IMPROVEMENT OF FACIAL RECOGNITION ACCURACY USING EYE-LIDS MOVEMENT AND TENSORFLOW MODEL

SYAZWAN SYAFIQAH BINTI SUKRI

UNIVERSITI SAINS MALAYSIA

2021

IMPROVEMENT OF FACIAL RECOGNITION ACCURACY USING EYE-LIDS MOVEMENT AND TENSORFLOW MODEL

by

SYAZWAN SYAFIQAH BINTI SUKRI

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

May 2021

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious, the Most Merciful, first and foremost, I would like to express my tremendous gratitude to Allah S.W.T for giving me His blessing, patience, guidance, and strength from the start until this end where I managed to finish this MSc thesis. The completion of this thesis also would be impossible without the assistance and full supports from various parties who have willingly lent me their hands. First of all, I would like to thank my dedicated supervisors, Dr. Nur Intan Raihana binti Ruhaiyem as main supervisor and Dr. Ahmad Sufril Azlan bin Mohamed as co-supervisor for their endless support, guidance, and valuable pieces of advice throughout my MSc study. Their patience and willingness to assist me are very much appreciated. Besides, from the deepest of my heart, I would love to thank both my beloved parents, Mr. Sukri bin Jaafar and Mrs. Faridah binti Abd Rahman for their endless love, prayers, and tolerance. They are always encouraging me during my hard times and difficulties and also never failed to always remind me about being a good and successful person. My next appreciation goes to the Dean, Deputy Dean, and all administrative staff of the School of Computer Sciences, USM for their assistance during my study. Not to forget, the panels, Prof. Dr. Putra bin Sumari, Dr. Ahmad Sufril Azlan bin Mohamed and Dr. Mohd Nadhir bin Ab Wahab for their valuable feedback and comments during the proposal review. Last but not least, I would like to thank all my colleagues for their help and useful ideas sharing to improve my research study. Thank you all for the tremendous positive feedbacks given along with the continuous support received.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

x_a	anchor image
x_p	image of the same subject
x_n	image of different subject
f	mapping learned by a model
α	margin enforced between the positive and negative pairs'

LIST OF ABBREVIATIONS

AE	Auto Encoder
AI	Artificial Intelligence
AMS	Attendance Management System
CNN	Convolutional Neural Network
CPU	Central Process Units
DBN	Deep Belief Network
DL	Deep Learning
FDD	Frequency Dynamics Descriptor
GPU	Graphics Processing Units
HoG	Histogram of Oriented Analysis
KLT	Kanade-Lucas-Tomasi
KNN	K-Nearest Neighbors
LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
LFW	Labeled Faces in Wild
ML	Machine Learning
MLP	Multi-Layer Perceptron
PCA	Principal Component Analysis
RBM	Restricted Boltzmann Machine
RNN	Recurrent Neural Network
VAE	Variational Auto Encoder

LIST OF APPENDICES

PENAMBAHBAIKAN PENGECAMAN MUKA MENGGUNAKAN PERGERAKAN KELOPAK MATA DAN MODEL *TENSORFLOW*

ABSTRAK

Dalam kajian penyelidikan ini, prestasi sistem pengecaman muka masa nyata dengan pembelajaran mesin, serta prestasi setiap pengelas (*classifier*) *Haarcascade* berdasarkan ketepatan dan kelajuan telah diselidik. Salah satu teknologi dalam pembelajaran mesin yang dipanggil pembelajaran mendalam (*deep learning*) digunakan untuk sistem pengecaman wajah masa nyata kerana teknologi pengecaman wajah yang mendalam telah meningkatkan prestasi terkini. Model pra-terlatih yang dikenali sebagai *FaceNet* telah digunakan dan teknik '*triplet-loss*' digunakan untuk mengenakan margin antara setiap sepasang wajah daripada orang yang sama dengan wajah lain. Dalam erti kata lain, ia mengurangkan jarak antara jangkar dan positif daripada identiti yang sama dan memaksimumkan jarak antara jangkar dan negatif daripada identiti yang berbeza. Selain itu, prestasi sistem ini juga diselidik dengan menggunakan rangka kerja *Tensorflow* bagi meningkatkan prestasi sistem dengan menggunakan Unit Pemprosesan Grafik (GPU). Data yang dilabel *Labeled Faces in Wild* (LFW) telah digunakan sebagai penanda aras untuk menguji prestasi sistem pengecaman wajah dalam kajian penyelidikan ini. Selain itu, satu percubaan awal telah dijalankan untuk menilai prestasi pengelas (*classifier*) *Haarcascade* supaya pengelas (*classifier*) terbaik dipilih daripada segi ketepatan dan kelajuan. Telah didapati bahawa *haarcascade frontal face default* mempamerkan prestasi terbaik berbanding *haarcascade frontalface alt* dan *haarcascade frontal face alt2* dengan bilangan muka yang tepat dikesan dan purata masa terpendek yang diambil untuk mengesan wajah.

