FACE RECOGNITION USING HYBRID GABOR FILTER AND STACKED SPARSE AUTO ENCODERS (SSAE) WITH DEEP NEURAL NETWORK MODEL

ABDULLAH GHANIM JABER

PROJECT SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF COMPUTER SCINCE (SOFTWARE TECHNOLOGY)

FACULTY OF INFORMATION SCIENCES AND TECHNOLOGY
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2021

PENGECAMAN WAJAH MENGGUNAKAN HIBRID PENAPIS GABOR DAN ENCODER AUTO TERSUSUN JARANG DENGAN MODEL RANGKAIAN NEURAL DALAM

ABDULLAH GHANIM JABER

PROJEK YANG DIKEMUKAKAN UNTUK MEMENUHI SEBAHAGIAN
DARIPADA SYARAT MEMPEROLEH IJAZAH
SARJANA SAINS KOMPUTER (TEKNOLOGI PERISIAN)

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2021

DECLARATION

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged.

30 June 2021

ABDULLAH GHANIM JABER
P 98659

ACKNOWLEDGMENT

I am indebted to Allah for always helping me to finish what I start, even harder than I thought, and for helping me to carry out this work. The completion of this thesis marks the end of a long and arduous process in my academic life.

I would like to give my great gratitude and appreciation to my supervisor, Assoc. Prof. Dr. Ravie Chandren Muniyandi, for his guidance and encouragement throughout this research work a grand success.

I would like to thank all other faculty staff and students at the computer science department, faculty of information sciences and technology, university Kebangsaan Malaysia, and all my friends from Algeria, Yamane, Iraq, Libya, Jordan, Malaysia. Moreover, I thank Olivetti Research Laboratory for their valuable ORL face image database that I have used in my thesis. Moreover, I thank Extended Yale-B for their valuable face image database that I have used in my thesis.

Finally, I heartily thank the members of my family whose unwavering support, care, and love have enabled me to complete this thesis. I am indebted to my mother and my family, who installed in me the belief that I could succeed in my journey.

ABSTRACT

Face recognition has been getting greater attention especially since most recognition systems are easily penetrated or copied. However, the accuracy of face recognition is very important in determining the success in identifying an individual. One of the hurdles that hampered the accuracy of face recognition is when the dataset is limited. Especially in traditional machine learning algorithms that are used for face recognition based on images requires sufficient training data to enable high levels of face recognition accuracy. Based on the afore-mentioned problem, a study was carried out to enhance face recognition technique by using hybrid Gabor Filter and deep learning Stacked Sparse Auto Encoders (SSAE). The experimental evaluation was carried out using two datasets which are Olivetti Research Laboratory OLR and Extended Yale-B databases. All face images are greyscale, and the resolution 92×112 for OLR database while 192×168 resolution for Extended Yale-B database. The result of the evaluation showed that the accuracy of face recognition has been improved by using proposed method and get best accuracy among all experimented State-of-the-Art methods on both of OLR and Extended Yale-B databases.

ABSTRAK

Pengecaman wajah telah mendapat perhatian lebih besar terutama kerana kebanyakan sistem pengecaman mudah ditembusi atau disalin. Namun, ketepatan pengecaman wajah sangat penting dalam menentukan kejayaan mengenal pasti seseorang individu. Salah satu rintangan yang menghalang ketepatan pengecaman wajah adalah apabila set data terhad. Terutama dalam algoritma pembelajaran mesin tradisional yang digunakan untuk pengecaman wajah berdasarkan gambar memerlukan data latihan yang mencukupi untuk membolehkan tahap ketepatan pengecaman wajah yang tinggi. Berdasarkan masalah yang disebutkan di atas, sebuah kajian dilakukan untuk meningkatkan teknik pengecaman wajah dengan menggunakan Gabor Filter hibrid dan pembelajaran mendalam Stacked Sparse Auto Encoders (SSAE). Penilaian eksperimental dilakukan dengan menggunakan dua set data iaitu Olivetti Research Laboratory OLR dan Extended Yale-B database. Semua gambar muka adalah skala kelabu, dan resolusi 92 × 112 untuk pangkalan data OLR sementara resolusi 192 × 168 untuk pangkalan data Extended Yale-B. Hasil penilaian menunjukkan bahawa ketepatan pengecaman wajah telah ditingkatkan dengan menggunakan kaedah yang dicadangkan dan mendapatkan ketepatan terbaik di antara semua kaedah Canggih yang diuji pada kedua-dua pangkalan data OLR dan Extended Yale-B.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

1.1 INTRODUCTION

Face recognition is a task so familiar to humans that the individual does not even notice the extensive number of times it is performed every day. Although research in automated face recognition has been conducted since the 1960s, it has only recently caught the scientific community's attention. Many face analysis and face modeling techniques have progressed significantly in the last decade (He et al., 2005). However, the reliability of face recognition schemes still poses a significant challenge to the scientific community.

Face recognition has been an active area of research in the past several decades. Initially, a branch of artificial intelligence to enable robots with visual perception is now part of a more general and more extensive discipline of computer vision. Computer vision applications can process images from a wide range of the electromagnetic spectrum (Alzubaidi & Kalita 2016).

Falsification of identity cards or intrusion of physical and virtual areas by cracking alphanumerical passwords frequently appear in the media. These problems of modern society have triggered a real necessity for reliable, user-friendly, and widely acceptable control mechanisms for the identification and verification of the individual (Heindl 1922).

In computer vision, face recognition applications are confined to the narrow band of visible light where surveillance and biometrics authentication can be performed. Biometrics, which is based on authentication on the intrinsic aspects of a specific human being, appears as a viable alternative to more traditional approaches (such as PIN codes or passwords).

In addition, biometrics is the term used to describe human characteristics metrics such as iris, fingerprint, or hand geometry. These metrics are used to identify and access control of individuals under surveillance (Alzubaidi & Kalita 2016). Among the oldest biometric techniques is fingerprint recognition. This technique was used in China as early as 700 AD for official certification of contracts. Later on, in the middle of the 19th century, it was used to identify persons in Europe (Heindl 1922). A currently developed biometric technique is iris recognition (J. Daugman 2002). This technique is now used instead of passport identification for frequent flyers in some airports in United Kingdom, Canada, and the Netherlands. As well as for access control of employees to restricted areas in Canadian airports and the New York JFK airport. These techniques are inconvenient due to the necessity of interaction with the individual to be identified or authenticated.

The face is becoming the preferred metric over current biometrics simply because it is a natural assertion of identity, and its non-intrusive nature provides more convenience and ease of verification. For example, in a fingerprinting system, the subject is required to interact with the system by placing a finger under a fingerprint reader, and an expert must verify the results. In contrast, using the subject’s face as a metric requires no intervention, and a non-expert can verify the results. This is one of the reasons why this technique has caught an increased interest from the scientific community in the recent decade.

Facial recognition holds several advantages over other biometric techniques. It is natural, non-intrusive, and easy to use. In a study considering the compatibility of six biometric techniques (face, finger, hand, voice, eye, signature) with machine-readable travel documents (MRTD) (Heitmeyer 2000), facial features scored the highest percentage of compatibility, see Figure 1.1. In this study, parameters like enrollment, renewal, machine requirements, and public perception were considered. However, facial features should not be considered the most reliable biometric.

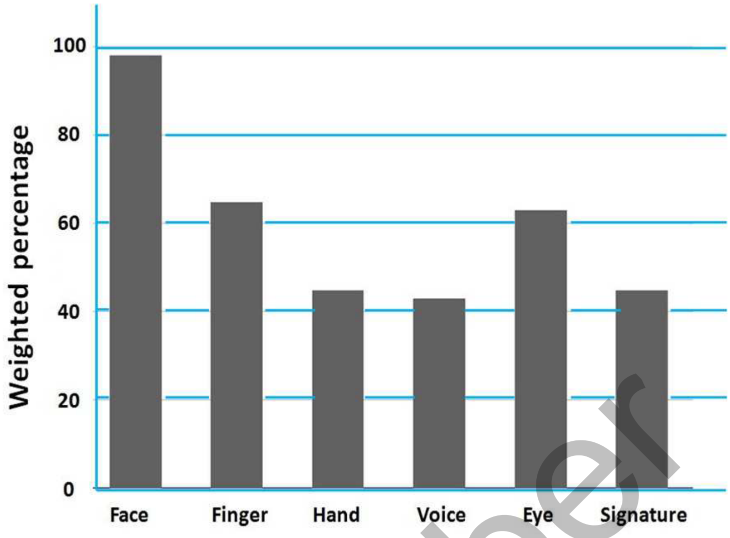


Figure 1. 1 Comparison of machine-readable travel documents (MRTD) compatibility with six biometric techniques; face, finger, hand, voice, eye, signature Courtesy of (Heitmeyer 2000).

However, automated facial recognition can be used in many areas other than security-oriented applications (access-control/verification systems, surveillance systems), such as computer entertainment and customized computer-human interaction. Customized computer-human interaction applications will in the near future be found in products such as cars, aids for disabled people, buildings, etc. The interest in automated facial recognition and the number of applications will most likely increase even more in the future. This could be due to increased penetration of technologies, such as digital cameras and the internet, and due to a greater demand for different security schemes.

1.2 FACE RECOGNITION SYSTEM

Face recognition is the process of labeling a face as recognized or unrecognized. The process has a life cycle based on a pipeline that goes through detection, extracted feature, and a recognition stage. A computer-based face recognition system consists of two phases: the enrollment and identification/verification phases shown in Figure 1.2.

