biometric applications, liveness detection based anti-spoofing method in face recognition is one of leading research areas in biometric authentication. It has an active field of research over the past few years [5].

Based on [6], it is very difficult to detect the characteristic properties of the face under compute vision uncontrolled environment situations. There are several ways that we can access the LDFR system without privilege. For example, a print attack is a type of attacker that use authorized user's image frame by showing a user imposter image to its advanced tools. As a result of the fact that advanced growth of the spoofing attack technology, the authentication system is highly vulnerable. It has become high computing for developing safe anti-spoofing approaches. The LDFR system has not permitted an unauthorized user to access other personal data, but it's also prohibits access the authorized user. Many users have suggested that LDFR system is too vulnerable to be attacked.

Several researchers have been done on LDFR system. Unfortunately, they are separated liveness detection as one system and recognition as another system. The development of the isolating system has achieved a good result. However, the separating systems are insecure, inefficient in terms of performance and resources.

Such issues could be addressed by automating the surveillance process using an image preprocessing and machine learning techniques. Different researchers have been experimenting with the problem of identification and classification for a long period of time. Most of the machine learning techniques used for the classification purposes where manual training in addition, has limited in accuracy. Another limitation was the set of features were used for classification that was generated with a specific pattern setup. These models do not generalize very well due the fact that the generated features are not high-level and weak features that describe the specific interested part of the face frame.

Consequently, accurate identification and classification of face image using deep neural networks is necessary to control the different type spoofing attack. High prediction accuracy of detection and classification is playing a vital role in biometric security application. The latest improvements in deep learning have increased its capability to solve such complex identification tasks. Therefore, using deep learning model we can improve the accuracy and, overcome the above challenges of the LDFR system. Finally, the following set of research questions will be answered.

- **RQ1 [Preprocessing for LDFR]**

Can the preprocessing method with enhanced techniques improve performance of face liveness detection and recognition system ?

- **RQ2 [Feature Extraction for LDFR]**

Can DCP outperforms the state-of-the-art feature extraction (SFE and LTP) for face liveness detection task ?

- **RQ3 [Deep Neural Network for LDFR]**

Can we improve the performance of face liveness detection and recognition using deep neural network ?

1.3 Objective

1.3.1 General Objective

The general objective of this thesis is to design and implement liveness detection based anti-spoofing methods in face recognition using CASIA ant-spoofing data set.

1.3.2 Specific Objective

It has the following specific objectives.

- Explore the LDFR system concepts and use the most up-to-data algorithms.
- Provide fair and rational comparison between various types of descriptors and classifiers.
- Study biometrics technology using face detection for ant-spoofing and recognition for authentication tasks.
- Offer solutions to several types of spoofing attacks.
- Construct deep neural network architecture as well as classifiers that use vector frame descriptors as input.
- To design and simulate the proposed architecture furthermore, evaluate the performance of the new approaches.

1.4 Significance of the Study

- It searches a new framework for computer vision researchers in the selecting robust descriptors, classifiers in addition, produces secure biometric applications.
- The research plays necessary role in understanding the limits of accuracy, challenges and, develop appropriately methods for face liveness detection and recognition systems.
- It will open a new direction towards the development of an unconstrained and degradation image processing toward the face liveness detection and recognition technology.

1.5 Methodology

- **Literature Review:** A literature review of the theoretical bases and different existing research works toward face liveness detection and recognition systems.
- **Select Descriptors:** By examining the existing system, robust feature extraction descriptor will be selected, that furthermore used to LDFR system. In addition, develop automatic features with the help of deep neural network.

- **LDFR Model Development:** By using auto-generated features as well as a descriptor with two classifiers build a new proposed architecture of LDFR model.
- **Data Collection and Preparation:** The eye frame image that necessary for experiments is collected and prepared for CASIA anti-spoofing data sets.
- **Evaluation of the LDFR model:** Model developed through the proposed architecture are evaluated using precision, recall, and f-score metrics and finally a discussion of the results, conclusions, and future work directions will do.

1.6 Scope and Limitation

The scope of the research is only limited to detect and classify of the face image using CASIA anti-spoofing data sets. It was difficult to get local public data sets due to the fact that the different quality of the camera not available. Up on agreement with the Ethiopian security research institute, the model developed can be tested with their data.

1.7 Contribution

Several methods combine multiple approaches such as motion and texture analysis of the framework image to overcome all types of attacks. However, it is not always clear how each component contributes. The objective of this research is to cascading the system and study several descriptors beside classifiers independently, further that help us to provide fair in addition, rational comparison between the state-of-the-art. The following contributions were made in the study.

- Review of advanced methods to establish the current limits of technical approaches.
- Development of a robust LDFR evaluation framework based on the study of the recent LTP, SFE, and DCP methods for the evolution of advance texture-based countermeasures.
- Investigated the preprocessing and different classifier approaches to further describe the different techniques that involve a real and fake face.
- Prepared data set from CASIA face ant-spoofing data set that used to assess the impact of face liveness detection in LDFR architecture.
- The face liveness detection and recognition proposed have surpassed the most modern method.

1.8 Thesis Organization

This thesis is organized as follows. A detailed understanding state-of-the-art, challenges of face liveness detection and recognition system are discussed in chapter two. The theoretical context of the LDFR is described in chapter three. Chapter four for an explanation of the proposed architecture. The experimental procedures followed and the result is presented in chapter five. Finally, Chapter six results of experimental observations and future work present to show other areas of improvement in LDFR system.

Chapter 2

2 Literature Review

2.1 Face Liveness Detection

These types of approaches are based on the face that identifies a fake face by distributing the statistical property of the real face(actual) frame using the eye blinking and the lip movement [7]. In [8] the researchers presented approaches to liveness detection against photograph attack in face recognition systems, recognizing spontaneous eye blinks. The main objective is to resist the attack identity theft a non-intrusive way without any external equipment with the exception of a generic camera. The physiological activity of eye blinking is to immediately close and open the eyelids. The camera may capture two or more pictures as the face looks at the camera. It does as a hint to use eye blinking against in anti-attack. The researchers postulate that the independent sequence is useful for liveness approach. The approaches outperform the cascading Adaboost and Markov pattern hidden in the blink detection task and 95.7% results achieved using a two-eye detection rate. The limitation of this research is that all frames are dependent may not be assumed free independents each frame.

Work in [9] the developers provide a component-based coding approach to face liveness detection. There are four stages of the proposed methodology. First, localize the facial components. Second, coding the low-level characteristics respectively, for all the components. Third, calculate the high-level face representation by clustering the codes with weights derived from the Fisher criterion. Fourth, beak down the histograms of all components in a classifier to identify them. Three types of descriptive method to be extracted like LBP, LPQ, and HOG and fed to the SVM classifier. The disadvantage of this research is that feature extraction that used was not robust and a one type of classifier was applied. Three types of data sets are used to assess the data sets of the approach, from four data sets, the high score performance found in the print-attach data set with 99.5%.

The researcher [10] applied the diffusion speed method to difference live and fake faces. Diffusion speed computing by utilizing the total variation flow scheme and extracting anti-attack features based on the local patterns of diffusion speed. The features are fed into SVM classifier to determine the liveness detection of the given face image. The researchers successfully perform when the images are captured in a wide range of indoor and outdoor environments. Experimental results on

various data sets show that the proposed method is robust as well as effective. Accurately identifies diverse malicious attacks with the performance of 98.5%. To evaluate the performance of the approaches they flatten various data sets. However, to classify they have not compare and contrast different kinds of classifiers.

2.2 Face Recognition

Face recognition approaches in [6] on anti-attack has been presented, but still now have several challenges. In [11] they present a new scheme for extracting the MDML-DCP from face frame image and calculate the extracting of the DCP descriptor at both the global and component scale. MDMLDCP are computationally more effective, but it merely doubles the cost of computing local binary models. However, it is extremely robust for posing, expression variations and without constraining in addition, uncontrolled face image. The SVM classifier is used to classify the descriptor method and preferred PLDA algorithm based on the data set. Experiments on four large-scale face data sets demonstrate the DCP has robust, expressive and moderate light variations with a 95.58% performance check rate. The descriptions provide successful. However, it has a high time execution.

For two combinations descriptors, the author [12] presented based on the local phase quantization LPQ and expanded it to a multi-scale framework (MLPQ) to make it more efficient. Regional characteristics are combined using kernel fusion. The combined Multiscale Local Binary Pattern (MLBP) descriptor was applied to core kernel fusion to decrease the light sensitivity of the combined feature extracted from discriminate information for face recognition. The experiments were conducted on six sets of sample data sets and the experiment showed greater with an average of 98 %.

In the research, [13] enhanced the combination of KPCA and LDA by proposed a new GKFD that is applied in face recognition. They take advantage of two categories discriminating information, not only utilize of discriminating information. And also unifies discriminant functions in two subspecies of dual-space linear discriminant analysis (DSDA) as well as improve the KFD method. Experiments have shown that GKFD works better than kernel Eigen face and kernel Fisher face when it’s two types of discriminant features are combined. The new GKFD was 93.5% accurate, but the total performance was not satisfactory.

2.3 Face Liveness Detection and Recognition

Technology to face liveness detection and recognition have recently been developed, but the security of the application of this approach is very limited when used practice [2]. In research [14] a new descriptor called shearlet transform was defined. The same feature extraction is applied to face liveness detection and recognition systems. The SFE is extracted from a face frame image, but before being sent in the stacked autoencoders, the input SFE is normalized. And then inserted into stacked autoencoders that are concatenated with the SoftMax classifier. The extracted descriptors are used to detect the liveness of the face frame. If it is a real face, these descriptors can be used directly for recognition otherwise they discard it. The result of the anti-attack data sets shows that it suits for both tasks.

Other works specified in [15]. The author presents and analyses the differences in optic flow properties that generated by 3D objects and 2D planes. The motion of the optical flux field is a combination of four basic types of motion. The first three basic types provide fairly similar optical flow fields for 2D and 3D images. The real optical flow field difference creates the fourth type. The approaches allow to deduce the reference field in addition, determine if the test area is flat or not. To do so, the difference between the optical flow fields is calculated. To determine whether a face is a real or fake, this difference is noted as a threshold. The experiment was performed on three sets of sample data sets and the experiment showed good results.

In other studies [16] face liveness detection and recognition approach, divide in to two parts such as liveness detection and face recognition. They applied Haar pre-trained classifier to detect the interested face party, PCA (principal component analysis) to generate the Eigenface as well as the Euclidean distance method used for recognizing the person based on the distance between the test and the train images. Two types of experiment configuration. First, the generation of attacks (five kinds of attacks). Second, face liveness detection and recognition occur simultaneously and the same frame is used to test on face liveness and recognition. In general, the system displayed a good precision ratio and manage to identify all trained people.

2.4 Face Anti-Spoofing Data set

Face anti-spoofing data set has been more attracted and intense attention, their main goal is to ensure the reliability of biometrics technology system. Institute of the automation Chinese

academy of sciences(CASIA) ant-spoofing data set developed by [17] which has significantly improved and has multi variation data collection compared to previous anti-spoofing data sets. Consequently, the data sets contain 50 subjects since each subject contain 12 videos in addition, individual subject includes three types of image quality as well as three types of fake face. The design test protocol consists of 7 scenarios to provide analysis of the various factors that can affect the accuracy of the anti-spoofing data set.

Research	Approach	Descriptors	Accuracy
Face Liveness Detection from a Single Image via Diffusion Speed Model	SVM	LSP+DOG	89.6
Face Recognition with Liveness Detection using Eye and Mouth Movement	KPCA+LDA and CNN	GKFD	93.5
Multiscale Local Phase Quantization for Robust Component-Based Face Recognition Using Kernel Fusion of Multiple Descriptors	MLPQ and KDA	LPQ +MLPQ	92.4
Multi-Directional Multi-Level Dual Cross Patterns for Robust Face Recognition	PLDA+SVM	DCP	91.59
Face liveness detection and recognition using shearlet based feature descriptors	Stacked autoencoder, SoftMax and SVM	SFE	92.1

Table 2-1: Summary results of different researches to LDFR

This research uses a machine-learning, similar to all previous research. However, this research differs from the above researchers how to cascading LDFR systems with deep networks in addition, a robust descriptor applied. Other previous systems have succeeded on anti-spoofing method, however, they cannot distinguish a live face from fake accurately using a multivariate anti-spoofing data sets. While this research is dependent a variety of the data sets.

Chapter 3

3 Theoretical Background

3.1 Face Liveness Detection and Recognition

Face liveness detection and recognition is an advanced biometric identification technology that has developed rapidly in recent year. Only face recognition has not secure sufficiently in biometric authentication systems as it cannot distinguish a live(true) face from fake face [5]. For the time being there is an easy way of spoof face recognition through photo print as well as portrait photographs. Therefore, to become safer in the biometric identify system, it required liveness detection approach that further helps to protect itself from the anti-spoofing approach [5].

3.2 Face Liveness Detection

Recently, a physiological life of sensing in biometric authentication become active fields in face recognition, fingerprint identification, and iris recognition. But, technology toward face liveness detection are not sufficiently secure, limited in accuracy and, it has many challenges tasks when used in practical application areas [18].

3.3 Face Recognition

Face recognition has been an active area of research due to the fact that scientific challenges in addition, a broad range of potential application. Face recognition methods have achieved satisfactory performance nevertheless, they are only in controlled environments [4]. It becomes productive and more widely utilized in our daily lives activity. Furthermore, when we compare and contrast with fingerprint and iris recognition, feature of the face is very easy to copy through photo print or from the video stream. Biometric authentication system is a challenge task due to the fact that multiple advanced face impersonations tools [6].