IMPROVEMENT OF FACIAL RECOGNITION ACCURACY USING EYE-LIDS MOVEMENT AND TENSORFLOW MODEL

ABSTRACT

In this research study, the performance of the real-time face recognition system with machine learning, as well as the performance of each *Haarcascade* classifier based on accuracy and speed were investigated. The subset of machine learning called deep learning was employed in the real-time face recognition system as the deep face recognition technology has improved the state-of-the-art performance. A pre-trained model named *FaceNet* was used and the triplet loss technique was employed to impose a margin between every pair of faces from the same person to other faces. In other words, it minimizes the distance between the anchor and the positive from the same identity and maximizes the distance between the anchor and the negative from different identities. Furthermore, the performance of the system was further investigated by implementing the *Tensorflow* framework to improve the system performance by the usage of the Graphics Processing Unit (GPU). Labeled Faces in Wild (LFW) dataset was used as the benchmark to test the performance of the face recognition system. Furthermore, a preliminary experiment was conducted to evaluate the performance of *Haarcascade* classifiers so that the best classifier can be chosen in terms of accuracy and speed. It was found that *haarcascade* frontalface default exhibited the best performance compared to *haarcascade* frontal face alt and *haarcascade* frontalface alt2 with accurate number of faces detected and shortest average time taken to detect faces.

CHAPTER 1

INTRODUCTION

In this chapter, the overview of the Attendance Management System (AMS), facial recognition system, machine learning, and deep learning are discussed. There are four general steps involved in the facial recognition system which include (i) face detection, (ii) normalization, (iii) feature extraction, and (iv) facial matching. The introduction of machine learning is briefly described in conjunction with the relationship between machine learning and deep learning. The brief implementation of deep learning in the facial recognition system from other works also are discussed. Furthermore, this chapter covers the research problem, research motivation, research objectives, and research scope. The research problem covers the changes from the traditional way of taking attendance to the biometric-based attendance management system and a comparison of the limitation of other biometric recognition systems compared to the facial recognition system. Research motivation covers the motivation by previous work done in improving the facial recognition system. There are three main objectives to be achieved at the end of this research work. Besides, the scope and limitations of this research work are specified in the research scope.

1.1 Overview

An Attendance Management System (AMS) is a system that handles and manages the attendance or presence, and it has been applied in many sectors including educational institutions, public and private sectors. The way of AMS implementation has been improvised from time to time to ease its usage and shorten its time taken. Before we heard of automated AMS, we had manual methods of taking and recording attendance by using paper-based such as signatures, calling

names, and mark on the paper. Other than that, the time clock method also had been introduced by punching the attendance's card into the time clock besides identification card scanning. However, these methods were considered to be time-consuming and easily exposed to errors. The errors can occur when the signatures were fabricated and misused of an identification card by somebody else besides its owner. Thus, the AMS now is automated by using biometric-based recognition methods as this method has obtained attention for its high-level security in the attendance management system to overcome the flaws of previous AMS.

Fingerprint, facial, iris, and retina recognition are several biometric-based recognition methods that have been introduced and implemented in today's technologies especially in access control purposes and also AMS. Fingerprint recognition is considered as one of the most popular and successful biometric recognition compared to others (Alsaadi, 2015). However, the drawback of this biometric method is the performance of the fingerprint system can be negatively influenced due to the issues of dirt, tear, and cuts that can make the fingerprint sensor becomes dirty and easily affect the ridges and minutiae of a fingertip (Meng, Wong, Furnell, & Zhou, 2015). Next, for iris recognition, the process of capturing the iris is done by using a special camera that does the iris scanning. Although the distinctiveness of the iris pattern provides an effective recognition scheme of individuals, it is intrusive and its accuracy can be decreased by occlusion factors such as glasses and eye lenses, apart from the high cost of implementation (Harakannanavar, Renukamurthy, & Raja, 2019). Eventually, the Attendance Management System (AMS) has now moved to facial recognition. The advantages of the facial recognition method include low-cost implementation and setup. Unlike fingerprint and iris recognition, facial recognition requires no additional scanner as

most technologies like smartphones and computers come with a pre-installed camera that can be used directly for the facial scanning process. Furthermore, facial recognition is also a contactless biometric recognition and easy to integrate into any authentication applications.

Biometric recognition that emphasizes the facial has gained a lot of attention from many sectors especially researchers, developers, and even the government. These are very useful in security measures and access controls in today's modern era. With the help of the latest and growing machine learning technologies, this biometric recognition system is more robust, accurate, and has faster performance.

1.1.1 Facial recognition

Facial recognition is categorized under physiological behavioral characteristic recognition, same as fingerprint, iris, retina, and hand recognition. These biometric recognition systems have a common mechanism where they are divided into two stages; (i) enrollment and (ii) release. First, the biometric features of an individual are captured before a unique template is created for the individual and stored in a database. This unique template is derived from the mathematical analysis using specific algorithms relevant to the type of biometric recognition system. Next, in the release stage, a comparison is made between the sample data and the unique data stored in the database. The result leads to either verification or identification type of decision.