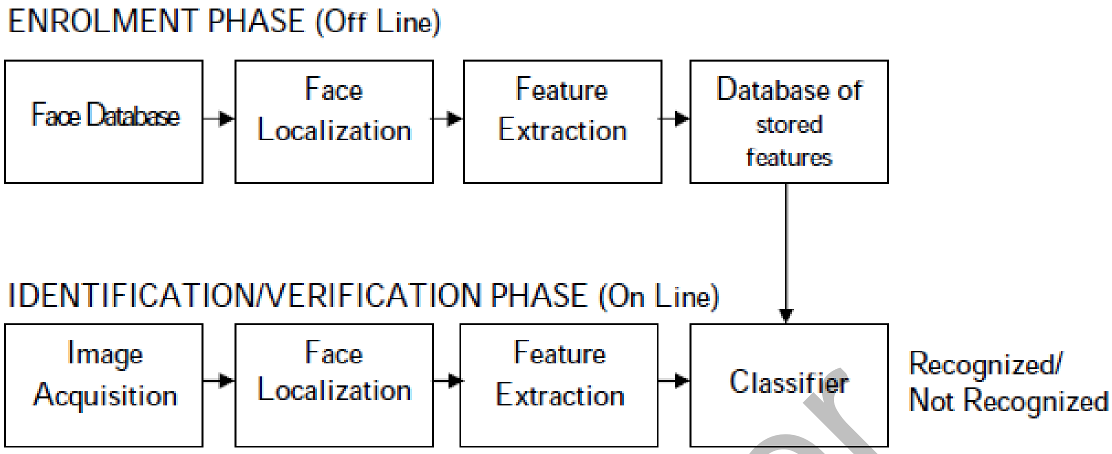


Figure 1. 2 A typical Face Recognition System

During the enrolment phase, the enrolled face images present in the face database are pre-processed and localized/detected. Then feature extraction techniques are applied to transform pixel images into facial features, as performing face recognition directly using raw images is an inefficient strategy. The extracted features are then used for analysis and recognition. In the identification/verification phase, the system captures a probe (test) image, pre-processes it, localizes the face region, and extracts the face features. The next stage takes the probe image features and compares them with a database of known people (previously enrolled), searching for the closest matching images. Next, the test image and the matched image are compared for similarity to give the decision as recognized or not.

1.2.1 Face Detection

The preliminary task of face recognition is to localize the face region and to condition the face image to normalize the effect of noise, illumination, scale, etc. Face localization aims at determining the position of a single face in an image. This is a simplified detection problem with the assumption that the input image contains only one face image. A facial image has to be pre-processed to geometrically normalize the face part in the image. The recognition system needs prior knowledge about the face to segment the image from the background.

1.2.2 Feature extraction

The major task in face recognition is to extract significant features capable of representing the face information from the image. It is known that humans use both low and high-level features such as shape, contextual information, colour, and mannerism to recognize a face.

Even though several feature extraction techniques are proposed in the literature to capture the low and high-level features, the current feature extraction methods find it difficult to extract the required features because of the wide variability of facial appearance due to the following factors

1. Presence or absence of structural components: Facial features such as beards, moustaches, and glasses may or may not be present, and there is a great deal of variability among these components, including shape, colour, and size. The features of the face change because of spectacles, as shown in Figure 1.3.

Figure 1. 3 Effect of structural components

2. Facial expression: The appearances of faces are directly affected by a person's facial expression. As shown in Figure 1.4, the interrelationship of the features gets distorted because of a change in facial expression.

Figure 1. 4 Face image with varying facial expressions

3. Occlusion: Faces may be partially occluded by other objects. Some faces may partially occlude other faces in an image with a group of people, which gives variation in the extracted features.
4. Image orientation: Face images directly vary for different rotations of the camera's optical axis.
5. Imaging conditions: When the image is formed, factors such as lighting (spectra, source distribution, and intensity) and camera characteristics (sensor response and lenses) affect the appearance of a face. In Figure 1.5, the same person with the same facial expression and pose looks different because of a change in illumination due to lighting conditions.

Figure 1. 5 Effect of illumination.

6. Pose: The images of a face vary due to the relative camera-face pose (frontal, 45 degree, profile, upside down), and some facial features change due to the partial or whole occlusion of the image. Figure 1.6 shows the variation in the same face image because of the pose.

Figure 1. 6 Face image with different poses.

1.2.3 Classification model

The next step in the face recognition system is to capture the distribution of the facial features using classification models. In general, classification models assume enough discriminatory information in the feature vectors and employ such information to discriminate the faces.

The performance of the face recognition system mainly depends on the effectiveness of the classification model. Parametric and nonparametric approaches are employed in the literature for classification problems. The difficulty in the parametric model is the requirement of assumptions regarding the distribution of features. The non-parametric model makes minimal assumptions regarding the distribution. However, models such as the nearest neighbour classifier give poor generalization, require large memory to store all the feature vectors, and time consumption is also very high.

The final stage is based on the information from the classification model to decide whether the person is a recognized one or not.

1.3 PROBLEM STATEMENT AND MOTIVATION

The field of face recognition has been a topic of study since the 1960s. It has stayed relevant both due to the practical importance of the topic and the theoretical interest from cognitive scientists. Face recognition aims to verify or identify an individual's identity using their face, either from a single image or from a video stream. It is used to accurately track patient medication consumption and support pain management procedures. Face recognition systems are used in many contexts, such as within security and healthcare.

Deep Learning models are commonly used in computer vision and significantly improves state of the art in many applications. As a result, researchers have recently given this field increased attention, conducting a multitude of studies and continuously improving the models that already exist. One of the most important ingredients is the availability of large quantities of training data (O'Mahony et al., 2019).
Building a face recognizer can be challenging, especially when the dataset is limited. When the dataset is limited, one of the major challenges is that an individual's face may look different if various lightning and those different persons may have similar-looking faces (Wójcik et al. 2016).

In addition to the above, it is well known that traditional machine learning algorithms used for face recognition based on images require sufficient training data to enable high levels of face recognition accuracy. To support the development of such algorithms, a large number of human face databases have been made available in the public domain. Such databases have thousands of images that can be used for training learning algorithms and subsequent testing. With the very recent advances of deep learning technologies that are driving the paradigm shift of learning technology from the traditional to the deep learning systems, one needs very large databases to train the systems to obtain the full potential improvements that deep learning algorithms can provide (Baralis et al. 2008; Kuncheva & Rodríguez 2007; Panigrahi et al. 2021; Yin et al. 2006).

1.4 OBJECTIVE OF RESEARCH

The research aims to develop a robust system for recognizing the face with high accuracy. Therefore, a hybrid model was proposed to satisfy this objective to improve face recognition using Gabor Filter and Stacked Sparse Auto Encoders (SSAE). This aim will be satisfied through the following objectives:

1. To improve the face recognition method based on the Gabor Filter (GF) naive combination and Stacked Sparse Auto Encoders (SSAE) deep learning.

2. To evaluate and compare the proposed hybrid face recognition model with state-of-the-art methods.

1.5 SCOPE OF THE RESEARCH

1. The novel hybrid recognition based on the Gabor filter (GF) of features method and Stacked Sparse Auto Encoders (SSAE) Model can be applied in the face research area.
2. Different datasets proposed for method hybrid recognition method have been tested on ORL, Yale-B datasets.
3. The hybrid recognition method is programmed by MATLAB R2020a based on windows 10.

1.6 RESEARCH METHODOLOGY

The research methodology consists of the following phases: theoretical, framework, and implementation.

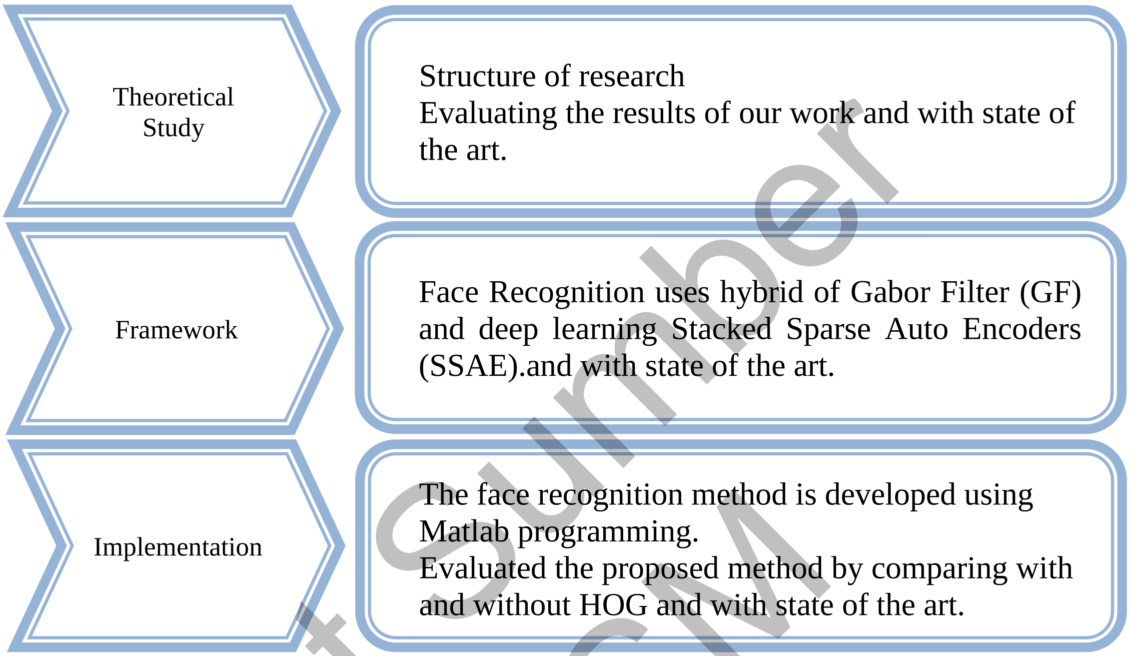


Figure 1. 7 Phases of research methodology

After that, the theoretical phase defines the problem statement and presents the structure of this research and the evaluation method for the proposed method. The framework stage consists of improving face recognition. The improvement of face recognition based on hybrid Gabor Filter (GF) and deep learning Stacked Sparse Auto Encoders (SSAE) is used. The evaluation of the proposed method is done by comparing the proposed method with and without using GF and comparing it with state-of-the-art.

1.7 THESIS ORGANIZATION

The layout of this thesis is as follows:

- Chapter two discussed the literature study of face recognition.
- Chapter three explains the proposed algorithms Hybrid Gabor Filter (GF) with Deep learning Stacked Sparse Auto Encoders (SSAE) Model.

- Chapter four discussed the results of the systems.
- Chapter five includes conclusions and future works.

CHAPTER II

LITERATURE REVIEW

2.1 INTRODUCTION

Human organs, with or without an attribute, can be used as variables of authentication or verification purposes. Technology has eased these processes by providing many ways to validate a person's identity using one or more human organs. This area of technology is what we know as “biometrics.” Biometric techniques refer to validating an individual's true identity by their behavioral patterns or using one’s physical characteristics. These techniques include fingerprint, keystroke, iris, palm print, DNA, voice, face, and handwriting styles. Using such biological traits in biometric systems to verify a person makes it difficult to alter or forge. Biometric methods can be classified into physiological (fingerprint, face, DNA) and behavioral methods (handwriting, voice, keystroke). The physiological methods are known to be more reliable than the latter because physiological traits are often unalterable except for acute injuries. The behavioral characteristics, on the other hand, may vary due to anxiety, fatigue or illness but also has the advantage of being less intrusive.