3.4 Application of Face Liveness Detection and Recognition

LDFR is a major task in the processing of advanced biometric imaging system [6]. It has many applications based either identification or checking the presence of the individual face frame. It plays a crucial in several biometric security applications from entertainment to security, such as information security, smart card based, entertainment, surveillance and law enforcement [19]. It

also plays a vital role in anti-spoofing biometric applications like logical access control, attendance system, and automatic recognition.

3.4.1 Information Security

There are multiple applications required either password or username to access privacy data. These kinds of techniques system are not always desirable. But, using a robust biometric identification of LDFR are able to access the data that only belong to the authorized person. A system like a computer log and ATM only allows it to be made possible.

3.4.2 Surveillance and Law Enforcement

Video monitoring is now part of the business of today's society activity. Although opinions on closed circuit television (CCTV) are varied, safety could be improved several times using an automatic face liveness detection and recognition system.

3.4.3 Smart Card-Based

Driving licenses, passports, airport checks and identity cards, all these applications can submit using a photograph of someone's face (attack). Such limitation in accuracy can overcome by a robust LDFR technique system.

3.4.4 Entertainment

Potential entertainment application like video games can recognize the player and load their preferences from human computer interaction system. Such an application needs like LDFR technology to become the current system safer and run smoothly.

The above list is a sample of possible applications that could benefit from the face liveness detection and recognition method. There are many effective approaches and methods. However, it is not robust and has low accuracy with the minimal entry.

3.5 Face Liveness Detection and Recognition Approach

The Steps leading to the face liveness detection and recognition according as per [20] can classified into two broad categories.

3.5.1 Rule Based Approaches

These approaches consist of a set of template using grammatical, syntactic and orthographic characteristic. Most face image descriptors are hand crafted (handmade), including local binary patterns (LBP) and Gabor wavelets are both representative methods. This type of system can give better results for restricted field. And it can detect complex face frames that might be hard to learn from the pattern. However, the main drawback of rule-based approaches is that lack robustness, probability, and high maintenance cost in a slight date change [21].

3.5.2 Machine Learning Approaches

Machine learning approaches use a collection of training data sets to automatically induce rules. These rules are sequence labeling algorithms for categorizing face images [22] Machine learning approaches are defined as statistical models to make predictions about face image in a given image frame or video stream. The significant advantage of learning based descriptors over handcrafted descriptors is their greater diversity in sample forms and the larger sample size. In general, there are three major approaches to machine learning.

Supervised Learning Approaches

The supervised learning approach uses labelled training data sets to allow the classifier to grow a model that can correctly classify. The goal of this method is to obtain latent data rule and relationships to predict labels of data that do not observe after learning. The performance of this approach depends on the number of practice or training sample. As indicated in [22] these systems cannot achieve high efficiency with a small training dataset. Recent supervised learning approaches are the most dominant approaches in face liveness detection and recognition system technologies. Some popular example of supervised learning include hidden markov, maximum entropy models, conditional random fields, and support vector machine.

Semi Supervising Approaches

Semi-supervised learning methods employed labeled and unlabeled face frames for the training set. The primary motivation behind semi-supervised learning is to overcome the scarcity of tagged training data by using a large unmarked face frame [6]. The primary method of semi supervised learning is called bootstrapping, which involves small controls early in the learning process. The

model consists of an initial set of label data, followed by a prediction with a distinct set of unlabeled data set. Model performance is then improved using predictions from that model already developed [23].

Unsupervised Approaches

Unsupervised machine learning algorithm use unmarked information for the training data set. Typically, a small amount of data labeled in large unlabeled data typically falls between supervised and unsupervised learning. It intends to find previously unknown data models, but most of the time these models are poor proxies of what supervised machine learning can accomplish [24]. Certain applications of unsupervised techniques include clustering, fault detection, association mining, and latent variable models.

3.6 Key Frame Selection

A video stream is split into a series of image called image frame. A frame is flashed on a screen for a short time, then immediately replaced by the following one [25]. An extracted key image is used to reduce the amount of memory space, the total time is taken for the processing of video flow data and the degree of complexity in the overall processing of images [26]. The key image used for face liveness detection should be calculated based on the spontaneous blinking rate of the human being eye [27]. By applying the histogram difference of a single image is mapped to the one of its neighbors to check the similarity of the frame.

3.7 Preprocessing

The unwanted signal or noise can found in any image processing in addition, it is perceived the deterioration of the quality of the face image. Particularly in key images may be exposed to a lot of potentially degrade such as blur, pose, and low resolution. That also decreases the quality of the face image and becomes hard to detect [28]. The main role of preprocessing on the video stream or face frame is an improvement of the image data by removing the unwanted disgrace of quality. There are a variety of advanced preprocessing algorithms. This paper briefly discusses certain preprocessing methods that used in this research.

3.7.1 Face and Eye Detection

Face and eye detection is a process where the system detects the presence of the face or eye with the corresponding regions in the given image frame. In biometric security systems, face detection should ensure that all images of data sets need to have a similar setting. The challenges of eye detection are similar to those of face recognition, like noise, pose, blur, lighting, etc. Most often, eye and face detection algorithm consist of extraction modules and classification of characteristics. Many other advanced face detection algorithms exist, such as [28] [29].

3.7.2 Noise Removal Using Median Filter

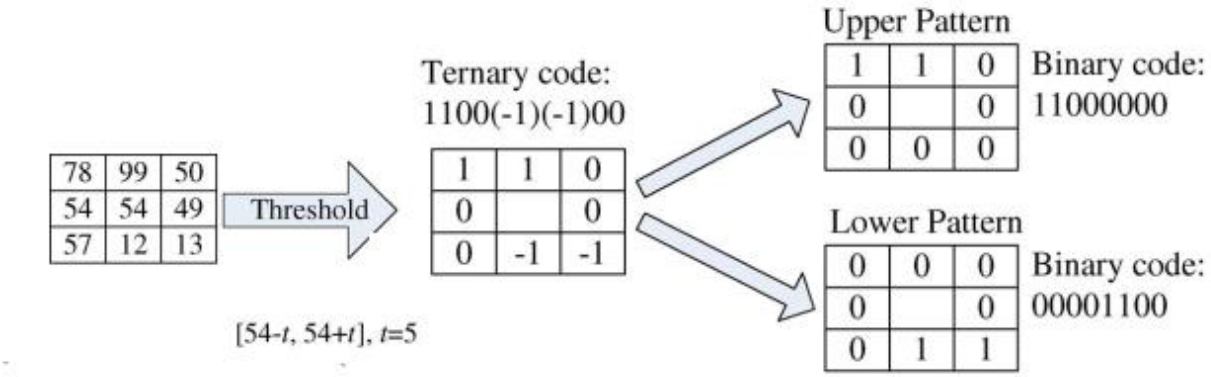
The noise suppression method is to delete the noise from image frame. The middle filter is an example of noise elimination that helps us to keep the shape in addition, the position of the edge image. It's most effective in terms of elimination, edge preservation, fine details of the digital image and applies to many works such as [30][31]. The algorithm determined the median value for each pattern of neighboring elements called the window or help, which is later used as a filtering result for the central element of the window [31]. In the presence of noise, especially image corruption, the median filter is well known for its computational efficiency and has the ability to maintain the edge compared to other linear filtering techniques [31]. Some of the latest preprocessing algorithms are [32][33].

3.8 Discriminative Feature Extraction

Feature extraction is the method of extracting meaningful and discriminatory descriptors from the face frame that will be used for identification systems. This is also a technique in which a high-dimensional space for face images can reduce to a space for low-dimensional face features. There are some issues when we conduct an analysis of complex data science, such as evaluating large variables with a large amount of memory, derived from the set of variables, and allowing the classification algorithm to overfit the training sample. It is very necessary to identify the face frame features that are sufficient to characterize a face frame and resolve the limitations above. There are many ways of state-of-the-art feature extraction. Some recent robust descriptors that were used in this research are briefly discussed.

3.8.1 Local Ternary Pattern

Extraction of the LTP function assigns a mark of each pixel face frame to the neighborhood with the center pixel value threshold 3x3 [34]. In this case study this [49, 59] is the pixel width, the range allocated to a pixel in window 0. When the threshold between c-t and c+ t where, c is the center intensity, and t is the user-specified threshold. All values above 54 are allocated to 1, any values below 54 are allocated to -1, and the s (u) parameter is replaced by a 3-value function. We have to break the code into upper and lower patterns when we determine the ternary tags. In general, values were allocated to -1 are reallocated to 0 for upper patterns and values were to -1 are reallocated to 1 in addition, any values that are 1 from the original window get mapped to 0 for



lower pattern.

$$. s'(u, i_c, t) = \begin{cases} 1, & u \geq i_c + t \\ 0 & |u - i_c| < t \\ -1 & u \leq i_c - t \end{cases} \qquad (3.1)$$

Figure 3-1: Splitting LTP code into positive and negative.

After this, the pattern is reading counter-clockwise starting from the east location with regard to the center (row2, column 3). Last, the output is the combination of both lower and upper clusters with a distance transform-based similarity measure in the local histogram. As shown in figure 3.1, LTP codes are more resistant to noise and have robust features.

3.8.2 Shearlet Transform

Multidimensional phenomena are mostly dominated by shearlet transform. It tries to overcome other traditional multiscale transforms due to the fact that there is no limitation on the scale shearing supports in addition number of shearing directions [36]. While shearlet transformation consists of

directional and anisotropic attributes, it has the ability to capture edge geometric information effectively.

This study used shearlet transformation through single mother functions. Which is parameterized by scaling and shear translation via capturing the direction of singularities [35].

By applying Laplacian pyramid and scheme directional filtering, shearlet transform is accomplished [36]. Shearlet is a sampling of the mother feature generated by shearing, dilating and translating $\psi \in L^2R^2$. DST (discrete shearlet transformation) is gained by sampling continuous shearlet transform, a discrete subset of the shearlet group S, which are associated to an orthonormal basis for $L^2(R^2)$. The mother function ψ for DST is defined as.

$$SH\ \{\psi_{j,k,l} = 2^{\frac{3}{2j}}\ \psi\ (B_k, A_{2j} - \mathrm{L}) : \mathrm{j}, \mathrm{k}\ \in \mathrm{Z}, \mathrm{L} \in Z^2\} \qquad (3.2)$$

Where, j, k, and j are the scale, orientation and location indexes and

$$A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}, B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Are corresponding to the dilation matrix and the shear matrix respectively. For the given image $f(N_{rows}xN_{colums})$. The DST can be written down as.

$$< f, \psi^d_{j,l,m} >= 2^{3/2j} \int_{\dot{R}^2} \dot{f(\xi)}\ \{\overline{V(ooo2^{2J(\xi)}w^d_{J,l}(\xi)}\ \ e^{2\pi i\xi}A_d^{-j}B_d^{-1}m\} \qquad (3.3)$$

Where,

$$V(\xi_1, \xi_2) = \psi_1\ (\xi_1)_x D_0(\xi_1, \xi_2) + \psi_1\ (\xi_2)_x D_1(\xi_1, \xi_2) \qquad (3.4)$$

Then x represented the indicator function of the set $D, D_0 and\ D_1$are the horizontal and vertical trapezoids respectively $d \in \{0,1\},\ \xi = (\xi_1, \xi_2) \in\ R^2, j\ \geq 0, l = \left(-2^j \ldots . 2^{j-1}\right)$ this is horizontal trapezoids of junction.

Where, $W^d_{j,t}\ (\xi)$ is a window function that located on a pair of trapezoids and

V is the coordinates of pseudo-polar.

$$D_0 = \left\{(\xi_1, \xi_2)\epsilon R^2 : |\xi_1| \geq \frac{1}{8}, \left|\frac{\xi_2}{\xi_1}\right| \leq 1\right\} \qquad (3.5)$$

$$D_1 = \left\{(\xi_1, \xi_2)\epsilon R^2 : |\xi_2| \geq \frac{1}{8}, \left|\frac{\xi_1}{\xi_2}\right| \leq 1\right\} \quad (3.5)$$

Thus, the shearlet coefficient can be write as

$$X = \iint 2^{-\frac{3}{2}j}\ g_j^{(u,v)}\overbrace{\left(w(2^{Jv-1})\right)} exp^{\left(2\pi i\left(\frac{n_{1+}n_2}{4_j}\xi_1 + \frac{n_2}{2_j}\xi_2\right)\right)} d\xi_1 d\xi_2 \quad (3.6)$$

Where,

$g_j(u,v)\left(\overline{w(2^{jv-l})} = f(\xi)\left\{\overline{\dot{V}(2^{-2J\xi})w_{j,l}^d}(\xi)\right\}\right)$, It is a pseudo-polar grid of discrete samples.

W is a window feature on a pair of trapezoids, located, $g_j(n_1, n_2)$. The value of a pseudo-polar DFT is the value of, $n_1\ and\ n_2$ are nite sequence of value for a given image $N_{rows}xN_{colums}$ [35] and u, v are the pseudo-polar coordinates $(u,v) \in R^2$ translated as follows.

$$(u,v) = \left(\xi_1, \frac{\xi_2}{\xi_1}\right) if\,(\xi_1, \xi_2)\ \epsilon D_0 \quad (3.7)$$

$$(u,v) = \left(\xi_1, \frac{\xi_2}{\xi_1}\right) if\,(\xi_1, \xi_2)\ \epsilon D_0 \quad (3.8)$$

DST uses the Laplacian pyramid algorithm to filter a given image [37], which is implemented in the spatial domain. The multiscale partition can be achieved by decomposing an image into a low-pass and high-pass filtered image. Sampling the outcome by 4, and then down. Directional localization for several directional components is achieved by translating a window function W to extract the frequency components of the input image. The decomposition of the first stage produces 4 or 8 sub-bands depending on the shearlet filter size chosen. Figure 3.2 demonstrates the frequency-domain implementation of shearlet support for 4 scales.