Also, facial recognition is a biometric recognition system that analyzes the characteristic of facial obtained from an image of a video. Facial recognition stands out from other biometric recognition systems because it is non-invasive and requires

no contact with the system, and these lead to high user acceptance. Compared to iris recognition, facial recognition can be operated with a longer distance between the camera and the individual. In today's technologies, facial recognition has been actively implemented purposely for access controls and security measures. The facial recognition process can be divided into four main steps, (a) facial detection, (b) normalization, (c) feature extraction and, (d) facial matching or recognition.

1.1.1(a) Face detection

In this process, the face(s) is/are captured from a photo or video. One or more faces can be detected in the image or video and marked with a bounding box to indicate the face(s).

1.1.1(b) Normalization

The face(s) located in the previous step (1.1.1(a)) is/are normalized by scaling and rotating to make it consistent with the training images in the database. The facial signature is obtained by using facial landmark estimation or detection with 68 specific points. The key to this facial signature is the distance between the eyes and the distance from forehead to chin.

1.1.1(c) Feature extraction

The meaningful data from the normalized face(s) (1.1.1(b)) are extracted and the 'noises' are left out and ignored. From the facial landmark in (1.1.1(b)), the facial signature is used to perform the computational comparison. Feature extraction is also referred to as encoding.

1.1.1(d) Face matching

All the gathered data are used to perform feature matching processes against the faces in the database. The facial recognition returns the matched face.

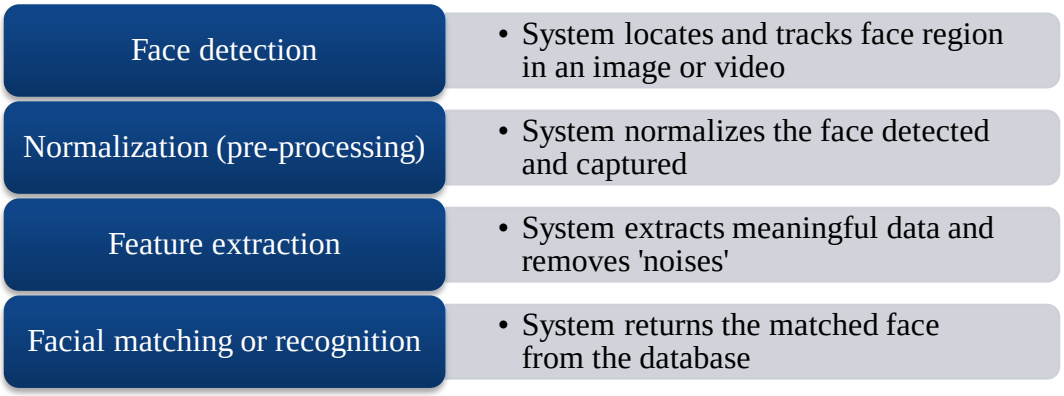

Figure 1.1 General steps in standard facial recognition

1.1.2 Machine learning and deep learning

Machine learning is part of Artificial Intelligence (AI) whereas Deep Learning (DL) is the sub-field of the Machine Learning (ML) family. The benefit of the machine learning algorithm is that it can learn from previous experience according to some tasks and performance measures (Amaro, Nuno-Maganda, & Morales-Sandoval, 2012). Machine learning generally involves classification tasks which can be referred to as pattern recognition. Classification allows the machines to learn to automatically recognize complex patterns, to distinguish data based on the patterns, and to make intelligent decisions (Amaro et al., 2012). Hence, facial recognition is summarized to be applying pattern recognition techniques (Bishop, 2006). Figure 1.2 shows the relations between Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL).

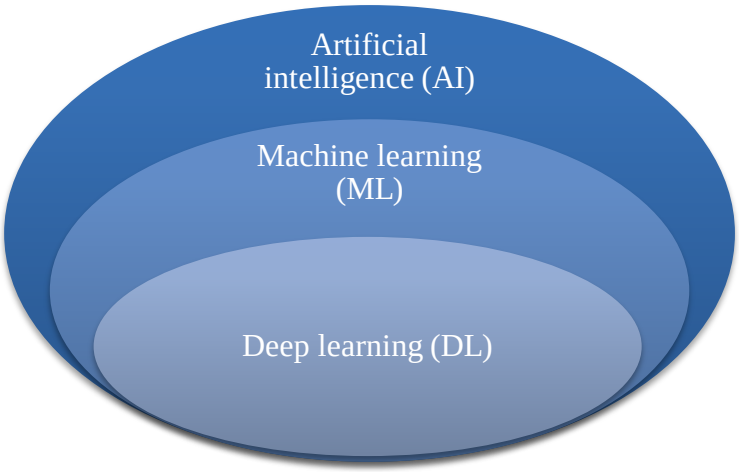

Figure 1.2 Relations between Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL)

On the other hand, compared to traditional machine learning algorithms that are linear, deep learning algorithms are stacked in a hierarchy of increasing complexity and abstraction (Kharkovyna, 2019). Deep learning consists of multiple layers of neural networks and uses a backpropagation method to learn. Some several networks and systems had been developed using deep learning algorithms such as AlexNet (Krizhevsky, et al., 2017), DeepFace (Taigman, et al., 2014), and FaceNet (Schroff, et al., 2015).