Recognition of face is a traditional challenge issue with a long path of the study. About since 1850, bureaucratic affairs and the criminal identity have been employed of the faces as codes. This contains the united police security crime scene photos suggested via Alphonse Bertillon 17 as well as the first passports photographic released after World War I 18 through this period, the lack of successful indexing was noticed where too many images were taken and recovered.

The study on facial recognition started in 1960. Woody Bledsoe, Charles Bisson, and Helen Chan Wolf have adopted the system device. This system demands the administrator to determine the pictures' nose, ears, mouth, and eyes. The proportion and distance among the features existing and the joint reference dots are then computed and compared (Petrescu 2019; Tikoo & Malik 2017).

In 1970 the experiment's work was further improved via Goldstein, Lesk, and Harmon by utilizing other characteristics like lip width and hair color for an automated recognition process. The first suggested the notion of (PCA) component analysis by Kirby and Sirovich in 1988 to dissolve the issue of face recognition. Until today several face-recognition experiments research has been performed continuously (Chin 2018).

In recent years, the utilization of facial recognition approaches has evolved increasingly in security, biometrics, defense, law enforcement, and other merchant sectors (Andrejevic & Selwyn 2020).

Moreover, summarized review literature on research areas linked to the approaches in this thesis is given in this chapter. To launch with, it presents the basics of the standard methods for recognition of a face. Accordingly, Section 2.2 reviews Face Recognition Methods, while Generic Face Recognition in Section 2.3, Section 2.4 reviews Artificial Neural Network (ANN) and Deep Neural Network for Face Recognition. Section 2.5 summarizes this chapter.

2.2 FACE RECOGNITION METHODS

There are four categories of face recognition methods (Choi et al., 2012; Dubey & Tomar, 2016; Yuille et al., 1992).

1 Knowledge-based approaches.
2 Features invariant approaches.
3 Template matching approaches.
4 Appearance-based approaches.

2.2.1 Knowledge-Based Approaches

They are known as rule-based methods. They mainly depend on a set of rules in the detection process (Rowley et al., 1998). This method considers our knowledge of face images and translates them into a set of regulations or rules. An example of a rule is a face having two eyes, a nose, a mouth, or a face having the eye area darker than the cheeks. These features are within specific distances in relation to each other.

There is a limitation to this method which is the ability to build an appropriate set of rules because very general rules might result in false positives, and too detailed a set of rules might result in false negatives. This actually could be resolved by using hierarchical knowledge-based methods, which are efficient with simple inputs. This method is generally limited since it cannot locate many faces in a complex image.

(J. Wang & Tan 2000) presented a system which was for images with uncomplicated backgrounds. The system's approach was to find the knowledge surrounding the geometry of face images and the design of the various features. They set rules that defined how to identify the face from the complicated background. These regulations provide an estimate of the geometry which is later used for recognition.

This method does not work effectively under varying positions or orientations. There is also the need for a method that can define human facial framework into clearly defined and meaningful regulations.

2.2.2 Feature Invariant Method

Method of Feature invariant identifies faces via used structural face characteristics to extracting features (Kjeldsen & Kender 1996; Leung et al. 1998; Yow & Cipolla 1996). This idea was developed to overcome the limitations of our instinctive knowledge of face images.

One of the earliest algorithms developed was by (Han et al. 1997). Normally, a statistical classifier, models, or edge detector is trained and then utilized to distinguish between non-face and face sections.

The method seeks to find distinct characteristics of an image of a face despite the position or angle. Its focus is to locate systemic characteristics like the fiducial points, the skin texture, and the colour of a face even though there are changes in the head pose, lightning variations, and viewpoint. Facial recognition uses different facial characteristics such as the mouth, cheekbones, the contour of the eye socket, nose, zone near the cheekbones, and the eyes.

Research has shown that the colour of the skin is reviewed to be one of the significant characteristics for face recognition since every person has a unique skin colour and recognition is explicit when the race is a criterion for detection (Sharifara et al. 2014). Feature invariant has a challenge when the characteristics of a face image are altered by noise, occlusion, or illusion. It is also demanding if there is a need for feature extraction.

2.2.3 Template Matching Approaches

Template matching methods use parameterized face templates to locate and detect faces. It compares test images with the template of images stored for detection. It can be computed using the relationship between the various characteristics obtained from input image and the predetermined estimate. Each feature could be defined independently, for instance, the hairline can be distinguished using filters or edge detectors.

This technique has limitations if face images are anterior or differences in scale, shape, and pose. Nevertheless, (Dubey & Tomar 2016) proposed the use of deformable templates to solve these limitations.

However, according to (Brunelli & Poggio 1993), template matching can be likened to a test face image which constitutes a two-dimensional array and can be compared with the use of acceptable measures like the Euclidean distance with one framework showing the whole face image.

The human face can also be represented by more than one framework. (Brunelli & Poggio 1993) chose a group of four feature templates: the mouth, eyes, nose, and the whole face, of 188 images of 47 individuals. They compared it to a geometrical matching algorithm, and it came out superior with a recognition rate of 100%. One limitation of the template matching approach is that it is computationally complex.

2.2.4 Appearance-Based Method

Appearance-based methods depend on delegated trained images to discover facial models (Osuna et al. 1997; Rowley et al. 1998; Viola & Jones 2001). It captures the representative variability of faces. In effect, appearance-based techniques have exhibited greater performance compared to other techniques (C. Zhang & Zhang 2010).

Generally, these methods depend on skills from machine learning and analysing statistically to search for the significant features of facial images. The method explains face recognition as an issue of two class image classification, that is the face or non-face class.

There are different recognition techniques that are based on appearance-based methods. They include eigen faces, LDA, SVM, PCA, etc. Whatever is considered face is contained in the face class, and all non-face characteristics are also grouped into the non-face. This method requires a large database and a high-quality image for the detection process.

2.3 GENERIC FACE RECOGNITION

As stated in Chapter I, four fundamental elements are included in a generic face recognition system, i.e., face alignment, detection of face, classification, and feature extraction. Representation of feature and recognition classifier plays a crucial role in face recognition and thus attracts much interest.

2.3.1 Extraction of Feature

In several applications, for face images, representation of the pixel raw is adopted, i.e., images of the face are represented as a 2-D composed matrix of intensity pixel values. There is much redundancy in the representation of the raw pixel, and it is susceptible to differences in modalities like occlusion, illumination, pose, etc. The raw image face is, therefore, typically pre-processed by different methods for the extraction of features.

A very significant aspect for systems of face recognition is feature extraction. The goal of this step is to acquire representations of faces that are appropriate for classification. Therefore, it is expected that the extracted characteristics emphasize the distinction among classes and minimize irrelevant variations inside each denomination. The optimal extraction of features would lead to a slight classification mission. This concept optimal is hard to achieve because many variants can exist together in images of the face.

This phase, however, describes a face with a collection of features characteristics called a "signature" vector that defines features prominent features of the face image with their geometry distribution, like eyes, nose, and mouth (Kortli et al., 2018; Smach et al. 2007).

Each face is distinguished by its shape, size, and structure, allowing it to be recognized. Many methods include extracting the outline of the nose, eyes, or mouth shape and employ the size and distance to define the face (Napoléon & Alfalou 2017). (Q. Wang et al. 2019) HOG, (Turk & Pentland 1991) Eigenface, analysis of linear discriminant

(LDA), analysis of independent component (ICA) (Annalakshmi et al. 2019; Seo & Milanfar 2011), scale-invariant feature transform (SIFT) (Vinay et al. 2015), quantization of local phase (LPQ), Gabor filter (Hussain et al. 2012), Haar-wavelets, Fourier transforms (Smach et al. 2007), and pattern of local binary (LBP) (HajiRassouliha et al. 2013; Napoléon & Alfalou 2017) methods are vastly utilized to extract the characteristics of the face.

A perfect exemplification of feature should be sturdy for both local and holistic variations while maintaining identity-preserving data with physical memory minimal.

In universal, current characteristics can usually be defined as representation local and representation holistic. Figure 2.1 below shows 1) local Features and (2) holistic (subspace) Features.

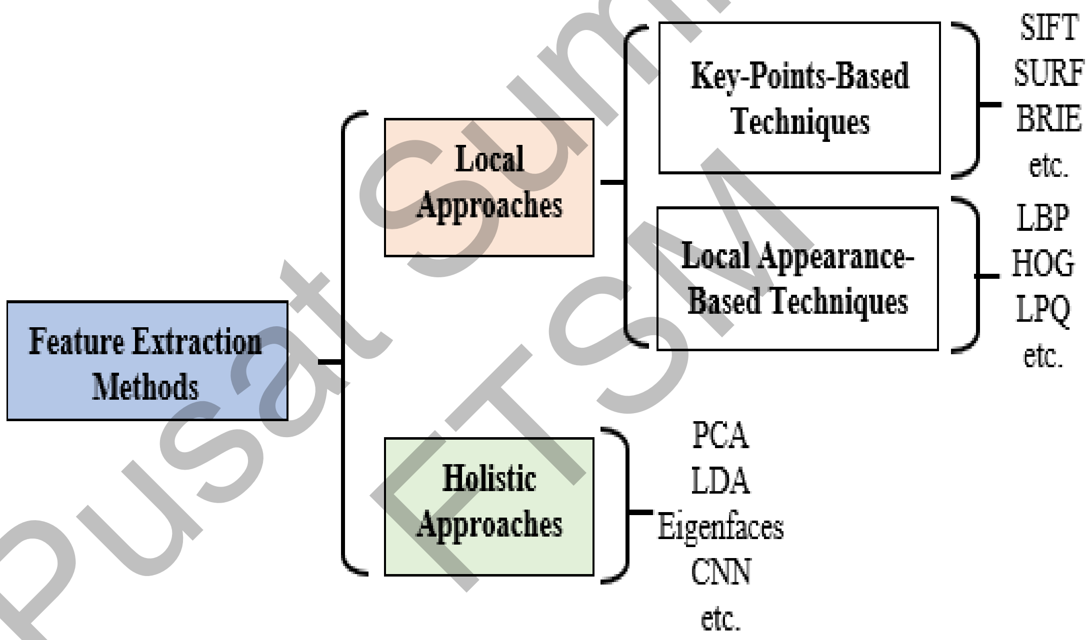


Figure 2. 1 Feature Extraction Methods. Features transform of scale-invariant (SURF); independent elementary of binary robust (BRIE); analysis of linear discriminant (LDA); histogram of gradients of oriented (HOG); feature transform of scale-invariant (SIFT); analysis of principal component (PCA); quantization of local phase (LPQ); pattern of local binary (LBP); neural network convolutional (CNN).