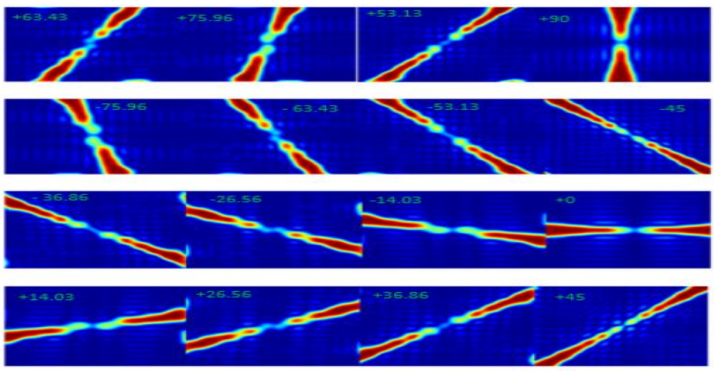


Figure 3-2: Shearlet with fourth level decomposition with 16 direction orientation.

3.8.3 Dual Cross Pattern

An extraction feature of DCP is a new generation image descriptor that is robust to changes in gray scale images and surface rotation. It uses local sampling and pattern encoding to work with the most informative directions within a face frame [11]. As shown in Figure 3.3, local sampling of the DCP is divided into 5x5 blocks and sixteen points are sampled around the central pixel I_c.

The points I_1 to I_8 of the sample are uniformly spaced on the inner circle of radius I_{in}, while I_9 to I_{16} be uniformly spaced on the outer circle of radius Iex. Then, assign a specific decimal number to each of the eight directions.

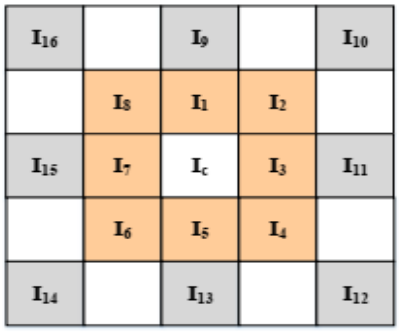


Figure 3-3: Local sampling of DCP.

By evaluating the [I_1-I_8] pixel and the [I_9-I_{16}] pixels separately, a typical histogram defines the dual cross patterns process. We apply the following equations to obtain the set of images.

$$A_{i,j}^{1} = s(I_1, I_c)x2 + s(I_9, I_1) \quad (3.9)$$

$$A_{i,j}^{2} = s(I_2, I_c)x2 + s(I_{10}, I_2) \quad (3.10)$$

$$A_{i,j}^{3} = s(I_3, I_c)x2 + s(I_{11}, I_3) \quad (3.11)$$

$$A_{i,j}^{4} = s(I_4, I_c)x2 + s(I_{12}, I_3) \quad (3.12)$$

$$A_{i,j}^{5} = s(I_5, I_{1c})x2 + s(I_{13}, I_5) \quad (3.12)$$

$$A_{i,j}^{6} = s(I_6, I_{1c})x2 + s(I_{13}, I_5) \quad (3.13)$$

$$A_{i,j}^{7} = s(I_7, I_{1c})x2 + s(I_{11}, I_7) \quad (3.14)$$

$$A_{i,j}^{8} = s(I_8, I_{1c})x2 + s(I_{16}, I_8) \quad (3.15)$$

$$A_1 = A_{i,j}^{1} * 4^3 + A_{i,j}^{3} + 4^2 + A_{i,j}^{5} * 4^1 + A_{i,j}^{7} + 4^0 \quad (3.16)$$

$$A_2 = A_{i,j}^{2} * 4^3 + A_{i,j}^{4} + 4^2 + A_{i,j}^{6} * 4^1 + A_{i,j}^{8} + 4^0 \quad (3.17)$$

Feature _set=conc (histogram$A_{1,}$ + histogramA_2)

Where conc (.) is a function of sub setting or concatenation features. DCP uses histograms of the feature image to formulate two feature images, by used set of features. As a consequence, the function conc (.) function is used to combine histograms of feature images.

3.9 Deep Learning

Deep learning is a series of algorithms for machine learning that help us to learn a layered input model commonly called neural networks. Multiple layers were used to learn the representation of data across multiple abstraction levels in a computational model. It would have the ability to find complicated structures from large data sets to modify their internal membership function using the backpropagation algorithm [38].

In this study [38][39] a model that learns representations from broad unlabeled data is being tried to develop. There are various deep learning including, deep belief networks, convolutional deep neural network, which were used for several applications and it has shown superior performance. Computer vision method, speech recognition, natural language processing and audio recognition are some practical areas [39]. The DBN network that was used in this research work is briefly discussed in this paper.

3.9.1 Deep Belief Networks

The Deep Belief Network (DBN) is a generative probability model that composed of many layers of latent stochastic variables. This allows us to calculate even more complex features of the input signals and uses unlabeled data the high-level extraction of features [40]. As shown in figure 3.4, it is a graphical model where you learn to derive a profound hierarchical representation of training data sets. From the top layers, the undirected bipartite Boltzmann machine forms a guided sigmoid belief network and the lower layers. The method of training consists of two steps. Unsupervised pre-training is the first step and fine-training is supervised in the second step. [40] As a result of the fact that DBN is an RBM stack, each RBM can be trained in DBN in an unsupervised greedy manner for the unsupervised pre-training of DBN. Like the Backpropagation learning algorithm, the supervised fine-tuning layer model structure is trained using different algorithm.

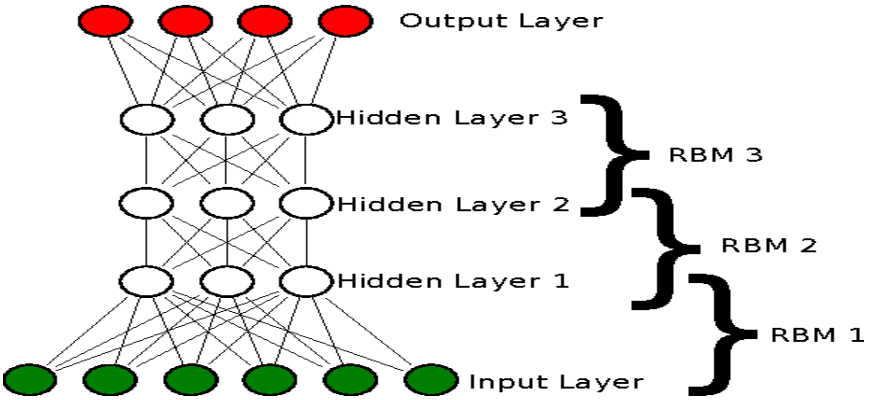


Figure 3-4: Deep belief network.

3.9.2 Restricted Boltzmann Machine

A two layer network knows as the Restricted Boltzmann Machine can used for training set of model [38]. In which stochastic binary pixels are linked by a symmetrical weighted relation to the stochastic binary function. It has a successful training procedure that makes it ideal for DBN essential elements.

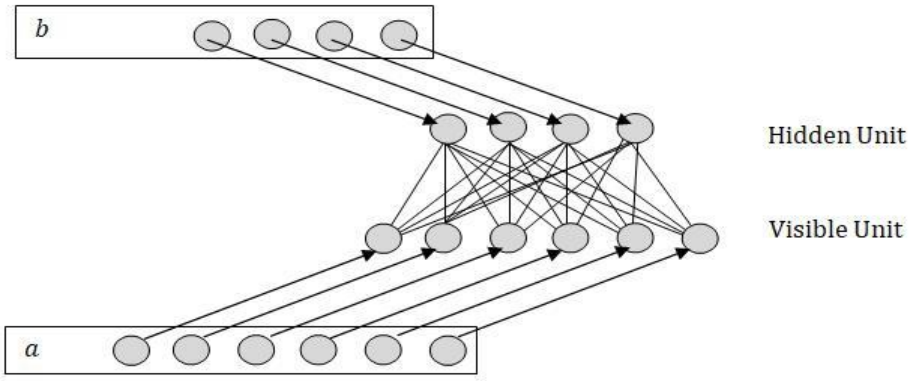


Figure 3-5*:* Restricted Boltzmann Machine

Where only a two-layer network was being used, such as visible units (v) and hidden (h). The pixels toward visible units, while hidden units are the feature detectors. The visible and hidden unit's joint configuration (v, h) has an energy E, as stated as follows.

$$E(v,h) = -\sum_{i\epsilon\ visible} a_i\, v_i \; - \sum_{j\varepsilon\ hidden} b_j\, h_j - \sum_{i,j} v_i\, h_i w_{ij} \qquad (3.18)$$

While v_i and h_j are respectively the binary state of the visible unit i and hidden unit j, their corresponding biases are a_i andb_j, along with w_{ij} is the weight between visible and hidden units.

Joint probability environment for the visible and hidden Restricted Boltzmann Machine energy function unit.

$$P(v,h) = \frac{1}{z}\, e^{-E(v,h)} \tag{3.19}$$

Where,

Z is the division function and can express all possible pairs of visible and hidden vectors summing them up.

$$z = \sum_{v,h} e^{-E(v,h)} \tag{3.20}$$

By summing up all possible hidden vectors, the probability that the network assigns to a vector v, is provided.

$$p(v) = \frac{1}{z}\sum_{v,h} e^{-E(v,h)} \tag{3.21}$$

In order to estimate the gradient on the log likelihood of RBM, Gibbs Markov chain sampling of the visible and hidden variable unit pair has been used. [38] It is the method of consecutive sampling of the visible unit given to all the hidden units, then the visible unit given to the hidden unit until after the end of the chain. The chain begins with t=0 from the input vector, then continues with ($v_{t,}h_{t,}$), where t is the sampling iteration number. The derivative of a training vector's log probability (data) with regards to weight can be written in the following form.

$$\frac{\partial log p(v)}{\partial w_{ij}} = < v_i h_j > data - < v_i h_j > model \tag{3.22}$$

The angle bracket has been used under the distribution defined by the subscript that corresponds to denote expectation. This resulted in a very basic learning rule for the stochastic steepest decent output in the log probability of the training data written as.

$$\Delta w_{ij} = \in (< v_i h_j > data - < v_i h_j > model) \tag{3.23}$$

Where,

$\in$ Is the weight update's learning rate.

This equation indicates that the weight update is the difference between the visible and hidden unit expectations under data distribution and model distribution.

Since there is no direct link between hidden RBM units, it is simple to get the unbiased sample $\in (v_i h_j) data$ data from the training vector. The state of each hidden unit h_j can be determined as set to 1 given the training vector.

$$p(h_j = 1/v) = \sigma(b_j + \sum_i v_i \, w_{ij}) \tag{3.24}$$

It is also possible to measure the sample of each visible unit v_i of the visible layer, provided the hidden layer as.

$$p(v_i = 1/h) = \sigma(a_j + \sum_j (h_i \, w_{ij}) \tag{3.25}$$

RBM trained with contrastive divergence (CD), which is an estimation of the gradient for some particular time of alternatively sampling starting from the visible layer unit. The Gibbs sampling of a single iteration consists of updating every hidden units using equation (3.24), followed by updating the visible unit using equation (2.24).

As the training method state [38] involves setting the state of the visible units to the training vector, the hidden unit is samples from the visible unit in parallel using the equation (3.27). The final stage is a reconstructive procedure that uses an equation (3.28) to sample the visible unit parallel to the hidden unit. As shown below, the weight equation (3.26) can then be simplified.

$$\Delta w_{ij} = \in (< v_i h_j > data - < v_i h_j > recon) \tag{3.26}$$

$$\Delta \alpha = \in (< v_i > data - < v_i > model) \tag{3.27}$$

$$\Delta b = \in (< h_i > data - < h_i > model) \tag{3.28}$$

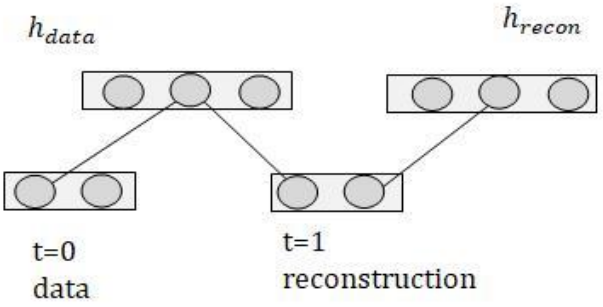


Figure 3-6: Constructive divergence with n=1 [38],

The last RBM is discriminate RBM [38] which model is the joint distribution of input vectors and target class as shown below, as previously DBN is constructed from RBM stucks.

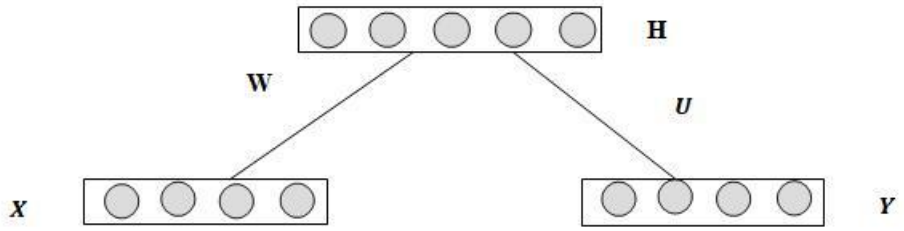


Figure 3-7: Discriminative restricted Boltzmann machine

Where, x is the input vector from which the hidden layer unit of the RBM model, y is the RBM target output, h is the hidden layer, w is the weight relation between the visible and hidden unit of the input vector, and u is the target class unit and hidden unit of the weight connection. The function of energy may also be written as.

$$E(v,l,h,\theta) = -\sum_{i=1}^{V}\sum_{j=1}^{H} w_{ij}\, v_i h_j - \sum_{y=1}^{L}\sum_{j=1}^{H} w_{yj}\, h_j i_y - \sum_{t=1}^{V} a_i v_i - \sum_{t=1}^{H} b_j h_j - \sum_{j=1}^{L} c_y l_i \qquad (3.29)$$

For which V is a visible layer, H is a hidden layer, L is a target class layer, v is a visible layer unit, h is a hidden layer unit, l is a target class layer unit, (a,b,c) is a bias signal for visible, hidden and target class layers, and θ =(w,b,c,d,u). The conditional probability of the hidden unit given by the target class is estimated at.

$$p(l_y = 1/h) = sigmo\left(\sum_{j=1}^{H} w_{ji}\, h_j + c_y\right) \qquad (3.30)$$

Boltzmann machine has two operating phases. Positive phase, this is a holding situation controlled by the machine that is under the control of the training sample and negative phase, this enables the machine to run freely, there is no environmental feedback.