1.1.3 Deep learning in facial recognition

Nowadays, developers and researchers are focusing on integrating and implementing the facial recognition framework with machine learning technologies. The modern and latest approach of developing one is using the deep learning approach which is based on learning data representations. Facial recognition has shifted to a deep learning approach when for the first time, *DeepFace* achieved an accuracy percentage of 97.35% which nearly to human performance of 97.53% on

the unconstrained conditions (Taigman et al., 2014). From the result, *DeepFace* has achieved state-of-the-art accuracy by using Labeled Faces in Wild (LFW) as the benchmark. Alongside deep learning, Google has created a framework named *Tensorflow* to create deep learning models (Martín et al., 2016).

1.2 Research problem

Traditional Attendance Management System (AMS) such as the signature method has many flaws especially to the administration such that the attendance can be fabricated where a person can simply sign the attendance form on behalf of another person. Besides, the usage of an identification card to mark the attendance also can be easily falsified by a person who could scan the card for another person as the card is a physical thing that can be passed around. Thus later, biometric-based security is introduced and has obtained attention for its high-level security in the attendance system.

Among the biometric-based recognition methods, facial recognition is a better option compared to the others such as fingerprint and iris recognition. Table 1.1 shows the simple comparison between facial recognition and the other two types of biometric recognition.

Table 1.1 Simple comparison between facial, fingerprint, and iris recognition

	Facial recognition	**Fingerprint recognition**	**Iris recognition**
Scanner	Contactless	Not contactless	Contactless
Distance from scanner	Not very close with a certain distance	Very close with no gap	Close
Cost	Low, can use a pre-installed camera or additional camera	Can be high, must use an additional scanner	High, iris scanner

From Table 1.1, facial recognition has stood out in terms of its easy implementation and setup. This is one of the reasons facial recognition has been chosen to be implemented with the Attendance Management System (AMS).

Many algorithms have been modeled to run facial recognition. The traditional facial recognition algorithms such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Local Binary Pattern (LBP) (Mahmood, et al., 2017) have been introduced to replace the manual ways of AMS before the wide usage of Artificial Intelligence (AI) technologies been spread into this field. These traditional ways are referred to as the holistic methods where the entire facial region is used and they relied on hand-crafted features such as edges and texture descriptors (Trigueros, et al., 2018). However, these algorithms show low accuracy. Wagh et al. (2016) stated that the accuracy was 79.65% while Sukri, et al., (2017) stated that the accuracy result of the facial recognition by using PCA was 70%. Besides, Ara, Simul, and Islam, (2018) stated that the facial recognition accuracy achieved 95% accuracy by using Convolutional Neural Network (CNN) while Saypadith and

Aramvith, (2018) stated that the accuracy achieved 90.29%. Both PCA and CNN algorithms show a huge difference in accuracy, thus proved that **PCA has low accuracy of recognition**. Apart from that, the problem with PCA is the need to train the model whenever a new training dataset is added into the model (Sukri et al., 2017). This is time-consuming and with the big dataset, it will be very troublesome and tedious to retrain the model manually.

Furthermore, another element within facial recognition is **spoof detection**. This is one of the problems of the current AMS where most of the current AMS does not implement spoof detection. A spoofing attack is usually done during the early stage of the facial recognition process which is facial detection. A simple example of a spoof attack is a person who shows another person's photo in front of the camera that capturing the facial, then the person will gain recognition from another person's credential. Spoof detection can use eye blink detection as one of the methods to detect spoof attacks (Sukri, et al., 2017). However, the performance was still considered as low as a full system. Other than eye blink detection, spoof detection can be done using additional external hardware such as a depth Infrared sensor (IR). Microsoft Kinect is an example of IR that captures the depth of images that contributed to the depth of image. Nevertheless, this external hardware is sensitive to sunlight and that makes it is not suitable to be used for outdoor applications (Karbasi et al., 2016).

1.3 Research motivation

The current Attendance Management System (AMS) is still using paper-based and this is time-consuming and also using a lot of papers that eventually affect the natural resources. Although some are using the biometric-based approach, many

of them are using fingerprint recognition which has some flaws like non-contactless with the scanner. This will question the hygiene factor. Furthermore, in the facial recognition field, the algorithms have been developing from time to time from simple algorithms to more sophisticated algorithms, from traditional to advanced technologies. The previous research was using Principal Component Analysis (PCA) as the recognition algorithm and the result exhibited was low in terms of accuracy. Apart from that, it also required to be retrained manually for every new dataset entry. Thus, with the evolution of machine learning technologies which now is deep learning, it has motivated this research to be carried out to improve the work on real-time facial recognition. Besides, spoof attacks also one of the major problems in the current system even in a real-time environment. Thus, spoof detection needs to be strengthened to avoid any security breaches.