The first method is labeled according to specific facial features, not taking into account the entire face. While the second method uses the face complete as input information and then award projects into a tiny subspace or plane of correlation.

2.3.1.1 Local Approaches

The local approach is extracting the features from zones of face-partial and evinces large soundness to variations local. Local methods address only specific facial characteristics in the scope of recognition of the face.

They are extra sensitized to poses, occlusions, and facial expressions (Liao et al., 2012). The primary aim of these techniques is to find out unique characteristics. These methods can usually be split into two groups: (A) methods-based-local appearance are utilized to debrief local characteristics. However, the face picture is split into tiny zones (patches) (Kortli et al., 2018; Napoléon & Alfalou 2017). (B) Key-points techniques-based are utilized to identify points of attraction in the picture of the face. After that, the characteristics located on the dots are extracted.

A. Methods-based-local appearance

It is a geometrical method, often referred to as an analytical technique or feature. In order to produce more information, methods-based-local appearance concentrates on key points of the face, like the eyes, mouth, and nose. In this situation, a group of vectors characteristic with small dimensions or tiny zones represents the face image (patches).

In order to define and utilize a limited set of parameters, it also considers face privacy as a natural shape. Furthermore, these methods define the local characteristics via histograms, orientations of pixel (Ouerhani et al. 2017; Rettkowski et al. 2017), geometric properties, and correlation planes (Benarab et al. 2015; Napoléon & Alfalou 2017; Q. Wang et al. 2019).

- Domestic extraction of the feature was first developed by (Ojala et al. 1996) and later enhanced by (Ahonen et al. 2006), where the authors used LBP for recognition of facial. By comparing the relationship of the value color variation between the pixels adjacent in the circular image symmetry, the LBP approach has a goal to alteration the grayscale appearance into a vector that contains digital codes. Therefore, if the parameters external do not change the positive & negative relations among the pixel predominant and the pixels neighboring, the appearance of the LBP example is enhanced. So, in link pixel, the LBP gauge of the pixel image can be calculated via equation (2.1)

$$\mathrm{LBP}_{P,R} = \sum_{t=0}^{p-1} s(g_t - g_c)2^t \quad (2.1)$$

where g_c and g_t state importance of the gray of focus pixel and P neighboring pixels in image district of domain R and s associates to threshold equation like defined by (2.2)

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

Every pixel is pre-arranged in the graph into an LBP 8-bit format, as showed in equation (2.2). In addition, a possibility assessment is used in fuzzy-LBP as a replacement for hard coding the format dissimilarity to characterize the prospect of a format dissimilarity to be a symbol like "1" or "0," such as a linear piecewise function of fuzzy membership and a function of Gaussian-like membership as specified in the format difference (N. Tan et al. 2009). A small picture dissimilarity will slightly change the histogram of the fuzzy-LBP as equated to the original histogram of LBP via infusing the fuzzy group into LBP as FLBP. However, membership is a type of pixel change, the degree of which can be readily changed via noise. Fuzzy-LBP is also slightly sensitive for noise.

- LBP has been commonly used in several problems of face recognition due to its low cost and reliability of computation (Chen et al. 2013; H. Li et al. 2013; Xiangyu Zhu et al. 2015). LBP aims to encode the information of local contrast into texture histogram patterns. In general, LBP contrasts each local patch's central pixel with its neighboring or adjacent pixels. For each comparison, it assigns a binary mark. The binary sequence is then converted to a digit decimal for every pixel, which is then utilized as a bin for the definitive histogram. Figure 2.2 shows the corresponding LBP process of extraction.
- Have been suggested several variants for further enhancement due to the popularity of LBP. LBP based Multi-scale Block (Liao et al. 2007), Texture Dynamic (Zhao & Pietikainen 2007), Pattern of Local Binary Gabor (W. Zhang et al. 2005), a feature of Locally Binary Assembled Haar (Yan et al. 2008), and such are exemplary examples.

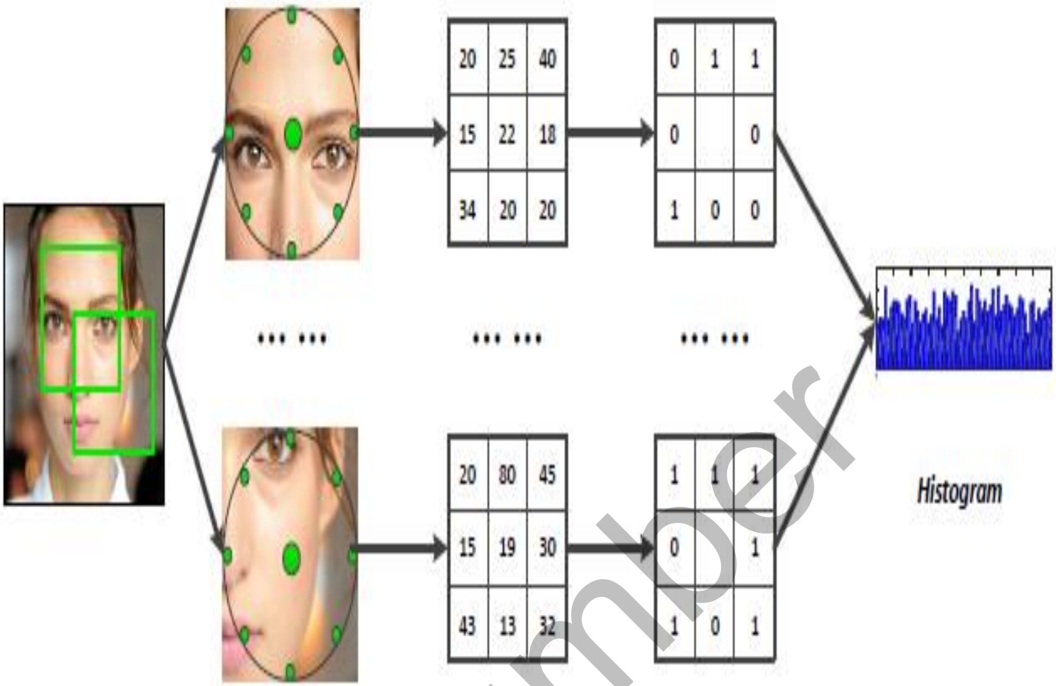


Figure 2. 2 Example of the LBP process of feature extraction. (Yan et al. 2008)

First, the picture of the face on the left is broken down into several sub-blocks. Every sub-block will be checked via pixel to pixel. Then the definition of the texture, shown as the block 3x3, is calculated by comparing the centric pixel with its circumferences. A histogram is determined to be the definitive feature after processed each of the sub-blocks.

- Histogram of oriented gradients (HOG) (Karaaba et al. 2015): For edge and shape classification, the HOG is one of the top descriptors utilized. Utilizing gradient intensity of light or the edge direction distribution the HOG method may define the face form. The method of this approach is carried out by sharing the entire face picture into cells (tiny area or zone); each cell produces direction gradients or direction of pixel edge histogram; and eventually, the entire cells histograms are merged to extract features of the face picture.
- The HOG descriptor calculates the vector of feature like follows (HajiRassouliha et al. 2013; C. Huang & Huang 2017; Ouerhani et al. 2017; Rettkowski et al. 2017): Segment the local picture into zones called cells firstly and then measure for every cell the capacity in both directions of the vertical and horizontal of the gradients first-order. A mask of 1D [-1 0 1] is the most common form to apply, as shown in equations (2.3) and (2.4).

$$G_x(x,y) = I(x+1,y) - I(x-1,y) \tag{2.3}$$

$$G_y(x,y) = I(x,y+1) - I(x,y-1) \tag{2.4}$$

where $I(x, y)$ are the value pixel of the dot (x, y) and G_y (x, y) and G_x (x, y) signify the magnitude of gradient vertical and the magnitude of gradient horizontal, respectively. The amplitude of the orientation and the gradient of every pixel (x, y) is calculated like follows in equations (2.5) and (2.6):

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{2.5}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \tag{2.6}$$

A group of various histograms of directed gradients (HOG) was suggested (Karaaba et al., 2015). With interpolation of tri-linear, the amplitude of the direction and the gradient of every pixel in the cell are measured in nine bins. Depending on the gradients of the direction, every cell pixel is created in histograms. Eventually, the entire cells histograms are merged to extract the face picture feature to execute a robust face recognition scheme. This approach is called "multi-HOG".

For identification, the authors establish a distances vector among the reference and the target face pictures. (Arigbabu et al. 2017) a new method of face recognition depends on the descriptor of pyramid-histogram of gradient (PHOG), and the filter Laplacian was suggested. Furthermore, a support vector machine (SVM) support vector machine is utilized with various kernel functions to examine the trouble of face recognition.

B. Key-Points-Based Approaches

According to the face surface for some geometric information (e.g., the distance among the width of the head and the eyes), the key-points-based approaches are utilized to detect particular geometric features. These approaches can be described via significant of two stages, the extraction of features and the detection of key dots (Calonder et al. 2011; Kortli et al. 2020; Napoléon & Alfalou 2017; Xiangxin Zhu et al. 2007).

The first phase concentrates on the efficiency of the key-dot features of the facial image detectors. The second phase concentrates on the exemplification of the data transported with the face image's key-dot characteristics. Since these approaches can address the occlusions and damaged bits, (BRIEF) robust features of binary independent elementary, (SURF) robust features of speeded-up, and (SIFT) feature transform of scale invariant approaches are commonly utilized to characterize the face picture feature

- Feature of Shift-Invariant Transform (SIFT). SIFT (Lowe 1999) is also a widely utilized feature local for recognizing face (Chen et al. 2013; H. Li et al. 2013; Simonyan et al. 2013). SIFT's merit lies in its translation, invariance to scaling, and rotation of the image. The contrast among a pixel and its adjacent ones is also considered by SIFT as the main factor in the exemplification of a face picture. Standard SIFT, unlike LBP, first locates the main dots, i.e., the corner dots of high-contrast, by defining the maximum through space and scale. In its 16x16 adjacent area, the descriptor for every main dot is then extracted via computing gradients amplitude and the direction. Furthermore, the area of adjacent is split into 16 of size (4x4) sub-groups and also calculated the 8 bin histogram of orientation in every of that. The dimension of the definitive histogram, therefore, is 128. In (Hu et al. 2014; Simonyan et al. 2013), they also employed SIFT descriptors and applied variants to exemplify the picture of the face, such as SIFT Dense.
- Speeded-up robust features (SURF) (Du et al. 2009; Işık 2014): the SURF approach is inspired via SIFT. However, it utilizes wavelets and a Hessian determinant approximation to realize better results (Du et al., 2009). SURF is a descriptor and detector which, relative to the descriptor of SIFT, appears to achieve even better, or the same, outcomes in terms of differentiation, repeatability, and robustness. SURF's primary benefit is the enforcement time, which is shorter than that utilized via the descriptor of SIFT. In addition, the descriptor of SIFT is more suited to characterize faces influenced via scaling, lighting conditions, rotation, and translation (Işık 2014). SURF aims to get the maximum approximation of the matrix Hessian to detect feature dots utilizing integral pictures to significantly decrease computational processing time. An epitome for a descriptor SURF for recognition of face utilizing datasets of 'AR face' is displayed in figure 2.3. (Mahier et al., 2011).