3.9.3 Greedy Layer-Wise Training of Deep Belief Networks

The deep neural belief network is an RBM stack and can be trained one layer at a time for pre-training in an unsupervised greedy manner. It is the way to train deep architecture by separate training each component and stacking them from the layered structure. For DBN training Greedy layer-wise training is the way of quick training for DBN in an unsupervised manner for pre-training. By sequentially training each layer of DBN and feeding lower layer output to the upper layer as input and initializing upper layer weight from lower layer weight trans-pose. Since each DBN layer is made up of RBM, the training of each DBN layer is equivalent to the training of the respective RBM layer. Since this results in a network being better optimized than conventional stochastic gradient descent training. The greedy layer-wise algorithm of unsupervised training for a DBN can be generalized as follows.

a. For model creation, the input vector is h (0) = X for the first RBM of DBN.
b. Using the derived data from the first layer as the second layer's input data. The extracted data may mean either activation data p (h(1) = 1jh(0)) or a set of p(h(1)jh(0)) samples.
c. Train, the second layer as the RBM and hold the mean activation or sample from the first layer as the visible layer of the RBM training results.
d. Repeat step 2 and step 3 for all layer numbers and feed either the mean activations or the samples upwards for each step.

It is unsupervised training, but a model that produces both the mark and the data can be applied to labeled data to learn. A stronger generalization is verified by the initialization of a local minimum that helps to formulate a representation of high-level abstractions of the network input.

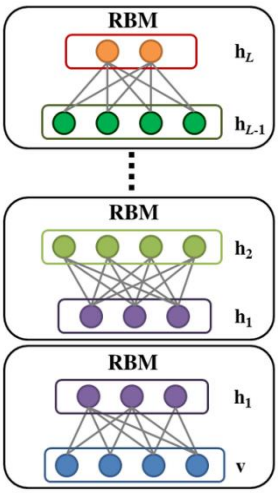


Figure 3-8: Greed layer-wise training procedure

3.9.4 Supervised Fine-turning

At the second stage in the DBN training phase, supervised fine-tuning is considered. It is the way the parameters of the pre-trained network are optimized for better machine efficiency. It utilizes machine learning algorithms supervised by the same training data set and network structure. Back Propagation is the most common supervised algorithm used for DBN fine-tuning. The following measures follow the Back Propagation algorithm.

a. Apply the real vector X of the input signal.
    a. Using the activation function to measure the output signal each hidden and output layer.
    b. Calculate the activation function gradient in each neuron of each layer with the activation function derivative.
b. Establish the network for back-propagation by reversing the signal transmission path.
    a. Replace the function of its derivative activation.
    b. The error between the true and the desired value is the input vector of the former output layer and the current input layer.
c. For all training samples, repeat one and two until the algorithm's stopping requirements are reached.

In its training phase, the back propagation algorithm can stack on the local minimum for some time while trying to decrease the energy function. There should also be some additive values to take off from the local minimum, which are functions known as momentum of the previous weight of the specified relation.

3.10 Classification

During either identification or verification, classification methods are applied and this system's preliminary objective. For LDFR, various machine learning algorithms are used to remove false positives detected as face frame regions for frame images with great success. Most systems use supervised or semi-supervised classification to detect and identify lightness [41]. Researchers have also shown that the combination of many classifiers increases the efficiency of the two systems. Several weak classifiers are also combined to create a stable and strong classifier [41]. There are several ways of method classification approach. The SVM classifiers used in this research paper are briefly discussed in this subsection.

3.10.1 Support Vector Machine-Based Methods

In the machine learning method, the linear support vector machine (SVM) is efficient and widely used. A support vector machine is a supervised method of learning that can be used for problems with both object classification and function approximation, mostly used for classification purposes. It is a two-class linear classifier, [42]. The dot product of two vectors, referred to as an inner product, is expressed as a linear classifier.

$$W^T * X = \sum_i W_i x_i \tag{3.31}$$

Where, X is the data set, w is the vector of weight, b is the bias, and in the data set, small x denotes components. The line classifying the information is known as the hyperplane.

$$f(x) = W^T * x + b \tag{3.32}$$

There are several different types of hyperplanes that classify the data, such as one and four hyperplanes. From all possible hyperplanes, only one of these achieves optimum classification. If we pick one hyperplane at random, it will end up close to one set of data sets compared to another and the outcome would be biased [42]. The notion of maximum margin classifiers was introduced to remove such a constraint. The difference between the support vector and the hyper plan is called a margin. The maximum margin offers better verifiable productivity. Ignore limited and enhanced local classifications [42].

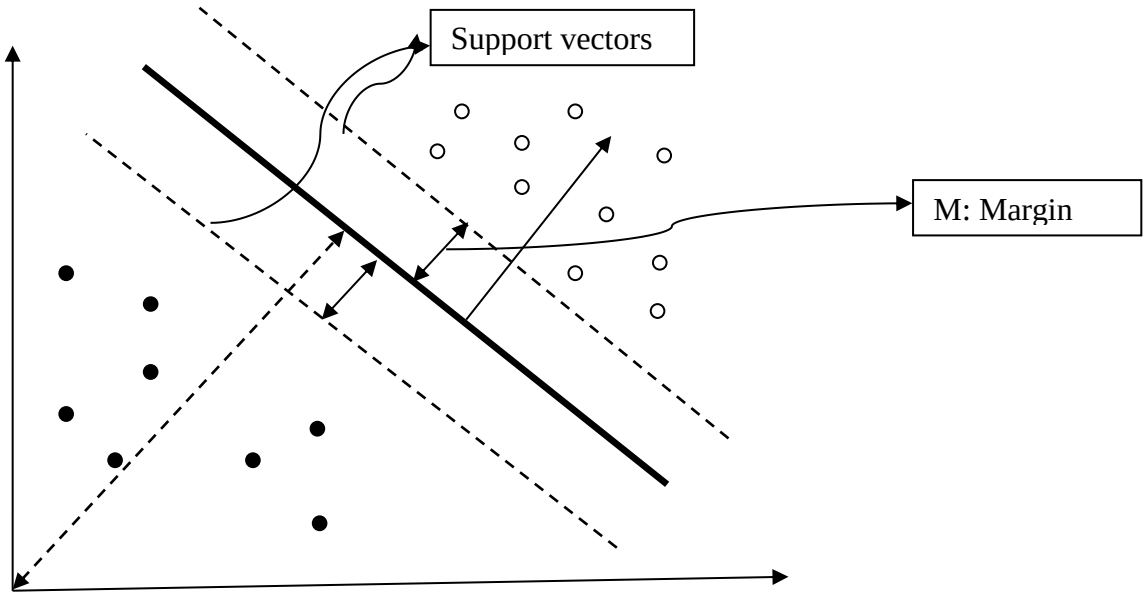


Figure 3-9: Support vector machine.

The hyperplane margin for a given hyperplane between positive and negative samples is defined by the weight vector w with regard to the data set, they may be written as.

$$M = \frac{1}{\|w\|} \tag{3.33}$$

The maximum margin classifier is the discriminating feature which maximizes the geometric margin.$\frac{1}{\|w\|}$ that is tantamount to reducing $\|w\|^2$[42]. This results in the issue of constrained optimization following [42].

$$minimize_{w,b}\ (\tfrac{1}{2}) * \|w\|^2 \tag{3.34}$$

$$\text{Subject to: } y_i(W^T x_i + b) \geq 1\ i = 1, \dots, n$$

Due to the fact that the assumption is that the data is linearly separable, the constraints in this formulation ensure that each example is properly categorized by the maximum margin. But the data is technically not linearly separable, even if it is, a wide margin can be accomplished by allowing certain points to be misclassified by classifiers [42]. Allowing errors, the equation above is replaced by.

$$y_i(W^T * x + b) \geq 1 - \xi_i \tag{3.35}$$

where ξ_i are slack variable that make the margin $1 \geq \xi_i \geq 0$ also known as margin error or misclassified ($\xi_i \geq 1$). The goal is to increase the margin, which is minimizing $\|w\|^2$ will be improved with a term $C \sum_i \xi_i$ to reducing misclassification and margin. The optimization problem becomes.

$$minimiz_{w,b}(1/2) * \|w\|^2 + C \sum_i \xi_i \tag{3.36}$$

The c>0 constant determines the value of increasing the margin and decreasing the amount of slack. This formulation is referred to as SVM Soft-Margin. The data to be categorized is also not linear and the data set cannot be isolated. The kernel is used for non-linear mapping of input data to a high-dimensional space to handle this data set. The one listed below is the common kernel function used in SVM [42]. A common technique for nonlinear modeling is polynomial mapping. The second kernel is typically preferable as it prevents the issue of being zero for the Hessian.

$$K(x_1, x_2) = (x_1^T x_2 + 1)^P \tag{3.37}$$

Gaussian or exponential radial base equation, most commonly with a Gaussian function for radial basis.

$$K(x_1, x_2) = \exp(\frac{-\left|\left|x_1 - x_2\right|\right|^2}{2Q^2}) \tag{3.38}$$

Linear, the long-established MLP, also has a valid kernel representation, with a single hidden layer.

$$(x_1, x_2) = (x_1^T * x_{2)} \tag{3.39}$$

Chapter 4

4 Proposed Approach

One problem found was the design and selection of robust descriptors after a review of studies conducted to face liveness detection and recognition. The key design approach aims to detect in addition, generate the extraction of features from the unlabeled input image frame. The developed image frame with vector is applied in the training process just after learning the image frame representation of large unlabeled data. This architecture can use include the image vector as a feature in LDFR system with different machine learning algorithms.

4.1 The Proposed Approach

The proposed architecture contains five main processes. Key frame extracting, preprocessing, feature extraction, training the model and, finally the cascade LDFR predicted by the trained prediction model. The workflow of the proposed approach is briefly described in the below figure 4.1.

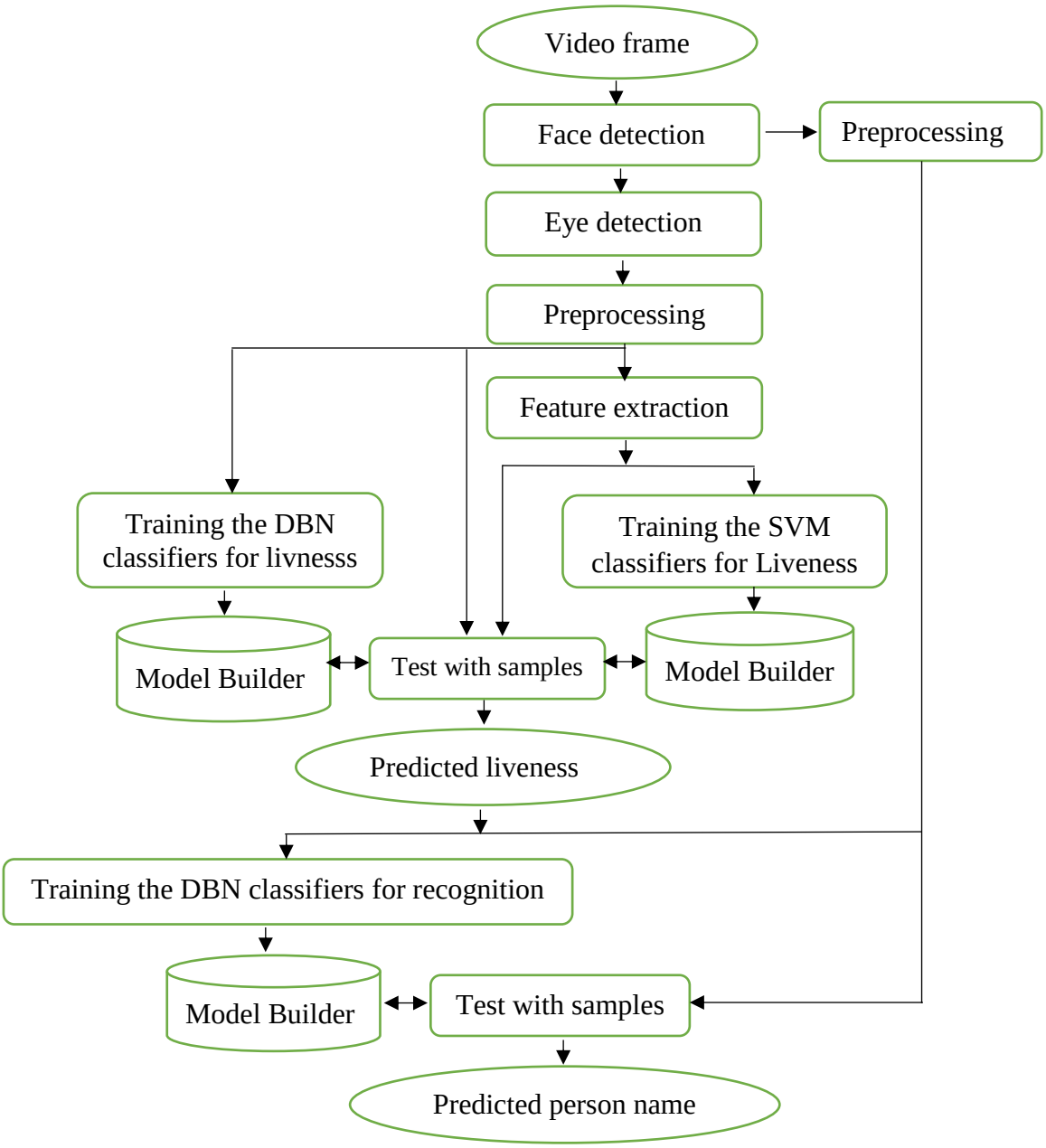


Figure 4-1: Proposed architecture for LDFR.