1.4 Research objectives

The general objective of this research is to propose real-time facial recognition using deep learning with spoof detection. Expected achievements to fulfill the objective are:

1. To model real-time facial recognition using Deep Learning (DL)
2. To investigate the effectiveness and reliability of using eye blink detection as spoof detection to detect spoof attack by observing the ability of the research work in detecting the spoofing attack
3. To evaluate the performance of facial recognition process using Deep Learning (DL) approach compared to Principal Component Analysis (PCA) by comparing the recognition accuracy from both DL and PCA approaches

1.5 Research scope

The scopes are outlined to ensure that the research is conducted within its expected and intended boundary and limitation. This is to guide and make sure that the research is moving in the right direction to achieve the specified objectives.

The **first** and main scope of this research is to propose real-time facial recognition using Deep Learning (DL) with spoof detection.

The **second** scope of this research is to detect human faces using the *Haarcascade* classifier as the first step of the facial recognition process. The face is detected from the real-time environment.

The **third** scope of this research is to detect spoof attacks using the eye blink detection method. The eye blink detection result which is either blinked or not will determine whether the person is real or is using a photo.

The **fourth** scope of this research is using the *Tensorflow* framework with Graphics Processing Unit (GPU) alongside the *FaceNet* Deep Learning (DL) algorithm.

1.6 Thesis organization

The thesis consists of five chapters. Chapter one of the thesis introduces the research background for overall research work along with an overview of the Attendance Management System (AMS), facial recognition system, Machine Learning (ML), and Deep Learning (DL). Apart from that, the implementation of DL by other research works, the transition from the traditional attendance system to the current DL implementation, and traditional ways of taking attendance to the biometric-based ways of taking attendance also have been introduced. The objectives and scope of this research work also have been stated in this chapter one.

Chapter two is about the literature review on face detection, spoof detection, and facial recognition. The methods used for detecting human faces are discussed based on four categories that include knowledge-based, feature invariant, template matching, and appearance-based method. Spoof detection is also discussed based on four categories that include motion-based, texture-based, image quality analysis, and other cues methods. Facial recognition is discussed based on three categories; holistic, feature-based, and deep learning method. Then, the Deep Learning (DL) method is further discussed by focusing on the effect of dataset size and method, model and framework towards the facial recognition performance.

Chapter three describes the overall methodology of the research work. This includes the selection of classifiers used for performing facial detection. A preliminary experiment setup for selecting the classifier is explained in this chapter. Apart from that, the method used for detecting spoof attacks also is discussed. The model and framework used for this research work are stated and the evaluation methods and experimental setup for each stage are explained in chapter three.

Chapter four presents important data and results in detail. The performance of each classifier in terms of accuracy and speed is presented. Based on the factors that could affect the performance of the facial recognition stated in chapter two, the result in terms of accuracy performance is portrayed in this chapter and further discussed. The result performance is compared with other research works to highlight this research contribution.

Chapter five summarized the research work that has been discussed in the previous chapters and the research objectives are restated to highlight the aim of this research work. The recommendations and suggestions for further improvement and

enhancements are highlighted for better performance of facial recognition using the latest machine learning technologies.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter discusses the process of the face is tracked and detected, the way of spoof attack is handled and the methods and techniques in recognizing and authenticating a face. Moreover, the methods used in the attempt of spoofing will also be discussed to give an overview of how spoofing has been done and the importance to overcome it. The related and previous works by other researchers that share the same goal in recognizing a face will also be discussed.

2.1.1 Face detection

Before the face can be detected by the system, the face acquisition process is done by using any available camera devices such as webcam and surveillance cameras. However, this face acquisition process will not be discussed in this chapter as it is just a simple and direct process. After the face acquisition process, the face detection process is done since a face needs to be located in the input source before it is recognized. Face detection involves two processes which first the whole input source is examined to find the "face" followed by a localization procedure to provide a more accurate estimation of the exact scale and position of the face (Neeraj & Sugandha, 2014). Face detection algorithms can be categorized into four categories:

a) knowledge-based methods
b) feature invariant
c) template matching
d) appearance-based methods

2.1.1(a) Knowledge-based methods

This method makes use of human awareness and knowledge about the typical human face geometry and facial feature arrangement, which are used to form common rules in defining the relation between facial features. However, his method has issues in finding the correct technique in converting human knowledge of face into well-defined rules and identifying face candidates based on the rules defined (Rao, Panakala, & Prasad, 2016). Also, the knowledge-based method is not well-performed under varying head poses. This method finds rules to describe the shape, size, texture, and other characteristics of facial features by taking advantage of natural face symmetry and natural top-to-bottom and left-to-right order according to the appearance of the features in the human face (Neeraj & Sugandha, 2014). Figure 2.1 shows the typical face image used in the knowledge-based method.