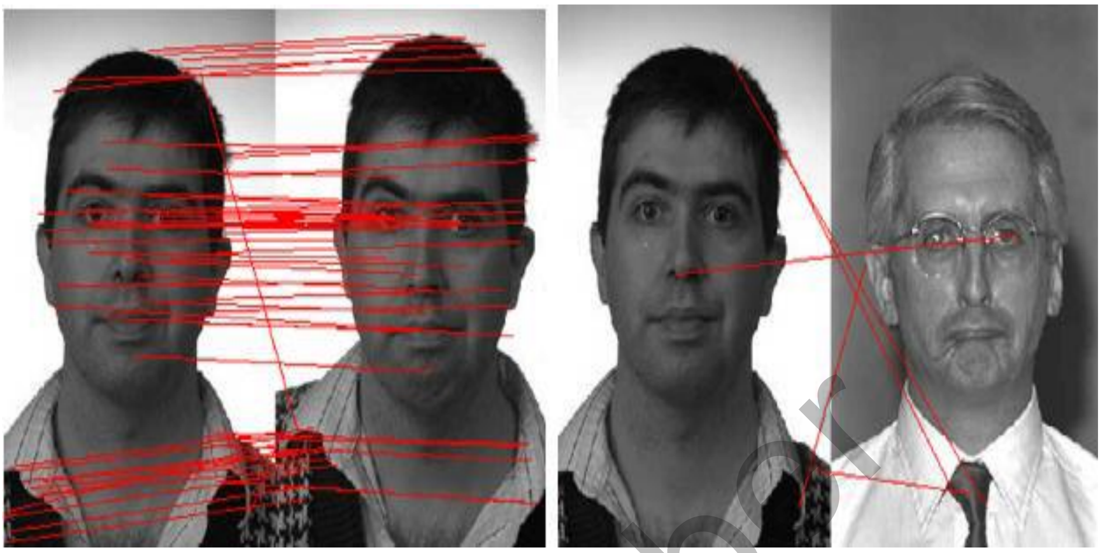


Figure 2. 3 Recognition of face depends on the descriptor of Speeded-up robust features (SURF) (Mahier et al., 2011).

- Robust independent elementary of binary features (BRIEF) (Calonder et al. 2011; Işık 2014): It is a descriptor of binary that can be computed easily and rapidly. In the idiom of appraisal, this descriptor depends on the variations among the pixel strength, which are close to the family descriptors of binary, like key-point of the fast retina (FREAK), and robust invariant scalable of binary (BRISK). The descriptor of BRIEF smooths the patches of an image to minimize noise. Afterward, the pixel strength differences are utilized to exemplify the descriptor.

- Key-point of fast-retina (FREAK) (Alahi et al. 2012; Işık 2014): the descriptor of FREAK suggested by (Alahi et al. 2012) utilizes a grid circular for a sampling of retinal. Patterns of 43 samplings utilized by this descriptor displayed in Figure 2.4 depend on visual fields of the retinal. visual fields of these 43 are sampled via factors reducing in order to obtain a descriptor binary such the space from the potential pairs of thousand to the middle yields of a patch. Gaussian function is employed to smooth every pair. eventually, via determining a threshold and looking the signs of variations among pairs, the descriptors of binary are exemplified.

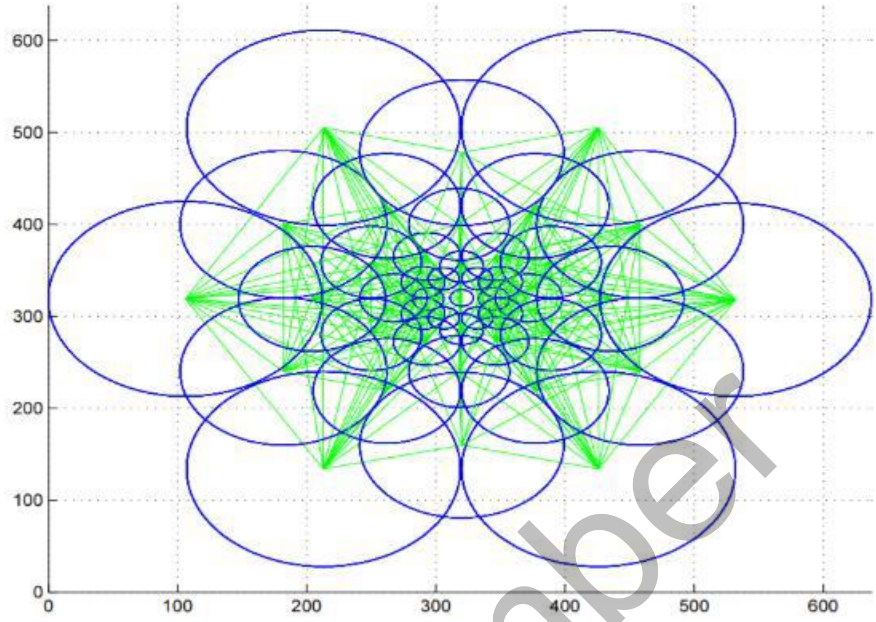


Figure 2. 4 Descriptor of key-point of fast-retina (FREAK) utilized patterns of 43 samplings (Pasandi 2014).

It is inescapable to accept the balance among efficacy and informativeness when applying descriptors of local. However, the following process cannot be retrieved during using local descriptors when lost any information is in local quantization. In addition, designing such characteristics typically demands tremendous manual effort and research practices, so they are not adaptive or suitable for set issues.

2.3.1.2 Holistic Approach

The entire face is processed through subspace or holistic methods, i.e., they do not involve face zone extraction or dots features (noses, mouth, and eyes). The main purpose of these methods is to exemplify the image of the face through pixels matrix and smooth their handling. This matrix is also transformed into characteristic vectors. Afterward, in space dimensional low, are executed these vectors of the feature. Even so, subspace or holistic methods are ticklish to differences (poses, expressions, facial, and lighting), and these benefits make these approaches commonly employed. In addition, examples of holistic methods depend on the system utilized to exemplify the subspace are described below.

- Eigenface (Turk & Pentland 1991) and principal component analysis (PCA) (Lima et al. 2004; Seo & Milanfar 2011): Eigenface (Turk & Pentland 1991) it is representing one of the premature techniques to holistic exemplification of features.
- Analysis of Principal Component (PCA) (Person, 1901) is the main part of Eigenface. The eigen vectors of the covariance matrix learn calculated from the set face of training, signified across as eigenfaces. As an eigenfaces linear mixture, every face picture is rebuilt.
- As face models or Eigenfaces, the principal components generated via the PCA method are utilized. The PCA approach converts a number of theoretical variables correlated into a tiny numeral of variables untrue named "principal components."
- The PCA aims to decrease (variables observed) the big dimensionality for space of the information to (variables independent) the junior essential dimensionality for space of the feature necessary to represent the data economically. Figure 2.5 displays how a tiny numeral of features can exemplify the face. PCA measures the covariance matrix's Eigenvectors and translates the original information onto a smaller dimensional space of the feature specified via big Eigenvalues of Eigenvectors. In face recognition and exemplification, has been utilized PCA, in which the measured Eigenvectors are known as Eigenfaces (as displayed in Figure 2.6).

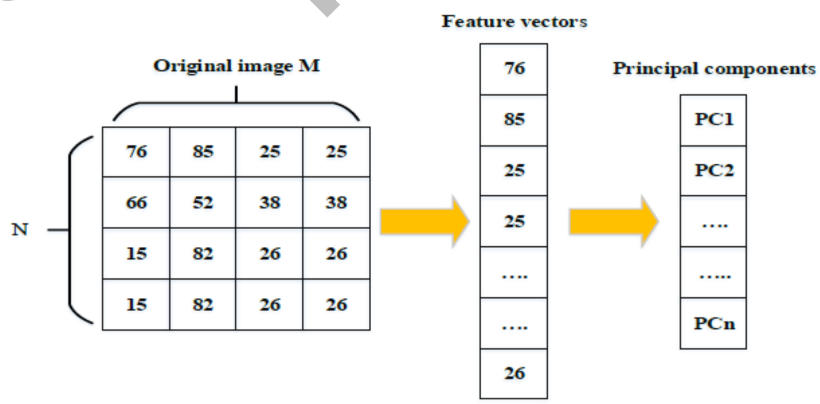


Figure 2. 5 Epitome of dimensional lowering when analysis of principal component (PCA) is applied (Lima et al. 2004).

A picture also may be reflecting the dimension vector of $M \times N$, consequently which an exemplary picture of size 4×4 turns into the dimension vector of 16. Let the collection of pictures of the training be $\{X_1, X_2, X_3 \ldots X_N\}$. The set's mean face is described as next in equation (2.7):

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \quad (2.7)$$

Compute the matrix of covariance estimate for exemplify the degree of dispersion for all vectors of the characteristics related to vector of mean. The matrix of covariance Q is described via the following equation (2.8):

$$Q = \frac{1}{N}\sum_{i=1}^{N} (\bar{X} - X_i)(\bar{X} - X_i)^{\mathrm{T}} \quad (2.8)$$

The Eigenvectors and corresponding Eigen-values are computed using equation (2.9)

$$CV = \lambda V, (V \in R_n, V \neq 0) \quad (2.9)$$

where V is the set of eigenvectors matrix Q associated with its eigenvalue λ. Project all the training images of i^{th} person to the corresponding Eigen-subspace as shown in the equation (2.10):

$$y_k^i = w^T(x_i), (i = 1,2,3 \ldots N) \quad (2.10)$$

where the y_k^i are the droppings of x, as well as, are named the components of principal, also signify as eigenfaces. The pictures of face are exemplified as a linear collection of these vectors "components principal ". LDA & PCA are two separate feature extraction approaches that are utilized to extract facial characteristics to recognize features facial, neural networks and fusion of wavelet are employed. For assessment, the database ORL is utilized. The five first faces-Eigen created from the database ORL are presented in Figure 2.6: (Devi et al. 2010).