4.2 Data set Used

We used the CASIA anti-spoofing dataset for evaluation of the study which containing 50 subjects. This includes three image qualities and three fake types' images. A suggested test protocol is also provided which consisting of 7 scenarios, each of them involves only certain samples and summarized below figure 4-2.

Quality Test: This test is intended to measure performance when the quality of the image is set. The samples, specifically, are:

1. Low (L) quality test: {L1, L2, L3, and L4}.
2. Normal (N) quality test: {N1, N2, N3, and N4}.
3. High (H) quality test: {H1, H2, H3, and H4}.

Fake Face Test: Similarly, when fake face types are fixed, this test is to evaluate the performance. These samples include:

1. Warped photo attack test: {L1, N1, H1, L2, N2, and H2,}.
2. Cute photo attack test: {L1, N1, H1, L3, N3, and H3}.
3. Video attack test: {L1, N1, H1, L4, N4, and H4}.

Overall Test: In this test, to give a general and overall assessment, all data are combined together.

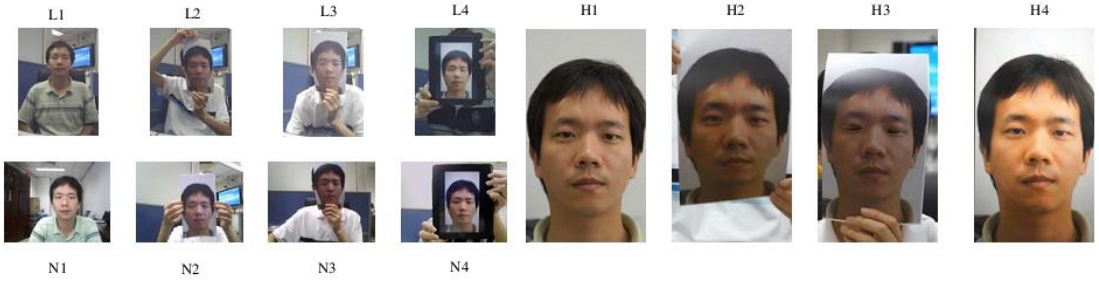


Figure 4-2: All set of subject videos.

4.3 Data set Preparation

There are two types of data sets in the case of liveness detection. The first categories include the right and left closed eye frame status, which has 27,000 total images. The second group includes both right and left the opened eye frame status, which contain 4,500 total images. In the event of recognition, the approach will recognize ten-persons with known dictionary names as well as from

each ten person's extracted 10 frame images. In addition, 40 individuals with 400 images are extracted for unrecognized groups.

4.4 Development Tools

There several development tools in this research work. The tools that have been used include Tensor Flow [43]deep learning library, Keras [43] deep learning library and Scikit library machine learning library.

4.5 Image Acquisition

The method of capturing image frame from video stream refers to image acquisition. On the LDFR system, it has an essential part and often it can involve some problems such as lighting, blur, noise, etc. [44]. In recent days, the efficiency of the camera has improved as a result of the fact that the evolution of technology. But the image quality taken from mobile phones, surveillance cameras and moving objects continues to pose a great deal of challenge.

4.6 Key Frame Extraction

The histogram difference method is used for both key frames, extraction and selection in addition, this popular approach is based on the threshold computing the mean and standard deviation values [45]. A human being's spontaneous resting blink rate is almost 15 to 30 eye blinks per minute when blinks about once every 2 to 4 seconds. Since the average blink lasts about 250 milliseconds [27] with not less than 15 fps, the new generic camera can easily capture the video frame. The video stream has 7 seconds length for all subjects of CASIA anti-spoofing data sets, then we can get 3 times eye blinking status in 30 key frames. Algorithm of Histogram difference is in the algorithm 4-1.

Algorithm 4-1: Key frame extractions using Histogram of difference [45]

$\boldsymbol{Input}: video\ stream$

$\boldsymbol{Output}: key\ frames$

$Total\ frames \leftarrow read\ an\ obj.Frame\ image\ of\ numbers$

$\boldsymbol{for}\ i = 1\ to\ Total\ frame\ image\ \boldsymbol{do}$

$I \leftarrow frame\ image.k;$

$J \leftarrow frame\ image.K + 1;$

$S \leftarrow absdif(I, J)$

end for

$mean \leftarrow mean2(S)$

$Standard\ Deviation \leftarrow std2(S);$

$threshold\ \leftarrow Standard\ Deviation + (Mean * b);$

$\boldsymbol{if}(S > threshold)\boldsymbol{Than}$

$The\ key\ frame\ image\ j\ is\ writen$

end if

In the key frame format, we used the python video reader feature to modify the video stream. The OpenCV library can use for multiple video operations and other powerful video functions. We can pick the threshold value that needs to be used to define key frames using the following equation.

$$\text{Threshold} \leftarrow \text{Standard Deviation} + (\text{Mean} * \text{b}) \qquad (4.1)$$

Where b is a constant value, and the desired value is chosen after the various values of b have been tried. The ladder frame is regarded as the key frame, if the gap between two consecutive frames reaches the threshold. Processes are terminated meanwhile the frame list is empty.

Although several frame image numbers are present in the ant-spoofing data set form. Comparing the sequence of face frames and further using it to see the characteristic distinction between each face frame sequence is simple and useful.

4.7 Face and Eye Part Detection

Methods of face and eye detect refer to the detection of the most concerned part from a given key frame. The biggest issue with extracted the interested section of the face as well as eye are out of focus, blurring, and poor contrast [46]. Techniques such as attentional cascade and integral image construct the algorithm of Viola-Jones [47] it has highly effective by feeding a valid sequence of images created from a standard video stream. The pertained Viola-Jones extracted the interested

part and saved it in JPC format. Algorithm of detecting faces and eyes with an Adaboost trained cascade classifier is in the algorithm 4-2.

Figure 4-3: Frontal view of face and eye detected.

Algorithm 4-2: Detecting face and eye frame using an Adaboost trained cascade [48]

Input: $key\ frame$

Output: $Detected\ face\ and\ eye\ part$

Input a set of gray image: $\{x_1, y_1\}, \dots, \{x_n, y_n\}\ Where\ \ y_i = 0,1\ for\ negative\ and\ positive$

Intialize the weights: $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}\ \ for\ y_i = 0,1\ \ respectively.$

$where, l\ and\ m\ are\ the\ numbers\ of\ negative\ and\ positive$

For $t = 1 \dots \dots \dots ., T$:

1. $Normalize\ the\ weights.$

   $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,i}}\qquad consquently,\ w_{t,}\ is\ a\ probability\ distribution\ function.$

2. $Each\ feature, j, train\ a\ classifier\ h_j\ which\ is\ restricted\ to\ apply\ a\ single\ feature.$

   $Such\ that, the\ \ error\ is\ evaluated\ with\ respect\ to\ \ w_t, \varepsilon_j = \sum_i w_i\ |h_j(x_i) - y_i\ |.$

3. $With\ the\ lowest\ error \varepsilon_t, choose\ the\ classifier,\ \ h_{t,}\ .$

4. $Update\ the\ weights$:

   $$w_{t+1,i} = w_{t,i}\beta_t^{1-e_i}\ Where, e_i = 0\ if\ the\ example\ x_i\ is\ classified\ correctly$$

   $$e_i = 1\ otherwise, and\ \beta_t = \frac{e_i}{1-c_i}.$$

5. $The\ final\ strong\ clasifier\ is$:

   $$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2}\sum_{t=1}^{T} \alpha_t \\ 0 & otherwise \end{cases} \qquad Where, \alpha_t = log^{\frac{1}{\beta_t}}$$

End for

By implementing the pre-trained Viola-Jones system, the input MXN grayscale images of face and eye key frames are collected using a Haar training module in open CV. The algorithm takes a training group of negative and positive images as its input. Finally, powerful characteristics have been produced that can then be used to detect the most interesting part of the face and eye from the image of the frame.

4.8 Preprocessing

Face and eye picture frames taken from the video stream of the ant-spoofing data set will suffer from variations in lighting, changes in size, resolution, to mention a few problems [25]. State-of-the-art LDFR system still struggle with all those problems [50]. It is easier to apply preprocessing methods to resolve those challenges.

4.8.1 Scale

Image scaling refers to the process of enlarges or reduces the physical size of the image frame by changing the number of pixels that contain it addition image scaling can speed up computations [51]. The identified eye status resizes its frame image size to 224x224, with scale factor 1.1, min size (30,30) and face frame image scale factor 1.2, min size to 640 x480 (50,50).

4.8.2 Noise Removal Using Median Filter

Noise removal media filter considers each pixel in the image frame, it looks at its neighbors to determine whether or not it is descriptive of its environs. This replaces the pixel value with the pixel values adjacent to the media. Noise removal is computed by sorting all values in numerical order from the environs neighborhood in addition, replacing the pixel as the middle pixel value. Algorithm of standard media filter is in the algorithm 4-3.

Algorithm 4-3: Standard media filter [31].

$\boldsymbol{Input}$: $Image\ with\ noise$

$\boldsymbol{Output}$: $Image\ with\ enhance\ quality$

let $\mathrm{p(x,y)}$$is\ the\ current\ pixel\ under\ consideration\ and\ W\ is\ sliding\ window.$

$whare : MxN\ is\ the\ size\ of\ the\ image;$

$\boldsymbol{for}\ i = 1 to N\ \boldsymbol{do}$

$\quad \boldsymbol{for}\ j = 1 to N\ \boldsymbol{do}$

$\quad\quad sort(W_{i,j} \dots W_{W,N})$

$P(x, y) \leftarrow media(W_{i,j} \dots W_{W,N})$

end

end

For each of the image frame pixels, the grayscale images are sorted into 3X3 neighborhoods by importing median from skimage.

4.9 Discriminate Feature Extraction

After performing some digital imaging preprocessing techniques, the feature extraction is presented with the normalized and grayscale frame image component. The process of features extraction specified the accuracy of the interested part of the image frame in the compact vector characteristic form. Detailed information about the image frame descriptors for the pixel scale is given in Section 3.8.1-3.80.3.

4.9.1 Local Ternary Pattern

It is an operator that was originally used to obtain a summary of the texture from the image frame and it's commonly used in the processing of images. LTP overcomes the LBP limits insensitive to noise, pose and smooth weak gradients of light that extend LBP to 3-value code [34]. And this has proved to be a highly discriminatory feature compared to LBP. Algorithm of local ternary patterns is in the algorithm 4-4.

Algorithm 4-4: Feature extraction of LTP [34].

Input : $Image\ \ frame$

Output: $Feature\ vectore(FeaVec)$

Let $M \times N\ image\ frame\ size$

for $i = 1\ to\ M$ ***do***

for $j = 1\ to\ N$ ***do***

$Pattern\ histogram\ \ initialize\ H = 0\ for\ each\ center\ pixel\ p(x, y) \in$ I **do**

$p(x, y)\ \leftarrow LTP\big(p(M, N)\big);$

if $Value\ of\ neighbor > Value\ fp(x, y)$ ***than***

setBit value $= 0$

$else$ setBit value $= -1$

else setBit value = 1;

end

end

end

4.9.2 Shearlet Transform

Shearlets transform is a method for directional representation and it's one of the most powerful systems for the effective representation of multidimensional data in recent years. The shearlet transform module includes all applicable transform functions for the 2D case given by the shearlet transformation 2D from the OpenCV Scikit-learn python module, such as forward and inverse transformation. For a broad class of multidimensional data set, this provides optimally sparse representations in addition, compactly supported analyzing functions can be used. Algorithm of shearlet discrete transform is in the algorithm 4-5.

Algorithm 4-5: Feature extraction of SDT [36].

Input : $Image\ frame$

Output: $Feature\ vector\ (FeaVec)$

Initialization: $Compute\ the\ extended\ DST: \widehat{S}\left(R^m(f)\right)\ by\ Set\ m = 0$

Best match: Find out projection of orthogonal Q_{K_m} , $\boldsymbol{p}$ such that

$$K_m = argsp_k \left|\left|Q_{k,p}\left(R^m(f)\right)\right|\right|$$

Update: Compute m + 1 order residual R^m(f);

$R^{m+1}(f) = R^m(f) - Q_{k_m}, p(R^m(f))$

Stopping rule: $If\ ||\boldsymbol{R^{m+1}(f)}||^2 = ||\boldsymbol{R^m(f)}||^2 - \left|\left|\boldsymbol{Q_{k_m}, p\left(R^m(f)\right)}\right|\right|^2 \leq \in^2 ||\boldsymbol{f}||^2$

Then terminate otherwise m=m+1 and go to1.

In this case, the number of iterations is K and in each iteration, P nonzero coefficients are introduced. As a result, the KP coefficient is created in the MP step after K iterations. The cumulative cost of the computation of this adaptive approximation method is therefore $O\left((2^{M+2} + 2)N\right)$. The cumulative cost of the computation of this adaptive approximation scheme also is $O\left((2^{M+2} + 2)N\right)K$.

4.9.3 Dual Cross Pattern

DCP decreases the effect of variations in illumination and computes the features at both the holistic and in part levels. The face frame should consist of two key sections of local sampling and pattern encoding, which are face frame descriptors, to design DCP in the most informative directions. Algorithm of dual cross pattern is in the algorithm 4-6.