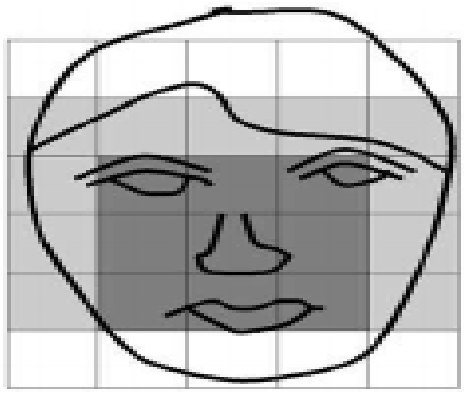

Figure 2.1 Human-coded rules based on human knowledge of the characteristics (Rao et al., 2016)

Under this method, a hierarchical approach may be used and can be divided into two levels; **higher** and **lower** level. At higher levels, a rough description of face geometry is used to find the possible face candidates while predefined rules on facial

characteristics and their arrangement are used to identify an input source region as face or non-face after facial features are extracted (Neeraj & Sugandha, 2014). Using the same hierarchy approach, a method of multi-resolution of the hierarchy of images and specific rules defined at each image is composed (Reddy, 2017). The method consists of three layers of hierarchy made by image sub-sampling.

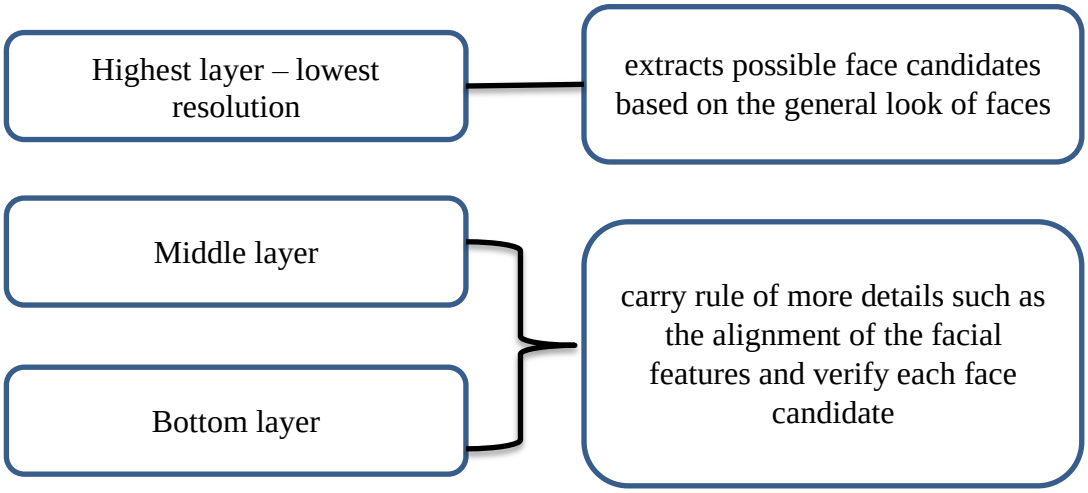

Figure 2.2 Three layers of hierarchy used by Reddy (2017)

Apart from the hierarchical approach, another method named horizontal and vertical projection has been proposed under the knowledge-based category by (Kotropoulos & Pitas, 1997). When the two projections are performed on the test image, local minimums are detected as facial feature candidates which together create a face candidate. This is based on the observations which that human eyes and mouth have lower intensity compared to other parts of faces. The detection rules such as eyebrow and nostrils are used to validate the face candidate.

2.1.1(b) Feature invariant

This method aims to find structural features despite the varied viewpoint and lighting conditions. The structural features include facial local features, textures, shape, and skin color. Besides, the local features such as eyes, mouth, and nose are extracted using multi-resolution or derivative filters, edge detectors, morphological operations, or thresholding. At last, relationships among facial features are determined and the existence of a face is verified by forming statistical models (Rao et al., 2016). Several features of faces are combined for face detection. At first, color information for skin color detection is used to extract the candidate of face regions. Different illumination conditions are handled by extracting five percent (5%) brightest pixels and used their mean color for lighting compensation. Next, invariant facial features are detected for region verification. Two detection schemes called ‘eyes map’ and ‘mouth map’ are designed based on chrominance contrast and morphological operations. Human eyes and mouth are selected as the most significant features of faces at that time. At last, a triangle is formed between the eyes and mouth, and verification is made based on luminance variations and average gradient orientations of eye and mouth blobs, geometry and orientation of the triangle, followed by the presence of a face boundary around the triangle (Jain, Hsu & Abdel-Mottaleb, 2002).

2.1.1(c) Template matching

In this method, the fairly elliptical head outline is detected using filters, edge detectors, or silhouettes. By using the same way of detection, the contours of local facial features are extracted. Then, the correlation between features extracted from the sources and predefined stored templates of the face and facial features are computed

to determine the presence of the face (Neeraj & Sugandha, 2014). Traditional techniques in template matching are mainly used for shape or boundary matching, not for texture matching and they are sensitive to scale, shape, and pose variations. Thus, deformable template methods have been proposed to cope with the variations.

2.1.1(d) Appearance-based methods

Different from the methods mentioned, the appearance-based method is not considering facial feature points but all regions of the face. From an appearance-based viewpoint, face detection can be classified as pattern classification problems with two classes which are '*face'* and '*non-face'*. In this method, it scans through the sources of given window size and analyzes each covered region. There are widely used methods to detect faces under appearance-based namely Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), Eigenfaces, Support Vector Machine (SVM), and Neural Networks.