Figure 2. 6 The five first faces-Eigen acquired from the database ORL (Devi et al. 2010).

- Analysis of linear discriminative (LDA) and face-Fisher (B. Li & Ma 2009; Simonyan et al. 2013): The method of Face-Fisheries depends upon the same theory of resemblance as the approach of Eigenfaces. Rather than the PCA approach, the purpose of this approach is to decrease the image space of high dimensional depend on (LDA) approach of the analysis of linear discriminant. The LDA approach is widely utilized for recognition of face and dimensionality lowering (R. Agarwal et al., 2019). Although PCA is approach of unsupervised, LDA is approach of learning supervised and utilizes information and data. The within-class matrix of scatter S_W and the among-class matrix of scatter S_B are described as follows in equations (2.11) and (2.12) for within all classes, all samples:

$$S_B = \sum_{I=1}^{C} M_i(x_i - \mu)(x_i - \mu)^T \tag{2.11}$$

$$S_w = \sum_{I=1}^{C} \sum_{x_k \in X_i} M_i(x_k - \mu)(x_k - \mu)^T \tag{2.12}$$

Where X_i exemplify the group of belonging samples to class i with x_k being the number picture of which class, μ is the samples mean vector belonging to class i, c is the classes featured number, and M_i is exemplify in class i the training samples number. For both face groups, S_B signify the characteristics scattering about the total mean, and S_w signify the characteristics scattering about the mean of every class of face.

The aim is to optimize the percentage $det|S_B|/det|S_w|$, in other words, minimizing S_w while maximizing S_B. The five first faces-Fisher and Eigen-faces (Devi et al. 2010) acquired from the database ORL are displayed in Figure 2.7.

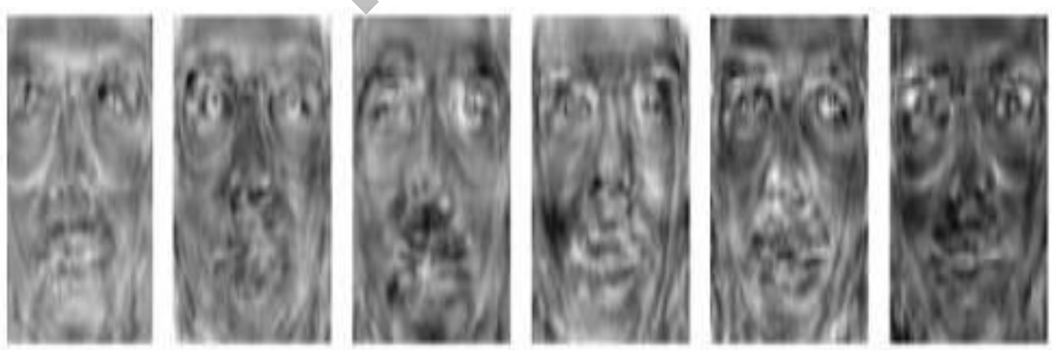


Figure 2. 7 The five first faces-Fisher acquired from the database ORL (Devi et al. 2010).

- Independent component analysis (ICA) (Annalakshmi et al. 2019): For computing the given basic space vectors, the ICA approach is utilized.

- It is decided which they are not mutually orthogonal. The purpose of this approach is to carry out a transformation of a linear in order to minimize the dependency of the statistical among the various basic vectors, that permitting components independent to be analyzed. Furthermore, in variables uncorrelated, the acquisition of pictures from various sources is pursued, which makes it easy to achieve greater quality since ICA obtains pictures within variables of independent statistically.
- Filters of Gabor: filters of the Gabor are sinusoids spatially located via a window of a Gaussian which, through determining their orientation, scale, and frequency, permit the features to be extracted from pictures. Filters of a Gabor are converted according to the pose and form to extract the characteristic vectors of the face picture together with the PCA in the work of the face picture to increase the output in unregulated environments for recognition of face (Perlibakas 2006). To delete the redundancies and obtain the superior description of face pictures, the PCA is employed to characteristics of the Gabor. Eventually, to test the resemblance, the metric of cosine is utilized.
- Analysis of Frequency domain (Z.-H. Huang et al. 2015; Sufyanu et al. 2016): As a function of frequency components depressed with energy aloft, the analysis approaches in the frequency give an exemplifying of the face human. The methods of (DWT) transformation of discrete wavelet, (DCT) transformation of discrete cosine, and (DFT) transformation of discrete Fourier are data-independent, and therefore do not demand training.

Holistic exemplification, however, is typically costly in characteristics calculation, so it cannot expand for large-scale issues.

2.3.2 Classification for Face Recognition

The purpose of classification is to learn a model for making forecasts in un-visual data depend on former observations. Several classifiers are taught in a supervised way, i.e., samples or prior observations are presented with tags of the ground-truth. Consequently, the generalization efficiency of the supervised classifier depends heavily on the adequacy of the named samples of training. In the following, several of the model supervised classifiers are presented.

- The classifier of (NN) Nearest-Neighbor: The 1-N-N classifier in machine learning is possibly one of the easiest classifiers. The mark of the un-visual sample dot is expected to be the same as the closest adjacent sample in the 1-N-N classification. The classifier of K-N-N is a common various1-N-N, allowing vote prediction for the preponderance dependent on the mark distribution of the closest adjacent points K rather. The notion of "NEAR" is generally described in terms of gauge distance in a feature space. Consequently, the efficiency of classifier of the nearest neighbor depends to a large degree on the usefulness of the exemplification of features.
- The classifier of Naive Bayesian: According (Domingos & Pazzani 1997), the classifier of Naive Bayes falls under the probabilistic classifier in general class depend on Bayes' theory. The classifier of Bayesian depends on the easy hypothesis, which says the vector feature dimensions are independent of one another. Regarding the theorem of the Bayes, can be formulated the probability posterior for classification trouble as next in equation (2.13).

  $$P(y_k \mid x_i) \propto P(y_k) \cdot \prod_{j=1}^{n} P\left(x_i^{(j)} \mid y_k\right) \tag{2.13}$$

  Where y_k exemplify the class k mark, and $P(\cdot)$ exemplify probability. Consequently, the forecast is made via selecting a class mark that gives the highest possible posterior likelihood.
- Forest Classification: Forest classification (Breiman 2001) is a special situation from random forest used to the classification trouble. Includes Forest Random with the notion of ensemble learning by preparation decision-making-trees in the form of a huge number, each serving as a poor classifier. The implementation of bagging results in a certain degree of freedom between trees, leading to a substantial enhancement after the results merge for all trees. During the process of experimenting, a sample of an un-visual is tested versus a chain of easy divides rules for the tree nodes over the track, and eventually dropping into a node of the leaf. The distribution of the category in the identical leaf node is used to measure the posterior likelihood. Eventually, the forecasting is made by calculating the outcomes mean of every tree. The mark of the category with the

greatest likelihood is then appointing to the un-visual sample. Forest classification (Bosch et al. 2007; Fette et al. 2007; Gall & Lempitsky 2013) is greatly scalable for the classification trouble of scale wide and has been exceedingly applied for different research areas.

- Support Vector Machine (SVM):(Boser et al. 1992) they first present SVM as a classifier of a non-probabilistic classification of the binary. Later then suggested classification of multi-category based on SVM extension (Vapnik 1999). The SVM stimulus is to know the decision-making surface, which distinguishes the samples of training by optimizing the decision-making margin. The kernel hoax allows SVM to be commonly implemented to the classification of non-linear troubles as well. makes the performance of Strong generalization SVM one of the most common classifiers. The maximal margin notion is also stretched to the field Learning of Semi-Supervised.

In several former studies, the selects both of classifiers and features are specified empirically or through experiments comprehensive. The convoluted fusing for these two key crucial elements remains a challenging problem in the implementation of facial recognition in particular scenarios. To fix this problem, researchers are working and trying together to learn about the classifier and feature in a common manner way. One of the most exemplifications approaches that pursue this theory is Deep Neural Network (DNN)

2.4 ARTIFICIAL NEURAL NETWORK (ANN) AND DEEP NEURAL NETWORK FOR FACE RECOGNITION

One of the multidisciplinary fields is (ML) Machine learning, which including of a broad variety of fields, like theory of approximation. statistics, theory of algorithm, and probability. One of the first described machine learning is Arthur Samuel as the " domain of study which grants machines "computers" the capacity without being frankly programmed to a learn" (Simon 2013).

According to (2003), machine learning is usually categorized into three groups: learning of unsupervised, learning of reinforcement, and learning of supervised. The variation is fundamentally based mostly on whether or not the machine is taught. Data in both of input and identical output are given to the machine in the learning of supervised. However, no marks are given in learning unsupervised because the machine has on its own to learn. Learning of unsupervised does not constantly have the objective of clear, that implies which it is permitted to detect a target on its own. Learning of reinforcement can be handled as a balance among the two methods described above. It has an evident target. However, it requires for interact in a complex context dynamic medium in that no teaching is offered.

(Kirby & Sirovich 1990) The usage of neural works is found attractive because of its non-linearity property and also is considered efficient. One of the earliest (ANN) networks of artificial neural approaches utilized for recognition of face is adaptive network of a single layer named WISARD that includes a split network for every stored individual (Stonham 1986). The way to creating the structure of a neural network is a decisive important for effective identification. It is quite much relied on the meant application. For detection of face, (Lawrence et al. 1997) convolutional neural network, and (Sung & Poggio 1995) multilayer perceptron has been implemented.

For recognition of the face, a pyramid structure of multi-resolution is employed (Weng et al. 1993). According (Lawrence et al. 1997) suggested a crossbred neural network that merging a neural network, a map of self-organizing (SOM), sampling of the local picture, and a convolutionary neural network. In a topological space, the SOM offers samples of the picture quantization. In the original space, the inputs are close and close in the output space, thereby offering invariance of small shifts and dimensional reduction in sample of the image. Finally, the convolutionary network extracts sequentially greater features in the layers hierarchical group and offers fractional invariance in rotation, deformation, gauge, and translation. The researchers recorded 96.2 percent accurate identification of 400 photos of 40 subjects in the ORL database.

As a completely computationally costly method, the procedure usually demands range 200-400 displays for each classifier to be educated. These preparation patterns involved variation and translation in expressions of facial, aiming to enhance recognition. Data set built utilizing 16 individuals were employed with sixteen classifiers. The neural network of the Kohonen is one of the extreme commonly utilized neural network architectures of unsupervised. (Kohonen 1988) one of the first algorithms of neural network was utilized in the field of facial recognition, is the map of Kohonen associative. The correct recall was recorded when using this procedure on a limited data set even though parts of the input image were either absent or noisy. Suggested (Seibert & Waxman 1992) a method for the identification of faces from their sections utilizing a neural network. This modular system integrated views of 2-D from various vantage dots, such that influential characteristics such as nose and eyes gamed a role of controlling into 2-D views.