Algorithm 4-6: Feature extraction of DCP [11].

Input: $Image\ \ fram$

Output: $Holistic\ feature\ vector\ (FeaVec)$

The input image frame parameter

$if\ \ narrgin < 4, \quad Radius = [4\ \ 6];\ end;$

$if\ nargin < 3, \quad$ RegionColNum $=\ 9;\ end;$

$if\ nargin < 2, \quad$ RegionRowNum $=\ 9;$ end;

if size (Input_Im, 3) $==\ 3$

Encode the input image frame

$feaVec = [];$

$OffSet = [0\ pi/4];$

$for\ i = 1{:}\ length(OffSet)$

feaVec(: , i) = calDCPVec_sub(Input_{Im}, RegionRowNum, RegionColNum, 4, Radius, OffSet(i))

Block by block arrange the feature vectore

$blockNum = RegionRowNum * ResgionColNum;$

$dimPerNum = size(feaVec, 1)/blockNum;$

$result = [];$

$for\ i = 1{:}\ size(feaVec, 2)$

$\quad temp = reshape(feaVec(:,1), dimPerBlock, blockNum)$

$\quad result = [result{:}\ temp];$

$\quad end$

$feaVec = result(:);$

end

In Figure 3.3, a local sampling of DCP is performed. In each direction, two pixels are sampled as I_1-I_8. In an inner circle of radius, it is equally spaced while I_9-I_{16} is evenly distributed with the

external radius on the external circle. Two measures are considered to encode the above sampling point. First, textural information is separately encoded in each of the eight directions. Second, to form DCP codes, patterns in all eight directions are combined.

4.10 Feature Selection

The method of selecting features is the step of selecting the best subset of relevant features, although not all features are necessary for the construction of the classification model. Reduced dimensional data sets, accelerated preparation and enhanced results. Just five of them have been considered special in this research from all extracted features. To eradicate the redundancy and irrelevance of the process of filtering features based on subset evaluation algorithms. Subset approaches are more effective in eliminating redundant feature [52]. Algorithm of subset feature selection is in the algorithm 4-7.

Algorithm 4-7: Feature selection.[52]

Input : $Total\ features$

Output : $Selected\ best\ features$

Input :

S − Functionality f with data sample X, |X| = n

J − maximized measure of assessment For m = 1,2, … . . n.

GS – Generation operator of successor

Output :

Repeat

L: = Strategy of search (L, GS (J), X);

X': = {best of L according to J};

If J (X') = J (Solution) or (J (X') = J (Solution) and |X'| < |Solution|) then

X': = {best of L according to J}

Solution: = X';

Until Stop(J, L).

A preprocessing approach that applies without any classifiers is the strongest subset of features. Two phases of the filter method consist of. First phase, feature relevance, graded by feature score based on some univariate or multivariate characteristic of feature evaluation. In the case of

univariate, irrespective of other features and multivariate ranked multiple features at the same time, each feature is individually ranked. The techniques rely solely on data features such as variance, correlation, mutual knowledge, accurate. Second phase, low ranked features is filtered out and the remaining features are kept. The image frames were available for training and testing after the feature selection process was finished. Break up between the proposed solution and machine learning models for the various models that were highlighted as options to complete this task.

4.11 Evaluation Metrics

Based on the similarity between the comparisons of the expected output of the model with that labeled data, the performance method of LDFR evaluation metrics is evaluated. A basic measure used in most studies is the f-score. It can measure a combined measure of accuracy and recall. Precision is the number of objects that are properly labeled as true positive (TP) over the number of elements that are labeled as a false positive (FP).

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{4.2}$$

The number of true positives over the number of true positives plus the number of false negatives will simply be determined by a recall (FN).

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}} \tag{4.3}$$

These quantities are also related to the (f-measure) ranking, defined as the precision and recall harmonic mean.

$$\text{F}_{\text{measure}} = 2 * (\frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}) \tag{4.4}$$

A classifier's overall accuracy, recall, and f-score is determined by using a weighted average. The weighted average is determined by summing, multiplying the score for each class by the number of instances and dividing the result by the number of instances. In all this experimentation and assessment train, test break, the experiment performed this research work using 75% frames as training data and 25% as testing data.

4.12 Model Training

The extracted descriptors were fed into SVM to train the model, as previously stated. The key component of this architecture is the training process. To construct a model, the training process uses various machine learning algorithms. Training of the eye image frame with its compact, feature vectors is the input to this process. The model building process begins to form a model after the training data is supplied. The features are expressed in the form of a floating number and various machine learning algorithms can be used.

4.12.1 DBN Model Training

DBN is a multilayer and the stochastic generative model that is created by training a stack of RBMs. Network learning is based on the probability of transmitting data processing by keeping patterns into the visible neurons and trained as training data using the hidden activities. It can be achieved good performance if there is regular structure in the data [53] . DBN model used in this research for both tasks of face liveness detection and recognition approach. Algorithm of deep belief network is in the algorithm 4-8.

Algorithm 4-8: Training deep belief network [40].

Input : $Selected\ keyframes$

output : $Predicted\ class$

$Input\ training\ dat\ set\ \ (X_t, Y_t), t = \{1,2,3 \ldots\ldots n\}, \qquad n = Input\ vector\ length$

Initialization

$Intilize\ \ M\_\ Epoch, LR_{Error}, Visible\ and\ hidden\ biase, Radome\ Weight\ Matrix$

Start positive phase

$$E(V, H1) = Sum(W_{ij} * V_j)$$

$$P\left(H1 = \frac{1}{V}\right) = Sigm\big(HB + E(V, H1)\big)$$

End V to H

From H1 to H2

$$E(H1, H2) = Sum(W_{ij}^{T} * V_i)$$

$$P\left(H2 = \frac{1}{H_1}\right) = Sigm\big(HB + E(H1, H2)\big);$$

End H1to H2

While $Error\ is\ not\ converged$

Start Negative phase

$H1 \sim \mathrm{P}(\mathrm{H1}/\mathrm{H2})$

$V1 \sim \mathrm{P}(\mathrm{V}/\mathrm{H1})$

Update weight and Biases

$$\mathrm{Updt}W_{ij} = W_{ij} + LR\left(+ veP\left(\frac{V}{H}\right) - veP\left(\frac{H}{V}\right);\right)$$

$$\mathrm{UpdtBV} = \mathrm{BV} + \mathrm{LR}(\mathrm{V1} - \mathrm{V2})$$

$$\mathrm{UpdtBH1} = \mathrm{BH1} + \mathrm{LR}(\mathrm{H1} - \mathrm{H2})$$

$$\mathrm{UpdtBV} = \mathrm{BV} + \mathrm{LR}(\mathrm{V1} - \mathrm{V2})$$

The unsupervised pre-training approach of DBN is established by applying the RBM training processes. The method begins by default to initialize the parameters while the value of the parameter is modified by an individual iteration of the training epoch starting. The RBM method first extracts the hidden unit of the input data, then uses the hidden layer units to create the visible unit, and then extracts the hidden unit from the reconstructed hidden units. In addition, it updates the weight of the connection as well as the respective visible and hidden bias, using the above consecutive feature extraction and construction results.

Deep belief neural network architecture for liveness detection, input is 32x32x2 patch size. The network has three hidden parts. Activation of all layers was set to ReLU, but sigmoid for the last layers. Since the architecture of face liveness detection has two outputs such as the real and fake person's classes. The parameters of the model layers were randomly initialized using glorot initialization, optimized using SGD, num_epoch =50, and learning rate =10^{-3}.

Deep belief neural network architecture for face recognition, input is 100x100x2 patch size. The network has three hidden parts. Activation of all layers was set to ReLU, but sigmoid for the last layers. As there are two output classes in the face recognition architecture those are recognized and unrecognized person's classes. The parameters of the model layers were randomly initialized using glorot initialization, optimized using SGD, num_epoch =30, and Learning rate =10^{-2}.

4.12.2 SVM Model Training

The SVM model applies to the 32x32x2 input patch size of liveness detection system. Since there are three types of methods of feature extraction that fed to the SVM model. The following steps are suggested for using of Svm classifier [54].

a. Transform the data to an SVM package format.

Each data instance needs to be represented in a vector of real number SVM. In the case of this study, feature extraction is float valued, as previously mentioned, so that there is no need for transformation.

b. Conduct easy data normalize.0

SVM input data is executed within a normal range, usually from 0 to 1 or from -1 to 1. In order to ensure that the values for each dimension are scaled to lie approximately within this range. The benefit of normalization such as preventing wide numeric range attributes from dominating those in smaller numeric ranges and avoiding numerical difficulties during the calculation. The default standardized parameters of the fitsvm function are set to true in this case.

c. Considers the Kernel RBF in addition, select best c and g parameters using cross validation to train the entire data set for testing.

For classification consideration, the RBF kernel is sometimes recommended [54]. The selected RBF kernel due to the fact that its ability to handle non-linear boundary classification issues. Then the value of Gamma (g) is set to 80 after many tests.

d. Test.

The entire 50 subject data sets, splits using, such as 75 % frame image as training data and 25 % is used as testing data to achieve the classification result.

In the Python Sklearn library, the svm classification algorithm is implemented. The sklearn svm library is used to import the svm class. Finally, testing is performed using test data sets.

4.13 Model Testing

The prediction input is the discriminate feature extraction output, which is a file containing the vector frames of the features. It predicts face liveness detection by taking this as input. From the

result of liveness detection, face recognition system will be done, and the final output is displaying the name of the person that corresponds to the feature vector.

Chapter 5

5 Experimental Result and Discussion

In this section, all experimental procedures, tools and experimental scenarios that used to assess our hypothesis are discussed. Can measure our new proposed architecture of LDFR system by any standard benchmarks.

5.1 Experimental Setup

The type of machine was used to implement the suggested method is Linux environment software and the hardware specifications are given below in Table 5-1.

Manufacturer	Dell
Model	OPTIPLEX 760
Processor	Intel ® Core (TM) 2Duo E8600 @ 3.40 GHz
Memory (RAM)	4.00 GB (3.89 Gb usable)
Operating System	Ubuntu 18.04 LTS

Table 5-1*:* Specification of the machine further used to experiments.

Two kinds of experimental studies. Firstly, an experiment in liveness detection. In this case, the training data set contains two kinds of categories, namely real and fake each folder has 4500 and 27000 eye image frames respectively in size of 224x224. Secondly, Cascading from the result of liveness detection experiment recognition approach was done. For every 50 subjects ten face frames are extracted, so that there are totally 500 face frames in the size of 640x480. From the seven scenarios of CASIA ant-spoofing data set, the overall protocol test is considered for the proposed assessment method. Assume that there is no overlap between the training and the testing of a data set. As mentioned above, we make the experiment by feeding three types of feature extraction to the SVM model, then classify the proposed approach of LDFR system using both SVM and DBN classifiers. The reason that two classifiers were used to give a rational and fair comparison between the state-of-the-art. For an evaluation of the statistical data, this research does not adopt the suggested set of training and testing a data set. However, we combine the 50 subjects in each training and testing phases.

5.2 Baseline Experiment

The performance of LDFR approaches was assessed through the institute of the automation Chinese academy of sciences (CASIA) anti-spoofing data set [17]. The reason for selecting this baseline research experiment is. Due to the fact that it has several variations in data collection compared to other ant-spoofing data sets and it's the most efficient of LDFR systems. The number of instances per class is shown in table 5.2.

	Number of instances for Liveness		Number of instances for Recognition	
Classes	Real	4,500	Recognized	100
	Fake	27,000	unrecognized	400
	Total=32,000 instances			

Table 5-2: Training data used in baseline experiments.

5.3 Experimental Scenarios

To assess the hypothesis, several experiments are conducted. In general, these experiments can be grouped into three. The purpose of the first set of experiments is to evaluate the accuracy of image enhancing techniques. The second group of experiments involves evaluating the significance of descriptors. The last group of experiments conducted on deep neural networks using the training data set. In addition, assessing the performance at each three experiments. The first, second and third experiments are used to respond for RQ1, RQ2, and RQ3 respectively.

In response to RQ1 [Preprocessing for LDFR], the RQ1 focuses on achieving high accuracy in the uncontrolled video frame. In response this research question, this research has implemented an efficient preprocessing system, namely the elimination of noise media filters. The advantage of the preprocessing step of LDFR system has the ability to remove some of the noise that may occur due to the fact that embedded image factors. Therefore, the use of an effective preprocessing algorithm makes the face liveness detection and recognition system robust.

In response to RQ2 [Feature Extraction for LDFR], this research implemented three types of feature extraction for LDFR classification purpose. The feature vector of each descriptor in the labeled training data is extracted and the training data contains a feature vector. The DCP descriptor outperforms the LTP and SFE descriptors in situations of multiple reduce scales. Since

LTP and SFE performance drop significantly as the resolution decreases, DCP operator performance remains high despite the radical loss solution. Due to the fact that the nature of DCP as a generalization of the LTP and the SFE, future work will focus on assessing many other generalizations identified by DCP. Their capacity to solve another problem in LDFR technology systems with constrains.

In response to RQ3 [Deep Neural Networks for LDFR], this research used the DBN network architecture. The preprocessing descriptors frame is used for training. By changing the neural network layers, the model is formed with 50 iterations, three times, then the layer that gives the best is selected. For the evaluation of all 50 subject data set, 75% images frame as training data and 25% frame images as testing data. Finally, a deep neural network of multilayer perceptron is developed to verify the coherence of the results. Using automatically generated features, an initial experiment was conducted to assess the information contained in the neural network.

5.4 Challenges of Face Liveness Detection and Recognition

The following challenges were addressed after some experiment on LDFR system are carried out in addition, after some literature review. However, it determines the possible actual situations where such systems would be used. The following challenges have been dealt with.