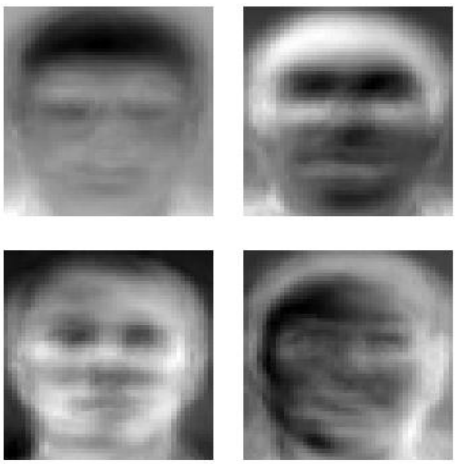

Figure 2.3 Sample of Eigenfaces

The eigenface-based method is based on the Principal Component Analysis (PCA) which reduces the dimension of data. "Eigenfaces" refers to the significant

features resulted from the projection of face image data into a feature space that covers the many variations among known facial images. However, the Eigenface is sensitive to lighting and head position while PCA is insensitive to variation of face position and facial expression (Saini et al., 2014). These weaknesses of PCA are supported by Solanki and Pittalia (2016). Besides, the learning process of the PCA method is time-consuming and it has restrictions on the size and face position (Gondhi & Kour, 2017).

2.1.1(e) Related works

Viola, Way, and Jones (2004) proposed a face detection framework that capable to process images extremely rapidly and with high detection rates. The framework used '*Integral Image*' also known as a summed-area table which is an image representation that allows the features to be computed very quickly. The input face image is converted into an integral image and this can be obtained by creating every picture element equal to the complete summation of all pixels on top of and to the left of the involved picture element (Hazim et al., 2016). Besides, the authors also proposed a classifier that was built using the AdaBoost learning algorithm (Freund & Schapire, 1995). The classifier is used to select a small number of critical visual features from a very large set of potential features. They used a method to allow background regions of the image to be quickly discarded which combining classifiers in a cascade. As a result, the authors yielded comparable performance with the best previous systems (Sung & Poggio, 1998; Rowley et al., 1998; Schneiderman & Kanade, 2000; Roth et al., 2000).

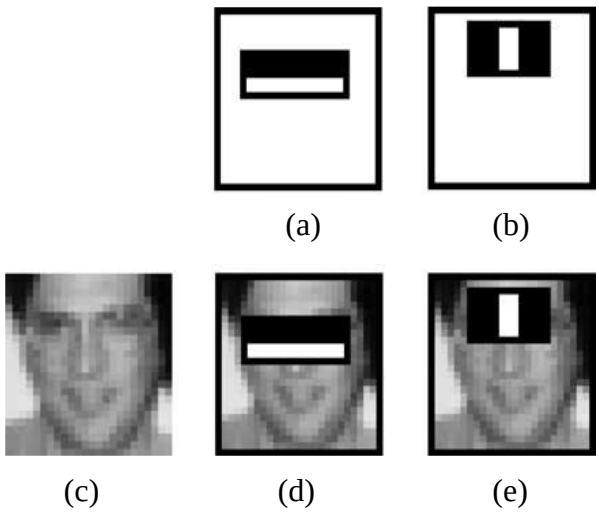

Figure 2.4 Two features ((a), (b)) selected by AdaBoost overlaid on the training image (c), as shown (d) and (e). The first feature (a) measures the difference in intensity between the region of the eyes and a region across the upper cheeks. The second feature (b) compares the intensities in the eye regions

The Haar features, integral image (i.e. summation of the pixel values of the original image), Haar feature classifier, and cascade are combined by (Dabhi & Pancholi, 2016) and the authors use cascading of the stage to make the process faster where it eliminates the unsuccessful candidate from the first stage and the process repeats until the final stage. The same algorithm also has been used by (Paharekari et al., 2017) in the automated attendance system to detect an object. Apart from using the original approach of the Viola-Jones algorithm, there is also a modified version whereas for the first step, the image is rescaled to different sizes and the size locator is running across the images. Then, a sub-window is swept across the selected image from the real-time capture to detect the faces. Compared to the Viola-Jones algorithm, the detector is rescaled rather than the images. In the end, a bounding box is drawn around the face detected depending on the size of the face in the video frame (Surekha et al., 2017). Another modification of the Viola-Jones algorithm is expanding the detection process. The face is tracked from frame to frame using a correlation tracker

from the *dlib* library. (Bhattacharya et al., 2018) stated that this approach helped in saving computational power as they do not have to detect face after transforming to the new frame and it acts as a face-log.

(Yi & Yi, 1998) proposed a facial features detection using a geometrical face model with three main steps; first is preprocessing which is to get rid of high-intensity noises and transform the image into binary, second is labeling and grouping process to generate facial feature candidate block by block and third is locating the actual face using geometrical face model. The success rate achieved by the authors was 86% in controlled conditions.