Utilize structure networks of the Hyper Base to recognition of facial was studied by (Brunelli & Poggio 1993). According to this study, using affine transformations of 2-D the picture was transformed first, to eliminate differences due to an alteration of view. Detected locations of the mouth and eyes in the face picture were utilized to achieve the characteristics of transformation and the locations of these features. Subsequently, the effects of lighting in the picture were minimized via adding the operator of directional derivative to the picture transformed. The outcome was multiplied through a function of the Gaussian and merged into the receptive area.

(Lin et al. 1997) proposed a probabilistic decision-based NN for face recognition. It adopts a hierarchical network structure with non-linear basis functions and a competitive credit assignment scheme. Hopfield model with pattern matching is proposed by (Dai & Nakano 1998) for face recognition. In (Lawrence et al. 1997), a hybrid NN called Convolutional Neural Network (CNN) is proposed, partially invariant to changes in the local image samples, scaling, and translation and has better classification than eigenface methods.

In (F. J. Huang et al. 2000), an ensemble of NN classifier is used to classify the view-specific eigenface features. (Gutta et al. 2000) describes the mixture of experts for gender classification, pose classification and showed their feasibility on the FERET database of the face image. The mixture consists of an ensemble of RBF. Inductive decision trees and support vector machine (SVM) are used for decision strategy. A separable Low complexity 2D HMM model combined with appropriate DCT feature block size is explored by (Othman & Aboulnasr 2003).

(McGuire & D'Eleuterio 2001) proposed an algorithm called eigen pixel. The image is broken down into a local receptive field called localized PCA, similar to the processes that occur in the visual cortex of many vertebrates and humans. A simple one-layer error-correcting Neural Network is used for classification.

To learn functions that are not linearly separable, expansion encoding is done in which additional degrees of freedom are introduced. Compared to CNN it exhibits less error.

A framework for facial recognition in monochromatic pictures still (L. G. D.-S. Huang 2003) was suggested depending on a neural network such as the Cresceptron structure. Cresceptron design was basically a structured network of a multi-resolution pyramid that uses automated, increasable learning. For locating faces into the picture, the face images were operated first via a rule-based algorithm before implementing the learning of Cresceptron. In order to achieve generalization a good, the patterns number in the network must be much greater than the number of the weights.

However, it is extremely difficult to produce a big sample. (Er et al. 2002) discussed the challenges for networks of multilayer feedforward for recognition of a face. To address the problems of over-fitting, over-training, and effects tiny-sample, the supervised clustering method was proposed to locate initial parameters and the structure for classifier of the neural network of the RBF prior to learning events. Modular memory-depend models were shown to be effective in solving limited sample size trouble (B. Zhang et al. 2006; B. L. Zhang et al. 2004).

According (Lin et al. 1997) utilized (PDBNN) neural network of probabilistic decision-based that patrimonial the structure of modular from its former, (DBNN) neural network of a decision based (Kung & Taur 1995). It can be easily implemented PDBNN to (1) detection of face, (2) locellate of eye, and (3) recognition of a face. Rather, the network is split into subnets of K. every subgroup is devoted to the identification of a single subject in the database. PDNN utilizes the activation function of the Guassian to its neurons, as well as the output of every "subnet of face" is the weighted sum of the output's neuron. In other way of explanation, subnet of the face estimates the probability density utilizing the famous Guassian mixture. The Guassian mixture offers a much more fluid and complicated model compared to scheme of the AWGN, for approximating the probability densities of the time in the space of the face. The PDNN learning framework involves of two stages. The first stage; trains every subnet was via its own pictures of the face. Learning of decision-based was named of the second stage, the parameters of the subnet may be trained via some samples of particular from another face groups. Framework learning of the decision-based does not utilize each sample of the training for purposes training. Only wrongly classified trends are utilized. If wrongly classified the sample to the incorrect subnet, the correct subnet can change its parameters thus which its decision-area can be shifted closer to the wrongly sample classified.

The biometric recognition method based on PDBNN has the traits of both statistical and neural networks methods, and its dialed computing concept precept is surprisingly simple to apply onto the parallel computer. According (Lin et al. 1997), it was recorded which PDBNN had the potential to identify up to 200 individuals and could reach up to 96 percent of the correct recognition score.

Even So, as the number of users increases, the cost of computation can become further challenging. Methods of neural networks, in general, are problematic as the number of groups (i.e., subjects) increases. Furthermore, they are not appropriate for a one model picture recognition exam since multiple model pictures per individual are required to the systems train to the "optimal" parameter tuning.

Another researcher (Latha et al. 2009) has demonstrated a visually proven computation for the identification of face view-points. The ability of the face image is reduced by the main part check (PCA), during which the confirmation is rendered by (BPNN) the neural back-inducing network. Existing 200 images face from the Yale database are obtained, and certain execution figures, such as execution time and approval, are resolved. Neural-depend facial confirmation is lavish and provides a greater degree of affirmation.

According to (M. Agarwal et al. 2010), the mechanism for identifying the face is based on clarification of the organization and deciphering the image of the face. Foreseeable device is cooperative with two phases for extraction of the feature implements guidance division enforcement and examination to utilize feedback well-sophisticated neural network perpetuating. This conspiracy is, however, self-determining the steep estimate. The attestation of the system structure is implemented based on eigenface, the ANN with the PCA. Significant section of reconnaissance for identifying the face relies on the knowledge of the clarification method under which the cumulative details in the face image are separated as competently as it may be understood to be protected by the priority.

The essential variation among machine learning and deep learning is how data is viewed in the framework system. Networks of deep learning depend on layers of ANN (artificial neural network), while algorithms of machine learning nearly always demand data structured. Deep learning in practice is a subgroup of machine learning which accomplishes large both of flexibility as well as power via learning to exemplify the world as an overlapping hierarchy of notions, with every notion described in connection to simpler notions and abstract more exemplification calculated in expressions of ones abstract less (Parkhi et al. 2015; Wu et al. 2018).

The fast exponential growth of Internet-based technologies applications, such as networks of social, engines of search, and websites of sharing video, has performed in a data influx in quantity. The sources large of data transfer community of the human to the epoch of data huge, and huge data fetches both opportunities as well as challenges in machine learning and computer vision for research areas.

Deep learning is simulated by such direction, also renowned as (DNN) the Deep Neural Network, which has arisen as a new science of interest in recent years and has drawn a great deal of concern from both academia and industry. The study (Hinton & Salakhutdinov 2006) has motivated many studies and simulated implementations in several areas. (Farabet et al. 2012; G. B. Huang et al. 2012; Krizhevsky et al. 2012; Nair & Hinton 2010; Sun et al. 2013).

Compared to traditional facial recognition systems, deep learning merges education feature exemplification and classifier learning commonly. The learning process is performed straight with respect to the goal of the trouble; consequently, the learning feature is appropriate to the target.

In addition, DNN embraces a hybrid structure with several extraction layers of features-the lower layer's intermediate exemplification as inputs are forwarded for the toper layer. In many instances, the non-linear activation function is implemented to obtain superior generalization for every layer's outputs.

In comparison for learning metric with shoaly linear transformation framework, the deep hybrid of projections of non-linear offers DNN with a higher degree of abstraction and discrimination ability, that has been shown to be efficient in several modern works (Cui et al. 2013; Davis et al. 2007; Guillaumin et al. 2009).

DNN proposes a generic solution for the classification trouble. It does not involve pre-processing of trouble--specific of information or extraction of feature. The raw data is used as the network's input by researchers in most instances. The network's learning is done via continuously alternating among back propagation and forward propagation frequently until convergence.

The inputs are moved via layer by layer of the network during the forward propagation, as well as, objective or function of a cost is determined to depend on the carryout of the definitive layer in the present iteration. The cost gradients are then measured with respect for intermediate inputs and the weights for every layer and further propagated for the previous layer with respect to the law of the chain. While, for updating parameters in back-spreading measures, Stochastic Gradient Descent (SGD) is used as a popular alternative.

In recent years, have been suggested different architectures for deep learning. (Schroff et al. 2015) introduced a structure, called FaceNet, in their work that easily takes a mapping from images of face to junior space of Euclidean that relates to face similarity percentage. The merit of their strategy is a lot further popular descriptive efficacy.

The purpose of this study is to confession confront from either indeed a solitary picture or a series of appearances accompanied in a recording "video" (Parkhi et al., 2015). Delayed progress has been attributed due to two factors elements:

(i) beginning to the end learning on an errand using the algorithm system of (CNN) convolutional-neural
(ii) the usability of large-scale dataset planning.

The automatic recognition of expression facial (FER) suggested in (Mollahosseini et al. 2016) has stayed intriguing and testing, the problem with respect for PC vision. Despite the efforts made to generate various FER approaches, current methodologies have required to be popularized when related to indistinct images or those captured in the wild.

A new approach for recognition of face was suggested (Kamencay et al. 2017) with three popular methods of picture recognition, such as (LBPH) histograms of local binary pattern , (KNN) histograms of K-nearest neighbor (K-nearest neighborhood), and (PCA) analysis of principal component, by using the coevolutionary-neural-network. In their presentation, they utilized KNN, LBPH, PCA, and CNN for precision. In this article, 98 percent were effective in achieving accuracy by utilizing CNN for superior results.

(Rothe et al. 2018) this study suggested a deep learning scheme to infer age from a solitary face image via not utilizing spots of facial tourism and to display dataset

IMDB-WIKI, the largest free dataset of gender and age labelled for face images. On the other hand, the investigation of the simple age prediction or age as seen by various people from the face image is an ongoing undertaking that manages the two assignments with our the VGG-16 (CNNs) coevolutionary neural networks engineering that on ImageNet are pre-prepared for features of the image. They portray the topic of estimation age as a deep issue of characterization trouble address through the softmax foreseeable amelioration of esteem.

2.5 OUTLINE

This chapter has offered a detailed review for recognizing the face and a description and review of the subsystem, such as Various methods of face recognition techniques from the literature. So, it is presumed, a superior feature extraction technique with a high-quality classifier based on deep learning can make a face recognition system effective. However, a recognition rate of 100% is still very difficult to achieve using prevailing approaches for images with illumination, expression, and variation in the pose. Therefore, it is also inferred that a classification technique capable of discriminating the overlapping features of the images can bring out an effective face recognition system.