- Two identical twins are hard to distinguish.
- A system is identified as a particular frontal view face image.
- As humans become aging (distrust), their texture or face shape was changed.
- Only expressionless front face will be shown on LDFR system.
- The recognition of the frontal image will be done using a single known image for each classification class.
- The automated face liveness detection and recognition system should be combined with fully automated LDFR technology. The face recognition subsystem displays a slight degree of invariance such as scaling and rotation errors when extracting key frames from the face detection subsystem.

Unfortunately, this may specify binding conditions in LDFR of challenges, it may not be possible to adhere strictly to these conditions during the implementation of a system in the real-world.

5.5 Results

5.5.1 Preprocessing for LDFR

The first experimental scenario is performed with preprocessing and non-preprocessing methods using CASIA ant-spoofing data sets. The face frame candidate classifier is trained and tested using the DBN model and the following results are obtained.

Method + Classifier	Class	Precision	Recall	F-Score
Preprocessing + DBN	Real	98.4	97.8	98.6
	Fake	97.2	98.2	97.9
	Weighted averages	**97.8**	**98.0**	**98.2**
Without Preprocessing + DBN	Real	79.4	70.6	97.9
	Fake	85.7	99.8	75.1
	Weighted averages	**82.5**	**85.2**	**86.5**

Table 5-3: Result of liveness detection using preprocessing (scale 100%).

To realis the impact of the preprocessing method on the performance of LDFR system, it conducts the experiment with non-preprocessing method. Table 5-3 shows that, the significant observation of the poor performances when the liveness detection system is applied with non-preprocessing method.

Summary results of liveness detection using preprocessing for RQ1 is depicted in Figure 5-1

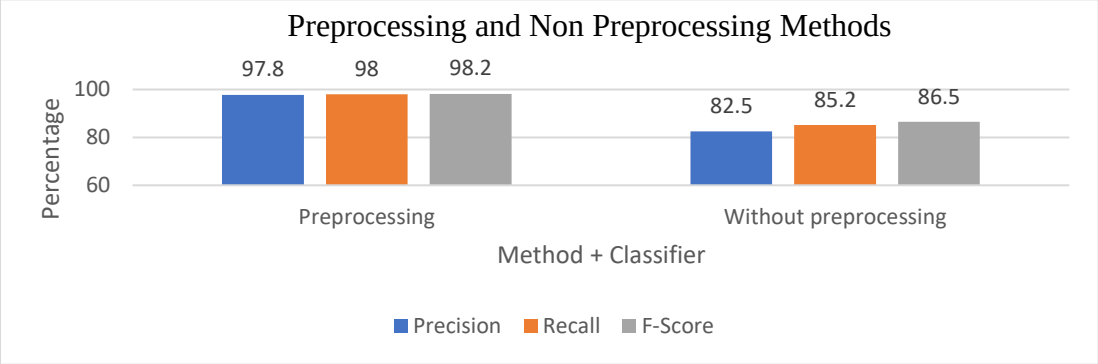


Figure 5-1: Summary results for RQ1.

Ultimately, it gains increase in performance of 15% in precision, 12.8% in recall and 11.7% in f-score, due to the fact that the addition of the preprocessing method. The above result of the experiments carries out to respond to RQ1. It shows an important observation that the methods have significantly poor performances when the frame image system is used without preprocessing.

5.5.2 Feature Extraction for LDFR

The method of feature extraction used to describe the data set with sufficient accuracy. The perfect optimized feature extraction is key to constructing the model. In general, these are dimensional reduction techniques for a set of data sets. This experimental result shows a complete comparison between three types of descriptors as far as the support vector machine connected.

Descriptors + Classifier	Class	Precision	Recall	F-Score
DCP+SVM	Real	95.7	94.1	96.7
	Fake	95.6	95.4	95.8
	Weighted averages	**95.6**	**94.7**	**96.2**
LTP+SVM	Real	90.7	85.3	91.3
	Fake	95.2	88.6	92.9
	Weighted averages	**92.9**	**86.95**	**92.1**
SFE+SVM	Real	93.4	92.4	96.2
	Fake	91.4	93.8	94.7
	Weighted averages	**92.4**	**93.1**	**93.9**
Without Descriptor + DBN	Real	98.4	97.8	98.6
	Fake	97.2	98.2	97.9
	Weighted averages	**97.8**	**98.0**	**98.2**

Table 5-4: Result of liveness detection using inputs of different features (scale 100%).

Summary results of liveness detection using various descriptors for RQ2 is depicted in figure5-2.

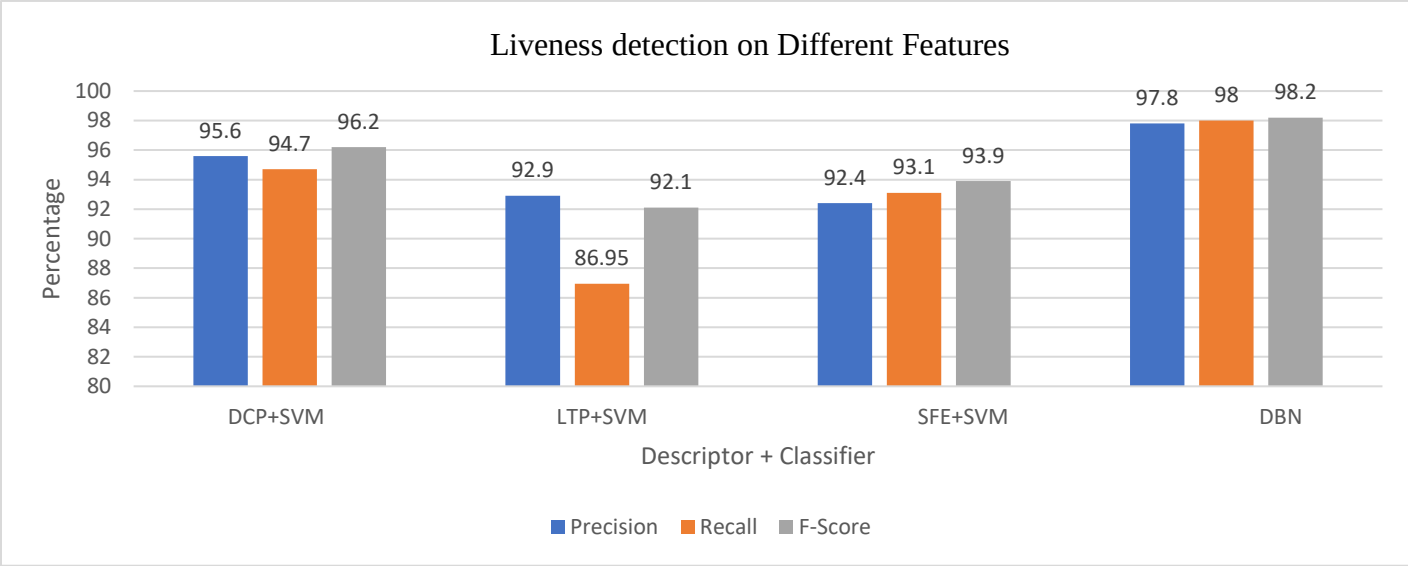


Figure 5-2: Summary results for RQ2.

This experiment conducted to respond RQ2 and the result demonstrated that discriminative feature extraction of DCP improves liveness detection system compare to the SFE and LTP descriptors. DCP descriptors show high performance on liveness detection with f-score of 96.2%, 92.1%, and 93.8% respectively with DCP, LTP and SFE. Consequently, DCP outperforms SFE and LTP by 4.1% and 2.4% respectively.

5.5.3 Deep Neural Network for LDFR

The deep belief network (DBN) is a network of generative stochastic artificial neural that can learn probability distribution with an input data set. In this thesis, the architecture of the deep neural network is constructed using DBN model alongside keras deep library by applying tensorflow as backend. The final result categorized used both DBN and SVM models. And these experiments result give rise to a full comparison between the two types of classifiers within the corresponding descriptors.

Classifier	Recognition	Precision	Recall	F-Score
DBN	**Weighted averages**	**94.3**	**95.1**	**94.7**

Table 5-5: Results of face recognition using DBN classifier (scale 100%).

Summary results of face recognition using deep belief network for RQ3 is depicted in figure5.3

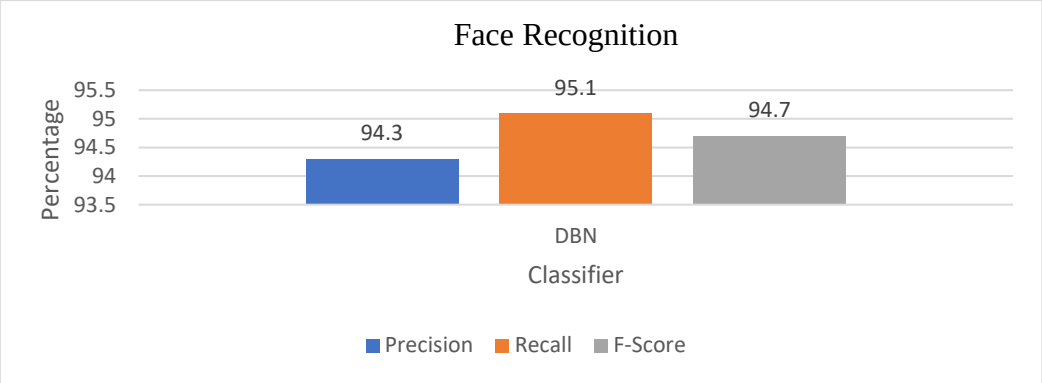


Figure 5-3: Summary results for RQ3.

In LDFR system, there are two types of classifiers and totally 50 iterations are ongoing. The DBN method yields better result compared to the state-of-the-art method with 97.8% precision, 98.2% recall and f-score 98.5% were achieved. Again liveness detection was also evaluated using an SVM model, where 95.6% precision, 94.7% recall and f-score of 96.2% were obtained using a DCP descriptor, as shown in table 5.4. DCP method has a good computational advantage over SFE and LTP descriptor. However, BDN with f-score performance evaluation gets high registered camper with DCP alongside SVM and it outperforms 2.3%. In addition, in case of recognition. We identify the face image for 50 subjects and the results as indicated in table 5.5, where 94.3% precision, 95.1% recall with an f-score of 94.7% achieved on the DBN classifier.

5.6 Threats to External Validity

Selecting the size of key frame and classifiers can threaten the validity of this research. The size of face and eye detection is in JPG format with 640 x480, 224 x224 pixel size respectively. There might be another size and format of the frames that will result in a better performance. We used the DBN deep network architecture with two classifiers, but there may be possibly other different deep learning classifiers with a different set of parameters to achieve high performance.

Chapter 6

6 Conclusion and Feature Works

6.1 Conclusion

The LDFR approach is an essential component and, it's consists of several biometric security systems. Adequate identification and verification of LDFR system will enhance the performance of other biometric security system. This research aims at liveness detection based anti-spoofing method in face recognition using CASIA ant-spoofing data set.

As part of this research, we have implemented and automated LDFR system. The cascading proposed architecture contains five main steps. Key frames extracting, preprocessing, feature extraction, training the model and, finally the cascaded LDFR predicted by the trained prediction model. In addition, two classifiers and three feature extraction were used the reason behind that is to provide a rational and fair comparison between the state-of-the-art.

In order to assess this hypothesis, various experiments were conducted. Each series of experiments aims to responses the research questions toward the use of preprocessing, feature extraction and deep neural network for LDFR.

We have constructed and automated liveness detection based anti-spoofing methods in face recognition. The first experiment conducted on the state-of-the-art preprocessing. Based on the selected baseline experiments, the preprocessing method where, precision 97.8%, recall 98.0%, and f-score 98.2% was achieved. And again without preprocessing method was evaluated, where 82.5% precision, 85.2% recall, and 86.5% was obtained. The preprocessing method showed higher outcomes and outperformed by 11.7% f-score. The second experiments carried out on the state-of-the-art descriptors. The first phase of liveness detection 95.6% precision, 94.7% recall, and 96.2% f-score were achieved on SVM classifier with DCP descriptor. Thus, DCP exceeds SFE and LTP descriptors by 4.1% and 2.4%, respectively using the evaluation performance of f-score. The last experiment conducted on the state-of -the-art between two classifiers beside that DBN given precision 97.8%, recall 98.2%, and f-score 98.5% for liveness detection method. Consequently, DBN classifier outperforms SVM classifier alongside DCP descriptor by 2.3% of

f-score. The second phase of face recognition method with DBN classifier was also evaluated, where 94.3% precision, 95.1% recall with f-score of 94.7% was achieved.

Finally, from the observations of the experimental results, we arrived at the following conclusion.

- A spontaneous eye blink rates can capture signs of real and fake face that further used to identify of liveness detection approach.
- The neural network architecture of DBN has a robust automatic feature extraction in addition, it was achieved effective result for labelling and specific pixel classification.
- Among the three kinds of feature extraction, DCP feature extraction exceed for both LTP and SFE feature extraction methods.
- Based on the result experiments of classifiers, DBN architecture models obtained highest f-score than SVM classifier.

6.2 Feature Work

This dissertation opens perspective for future research.

- The techniques proposed in this thesis are designed and choose robust descriptors for LDFR approaches with CASIA anti-spoofing data set. We pulled key frames from the data set by breaking down a video stream into frames. But such method of translating is not recommended. Instead, extract the key frame of the video stream as image sequence just as we did before. The extraction step must be replaced by a real time. Consequently, real-time system more necessary index of research work in the further.
- The feature descriptor proposed in [6], this research has proven their worth in addition, other works have been successfully used [22]. We strongly advised the next researcher to use in image processing method DCP description. Due to the fact that there is the large sample size of DCP descriptor that can compute using full-length or integral images. Make the descriptor more computationally effective since it's included more texture around the neighborhood. The descriptor could also be extended to a multidimensional descriptor that includes more than one frame in the video sequence.
- LDFR approach, three types of feature extraction with two classifiers were used to compute the features and to perform the classification respectively for the cascading systems. It's

preferable to use the high execution time approaches like parallel computing system. Further it used to score superior result in system performance.