2.1.2 Spoof detection

Spoof detection is the aliveness detection of faces that have been introduced as a countermeasure against spoofing attacks. As a user-based access system, a facial recognition system should be robust against spoofing attacks as human face images can be easily obtained from smartphones and social networks. Also, spoof attacks can be categorized into three categories; print-attack, replay-attack, and 3D mask attack. The bad effect of spoof attack exposure is the system might accept the malicious user as the authenticated user and this will make the reliability of the system in terms of its security measure be questioned. According to (Kaur, 2017), face spoof detection can be categorized into four categories:

a) motion-based methods
b) texture-based methods
c) methods based on image quality analysis
d) methods based on other cues

2.1.2(a) Motion-based methods

These methods aim to capture an important cue for vitality which is the subconscious motion of organs and muscles in a live face. (Pan et al., 2007) proposed an eye-blink-based anti-spoofing method which is claimed to be a spontaneous and non-intrusive approach. The authors' aim was for photograph spoofing which falls into the print-attack category. The proposed method does not require any additional hardware except a generic web camera. The approaches taken by the authors include modeling blink detection as a reference in a Conditional Random Field Framework. With this modeling process, long-range dependencies among observations and states are enabled. Apart from that, the authors also introduced Eye closity (i.e. a discriminative measure derived from the adaptive boosting algorithm) and embedded it into the contextual model. This method is for computational efficiency and accuracy detection consideration. At the end of the experiments, the authors' proposed method has outperformed the cascaded AdaBoost and Hidden Markov Model (HMM) by just using one generic web camera under uncontrolled indoor lighting conditions, and glasses-wearing was allowed in the experiment.

(Killioğlu, Taşkiran & Kahraman, 2017) presented anti-spoofing in face recognition with liveness detection using pupil tracking. The authors extracted the eye area from the real-time camera by using the Haar-Cascade Classifier. Then, Kanade-Lucas-Tomasi (KLT) algorithm is used to get stable eye regions, and feature points are extracted and traced to minimize head movements. They have introduced an algorithm with five main steps which are; first, eye area extension from live camera vision, second, eye area tracking using KLT algorithm, third, pupil localization to find eye direction, fourth, activating random LED on the square frame, five, verifying

pupil direction. This proposed method has a success ratio of 89.7% on all databases while 94.8% on the database that excluded persons with glasses.

2.1.2(b) Texture-based methods

These methods are proposed to extract image artifacts in spoof face images. Texture features like Histogram of Oriented Gradients (HoG) are capable of differentiating artifacts in spoof faces from the success of the Idiap and CASIA databases (Kavitha & Vijaya, 2017). (Raut, Borkar & Nikam, 2018) proposed a face spoof detection system that involves the use of K-Nearest Neighbors (KNN). The proposed methods involved face detection, Gabor filters, blur detection, Local Binary Pattern (LBP), and color moments. The authors highlighted that the recaptured images have less color compared to the original images. Images from video attack plotted distributed histograms while the real images had a histogram concentrated in one area. Hence, the histogram of real images had higher peaks than the spoof images. The authors used two widely used databases for face liveness detection named NUAA Database and CASIA Database and experimented on both models. LRF-ELM produces more accurate results compared to CNN on both databases at 84.04% and 88.75% respectively.

2.1.2(c) Methods based on image quality analysis

(Li et al., 2004) presented a live face detection based on Fourier spectra to classify live faces or faked images. The authors assumed that photo has less high-frequency components that those of live face images. Unfortunately, according to the authors, their method can be defeated by a very clear and big-size photo. In the paper,

the authors stated that it was an effective approach in liveness detection by monitoring temporal changes of facial appearance over time (i.e. motion images from video sequences), where energy value defined in frequency domain represented the facial appearances. Therefore, they proposed a three-step algorithm to solve the problem; first, construct a subset by extracting an image from an input image sequence every four images, second, the energy value is computed for every image in the subset, third, the Frequency Dynamics Descriptor (FDD) which is the standard deviation from the flag values is calculated to represent temporal changes of the face. From the results, the authors made a conclusion which it is very effective to prevent small-size fake images by using a high-frequency descriptor while frequency dynamics descriptor is useful to discover spoof using large size of a fake image.

2.1.2(d) Methods based on other cues

(Kollreider, Fronthaler, & Bigun, 2008) has proposed an optical flow-based method to the input video to observe the motion of the face where the information of face motion for liveness judgment can be obtained. The authors collected evidence for the liveness of each faces by observing the 3D properties of the faces. Besides, they also observed the eye-blinking or mouth movements such that these two features are non-intrusive modes. Apart from that, they also presented a method which to ask and check for responses at random; applied to eye-blinking or utterances. In the paper, the authors stated that in non-interactive mode, eye-blinking was easier to be observed compared to mouth movements.

(Akbulut, Sengur & Ekici, 2017) presented a deep learning-based face liveness detection in videos. The authors used Local Receptive Fields (LRF)-ELM and Convolutional Neural Network (CNN). In this proposed method, the LRF-ELM model