- Based on the principle of splitting to conquer algorithms, we recommended that the next researcher must be use multi-stage classification rather than binary classification.
- The next researcher is expected to develop a new framework that uses an end-to-end identification learning method that will be used for pixel-level imaging methods.
- Performance of the proposed approach evaluated using only CASIA face anti-spoofing data sets. But, these are not good guidelines for researchers, next researchers must be using another anti-spoofing data sets, this help us to fair and rational compared the performance of reference algorithms.

Reference

[1] J. Komulainen, A. Hadid, and M. Pietik, “Context based Face Anti-Spoofing.”

[2] S. Marcel, M. S. Nixon, and S. Z. Li, *Handbook of Biometric Anti-Spoofing-Trusted Biometrics under Spoofing Attacks*. 2014.

[3] S. A. C. Schuckers and D. Ph, “Spoofing and Anti-Spoofing Measures,” vol. 7, no. 4, pp. 56–62, 2002.

[4] A. Hadid, “Face biometrics under spoofing attacks: Vulnerabilities, countermeasures, open issues, and research directions,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, no. Cmv, pp. 113–118, 2014, doi: 10.1109/CVPRW.2014.22.

[5] J. Bigun, H. Fronthaler, and K. Kollreider, “Assuring liveness in biometric identity authentication by real-time face tracking,” *Proc. 2004 IEEE Int. Conf. Comput. Intell. Homel. Secur. Pers. Safety, CIHSPS 2004*, no. July, pp. 104–111, 2004.

[6] W. Zhao and A. Rosenfeld, “<Zhao, W., & Rosenfeld, A. (2003). <Zhao - Face Recognition.pdf>, 35(4), 399–458.Zhao - Face Recognition.pdf>,” vol. 35, no. 4, pp. 399–458, 2003.

[7] J. Cassell *et al.*, “Animated conversation: Rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents,” *Proc. 21st Annu. Conf. Comput. Graph. Interact. Tech. SIGGRAPH 1994*, no. May, pp. 413–420, 1994, doi: 10.1145/192161.192272.

[8] G. Pan, L. Sun, Z. Wu, and S. Lao, “Eyeblink-based anti-spoofing in face recognition from a generic webcamera,” *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, doi: 10.1109/ICCV.2007.4409068.

[9] J. Yang, Z. Lei, S. Liao, and S. Z. Li, “Face liveness detection with component dependent descriptor,” *Proc. - 2013 Int. Conf. Biometrics, ICB 2013*, 2013, doi: 10.1109/ICB.2013.6612955.

[10] W. Kim, S. Suh, and J. J. Han, “Face liveness detection from a single image via diffusion speed model,” *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2456–2465, 2015, doi: 10.1109/TIP.2015.2422574.

[11] C. Ding, J. Choi, D. Tao, and L. S. Davis, “Multi-Directional Multi-Level Dual-Cross Patterns for Robust Face Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 518–531, 2016, doi: 10.1109/TPAMI.2015.2462338.

[12] C. H. Chan and M. A. Tahir, “Multiscale Local Phase Quantization for Robust Component-Based Face Recognition Using Kernel Fusion of Multiple Descriptors,” vol. 35, no. 5, pp. 1164–1177, 2013.

[13] Y. Shi, X. Ren, S. Yang, and P. Gong, “A generalized Kernel Fisher Discriminant framework used for feature extraction and face recognition,” *2016 12th Int. Conf. Nat. Comput. Fuzzy Syst. Knowl. Discov. ICNC-FSKD 2016*, pp. 1487–1491, 2016, doi: 10.1109/FSKD.2016.7603396.

[14] X. L. Dl *et al.*, “Face Liveness Detection and Recognition Using,” *2016 IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 874–877, 2016.

[15] W. Bao, H. Li, N. Li, and W. Jiang, “A liveness detection method for face recognition based on optical flow field,” *Proc. 2009 Int. Conf. Image Anal. Signal Process. IASP 2009*, pp. 233–236, 2009, doi: 10.1109/IASP.2009.5054589.

[16] A. K. Singh, P. Joshi, and G. C. Nandi, “Face recognition with liveness detection using eye and mouth movement,” *2014 Int. Conf. Signal Propag. Comput. Technol. ICSPCT 2014*, pp. 592–597, 2014, doi: 10.1109/ICSPCT.2014.6884911.

[17] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, “A face antispoofing database with diverse attacks,” *Proc. - 2012 5th IAPR Int. Conf. Biometrics, ICB 2012*, pp. 26–31, 2012, doi: 10.1109/ICB.2012.6199754.

[18] S. T. V. Parthasaradhi, R. Derakhshani, L. A. Hornak, and S. A. C. Schuckers, “Time-series detection of perspiration as a liveness test in fingerprint devices,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 35, no. 3, pp. 335–343, 2005, doi: 10.1109/TSMCC.2005.848192.

[19] O. Kahm and N. Damer, “2D face liveness detection: An overview,” *Proc. Int. Conf. Biometrics Spec. Interes. Group, BIOSIG 2012*, 2012.

[20] A. Alotaibi and A. Mahmood, “Deep face liveness detection based on nonlinear diffusion using convolution neural network,” *Signal, Image Video Process.*, vol. 11, no. 4, pp. 713–720, 2017, doi: 10.1007/s11760-016-1014-2.

[21] R. F. Nogueira, R. De Alencar Lotufo, and R. Campos MacHado, “Fingerprint Liveness Detection Using Convolutional Neural Networks,” *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 6, pp. 1206–1213, 2016, doi: 10.1109/TIFS.2016.2520880.

[22] H. Larochelle and Y. Bengio, “Classification using discriminative restricted boltzmann machines,” *Proc. 25th Int. Conf. Mach. Learn.*, pp. 536–543, 2008.

[23] J. T. Wassan *et al.*, “Intelligent Computing Theories and Application,” vol. 10362, pp. 421–427, 2017, doi: 10.1007/978-3-319-63312-1.

[24] L. Aryananda, “Recognizing and remembering individuals: Online and unsupervised face recognition for humanoid robot,” *IEEE Int. Conf. Intell. Robot. Syst.*, vol. 2, no. October, pp. 1202–1207, 2002.

[25] E. Hjelmås and B. K. Low, “Face detection: A survey,” *Comput. Vis. Image Underst.*, vol. 83, no. 3, pp. 236–274, 2001, doi: 10.1006/cviu.2001.0921.

[26] C. V. Sheena and N. K. Narayanan, “Key-frame Extraction by Analysis of Histograms of Video Frames Using Statistical Methods,” *Procedia Comput. Sci.*, vol. 70, pp. 36–40, 2015, doi: 10.1016/j.procs.2015.10.021.

[27] C. N. Karson, “Spontaneous eye-blink rates and dopaminergic systems,” *Brain*, vol. 106, no. 3, pp. 643–653, 1983, doi: 10.1093/brain/106.3.643.

[28] V. Paul and J. Michael, “Prefacio Prólogo,” 2001, doi: 10.1109/CVPR.2001.990517.

[29] H. Schneiderman, “Applied to Faces and Cars,” *Faces*, 2000.

[30] D. Kuykin, V. Khryashchev, and I. Apalkov, “Modified progressive switched median filter for image enhancement,” *19th Int. Conf. Comput. Graph. Vision, Graph. - Conf. Proc.*, pp. 303–304, 2009.

[31] Z. Wang and D. Zhang, “Progressive switching median filter for the removal of impulse noise from highly corrupted images,” *IEEE Trans. Circuits Syst. II Analog Digit. Signal Process.*, vol. 46, no. 1, pp. 78–80, 1999, doi: 10.1109/82.749102.

[32] D. S. Bolme, J. R. Beveridge, M. Teixeira, and B. A. Draper, “The CSU face identification evaluation system: Its purpose, features, and structure,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2626, pp. 304–313, 2003.

[33] D. Wen, H. Han, and A. K. Jain, “Face spoof detection with image distortion analysis,” *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 4, pp. 746–761, 2015, doi: 10.1109/TIFS.2015.2400395.

[34] W. Tan, X., & Triggs, “Recognition Under Difficult Lighting Conditions To cite this version : HAL Id : hal-00565029 Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions,” 2011.

[35] E. S. Han and A. goleman, daniel; boyatzis, Richard; Mckee, *済無No Title No Title*, vol. 53, no. 9. 2019.

[36] G. Easley, D. Labate, and W. Q. Lim, “Sparse directional image representations using the discrete shearlet transform,” *Appl. Comput. Harmon. Anal.*, vol. 25, no. 1, pp. 25–46, 2008, doi: 10.1016/j.acha.2007.09.003.

[37] Y. Teng, F. Liu, and R. Wu, “The research of image detail enhancement algorithm with laplacian pyramid,” *Proc. - 2013 IEEE Int. Conf. Green Comput. Commun. IEEE Internet Things IEEE Cyber, Phys. Soc. Comput. GreenCom-iThings-CPSCom 2013*, pp. 2205–2209, 2013, doi: 10.1109/GreenCom-iThings-CPSCom.2013.416.

[38] C. M. Bishop, “Pattern Recognition and Machine Learning Springer Mathematical notation Ni,” *Springer-Verlag New York, Inc.,, Secaucus, NJ, USA*, p. 9, 2006, [Online]. Available: http://cds.cern.ch/record/998831/files/9780387310732_TOC.pdf.

[39] L. Deng *et al.*, “IEEE Signal Processing Magazine - November 2007,” *IEEE Signal Process. Mag.*, vol. 24, no. 99, pp. c1–c1, 2008, doi: 10.1109/msp.2007.4317458.

[40] G. E. Hinton, “A practical guide to training restricted boltzmann machines,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7700 LECTU, pp. 599–619, 2012, doi: 10.1007/978-3-642-35289-8-32.

[41] H. Hepağuşlar, D. Özzeybek, S. Özkardeşler, A. Taşdöğen, S. Duru, and Z. Elar, “Propofol and sevoflurane during epidural/general anesthesia: Comparison of early recovery characteristics and pain relief,” *Middle East J. Anesthesiol.*, vol. 17, no. 5, pp. 819–832, 2004.

[42] V. Jakkula, “Tutorial on Support Vector Machine (SVM),” *Sch. EECS, Washingt. State Univ.*, pp. 1–13, 2011, [Online]. Available:

http://www.ccs.neu.edu/course/cs5100f11/resources/jakkula.pdf.

[43] J. Fox, Y. Zou, and J. Qiu, “Software Frameworks for Deep Learning at Scale,” *Intern. Indiana Univ. Tech. Rep.*, 2016.

[44] T. E. Boult and W. Scheirer, “Long-Range Facial Image Acquisition and Quality,” pp. 169–192, 2009, doi: 10.1007/978-1-84882-385-3_7.

[45] G. Rathod and D. Nikam, “An Algorithm for Shot Boundary Detection and Key Frame Extraction Using Histogram Difference,” *Int. J. Emerg. Technol. Adv. Eng.*, vol. 3, no. 8, pp. 155–163, 2013, [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:An+Algorithm+for+Shot+Boundary+Detection+and+Key+Frame+Extraction+Using+Histogram+Difference#0.

[46] P. Viola and M. Jones, “MERL-A MITSUBISHI ELECTRIC RESEARCH LABORATORY Rapid Object Detection Using a Boosted Cascade of Simple Features Rapid Object Detection using a Boosted Cascade of Simple Features,” 2004, [Online]. Available: http://www.merl.com.

[47] W. M. Fang and P. Aarabi, “Robust real-time audiovisual face detection,” *Multisensor, Multisource Inf. Fusion Archit. Algorithms, Appl. 2004*, vol. 5434, p. 411, 2004, doi: 10.1117/12.545934.

[48] J. Li and Y. Zhang, “Learning SURF cascade for fast and accurate object detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3468–3475, 2013, doi: 10.1109/CVPR.2013.445.

[49] A. Mohssen, J. Hossen, R. Abdul, S. Ogla, and M. M. Ali, “Face Detection by Using OpenCV ’ s Viola -Jones Algorithm based on coding eyes ةيمزراوخ لاوي ف زذوج د لع سا سا بو ساحلا ي ف رف ش ة ذ وي علا ف شك هجو لا مادخ ت ساب ةب ت كم ةي جمرب ةحوت فم ةي ؤرلا ل ةي,” vol. 58, no. 2, pp. 735–745, 2017.

[50] R. Gross and V. Brajovic, “An Image Preprocessing Algorithm for Illumination Invariant Face Recognition,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2688, pp. 10–18, 2003.

[51] R. Wu, S. Yan, Y. Shan, Q. Dang, and G. Sun, “Deep Image: Scaling up Image Recognition,” 2015, [Online]. Available: http://arxiv.org/abs/1501.02876.

[52] E. Proceedings *et al.*, “QUT Digital Repository : Copyright c 2007 , Australian Computer Society , Inc . This paper appeared at the Probabilistic Logic under Uncertainty,” vol. 65, no. January, pp. 101–110, 2007.

[53] G. Batres-Estrada, “Deep Learning for Multivariate Financial Time Series,” *Master’s thesis, KTH, Math. Stat.*, p. 94, 2015.

[54] C. Yang, G. N. Odvody, C. J. Fernandez, J. A. Landivar, R. R. Minzenmayer, and R. L. Nichols, “Evaluating unsupervised and supervised image classification methods for mapping cotton root rot,” *Precis. Agric.*, vol. 16, no. 2, pp. 201–215, 2015, doi: 10.1007/s11119-014-9370-9.