CHAPTER III

METHODOLOGY

3.1 OVERVIEW

Technology has acquired a significant turn with the inculcation of nearly every type of Endeavor of human inside technology of information. Recent juveniles have demonstrated the need for increased protection, especially in this era of the Internet, where everything needs to be done to strengthen authentication and security.

In computer vision, the most critical aim is to obtain a visual perception confession capability equal to humans, as discussed in the study (Sarawat Anam et al., 2009). In the sight of this, several researchers have adventured to utilize the Face as a quick and safe means of authentication, with practically no claims on the person who utilizes it.

Recognition of Face has been promoted to be a primary source of encryption and improved verification to substitute fingerprint identification with a diversity application in ID card of national, passport biometric, airport security, forensics, e-commerce & banking, and other applications. Even now, the list is growing, with researchers exploring different ways in which the peculiar and naturally distinct features of Face may be used.

The Recognition system of the face is still not perfect in terms of accuracy. Therefore, several research concerns simultaneously focus on large-scale identification, localization of face, design improved face recognition algorithms, and evaluation in contexts.

This study is one such potential effort to enhancing face recognition. The research methodology framework used in this thesis is based on a hybrid proposed model of

Gabor Filter and Stacked Sparse Auto Encoders (SSAE) to obtain the best face recognition rate.

3.2 RESEARCH METHODOLOGY

The main objective of this research is to develop a robust system for recognizing the face with high accuracy. Therefore, a hybrid model was proposed to satisfy this objective to improve face recognition using two datasets of face images; the OLR dataset and the Extended Yale-B database. The proposed hybrid model was applied by using Gabor Filter and Stacked Sparse Auto Encoders (SSAE). Thus, the framework methodology of the proposed model includes three main stages: pre-processing (resize the input face image to half), feature extraction using Gabor filter, and face prediction using SSAE Deep Neural Network. Figure 3.1 shows the framework diagram of the proposed hybrid model.

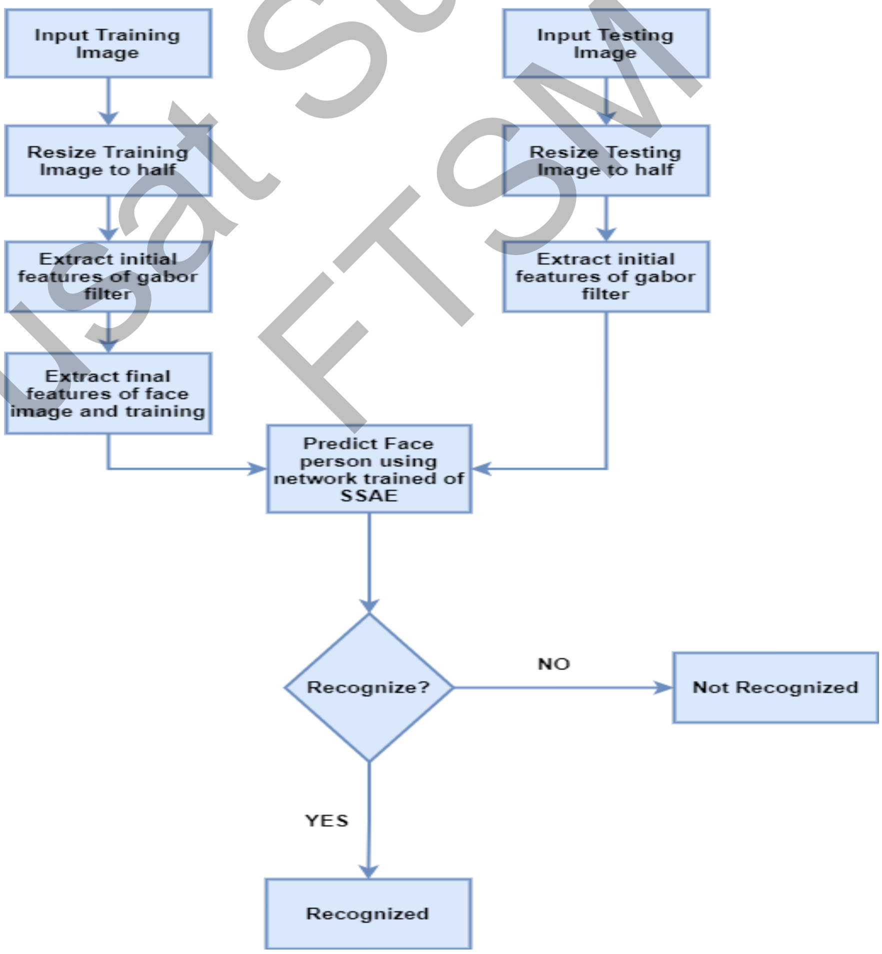


Figure 3. 1 The framework methodology of the proposed face recognition model

3.3 FACIAL DATASETS

The collecting of facial databases for purposes benchmarking was an integral section of the fixed growing made in automatic recognition expression and facial expression. In the 1990s, for automated face recognition, new methods have been motivated due to major developments in sensor and computer technology. There are presently many databases utilized for facial recognition that range in expressions, conditions, pose, lighting, occlusions, size, and the number of subject's image.

The two datasets were used in this thesis. The first dataset is the ORL database containing a compilation of faces taken at the Olivetti Research Laboratory between April 1992 and April 1994 in Cambridge, United Kingdom (Cambridge 1994).

There are 40 different distinct subjects, and with each subject are 10 different photographs. On certain subjects, the photographs were taken at various times. There are differences in detail of facial (no glasses/glasses) and appearance of facial (non-smiling/smiling, closed eyes /open eyes). All photographs were taken versus a dark homogeneous backdrop with subjects in a frontal posture, upright, tolerance for any rotation, and tilting up to around 20 degrees. There is some variance in the scale range of up to around 10%. Figure 3.2 presents the sample images of the OLR dataset that. The pictures are greyscale and have a resolution of 92 × 112. In this work, resize the OLR images to half in order to increase the time computation.

Figure 3. 2 Samples of OLR dataset images through two persons in different poses.

Figure 3. 3 Samples of OLR dataset images from every person in different poses.

However, the second database used in this thesis is Extended Yale-B, consisting of 2,432 frontal-face pictures with dimentions192×168 across through 38 various people (Georghiades et al. 2000). Thus, there are 64 photographs per subject, and they differ in illumination.

The photographs were taken under different expressions of facial and varying conditions of lighting. Face photographs differ widely in lighting across objects, so much so which only a tiny piece of the face is apparent at times. The version of the close-cropped dataset is utilized, where each photograph is cropped to have a face with no hair or backdrop. Also, in this thesis, resize the Extended Yale-B database pictures to half to reduce the time computation.

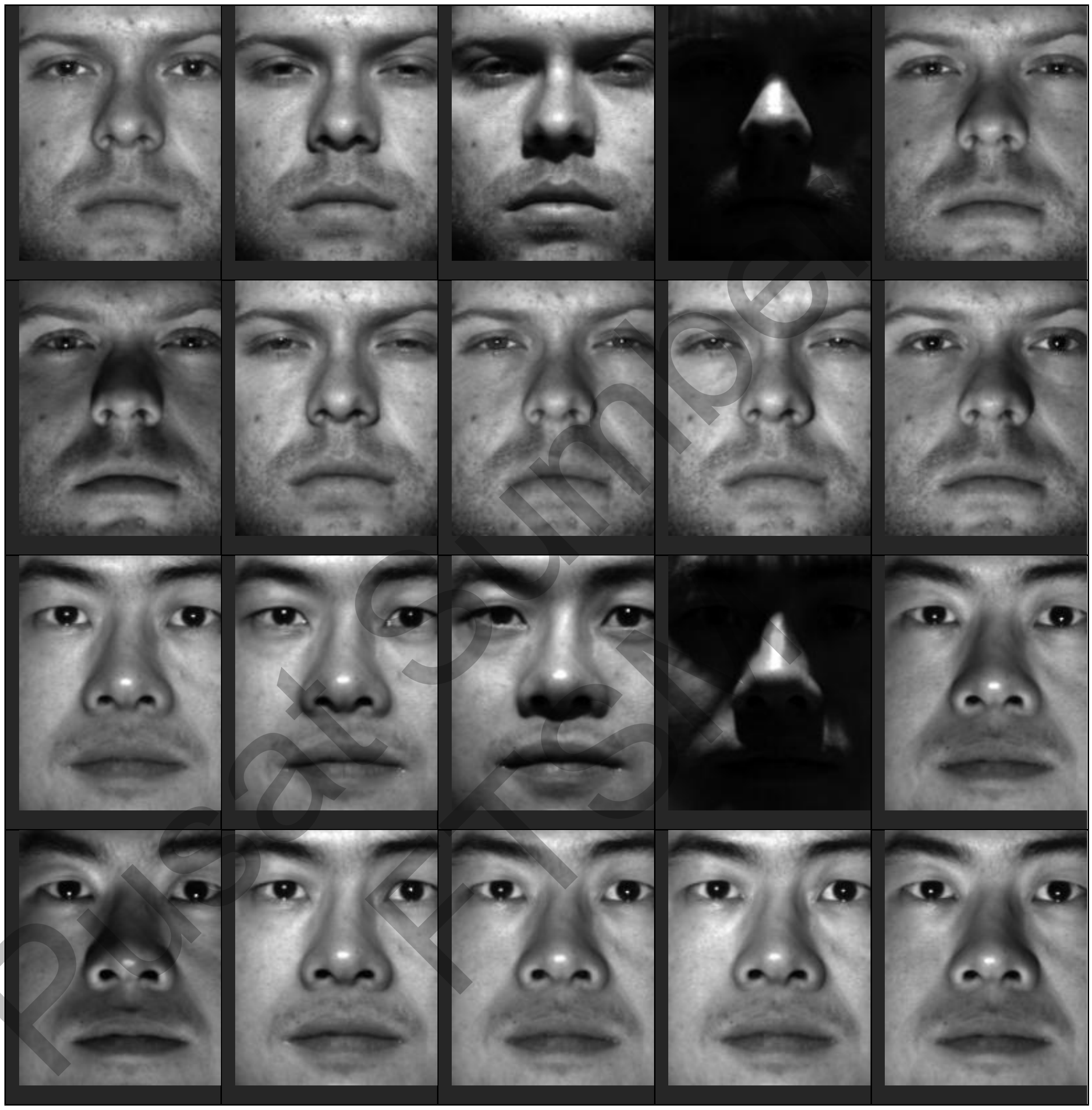


Figure 3. 4 Images of Extended Yale-B database through two poses.

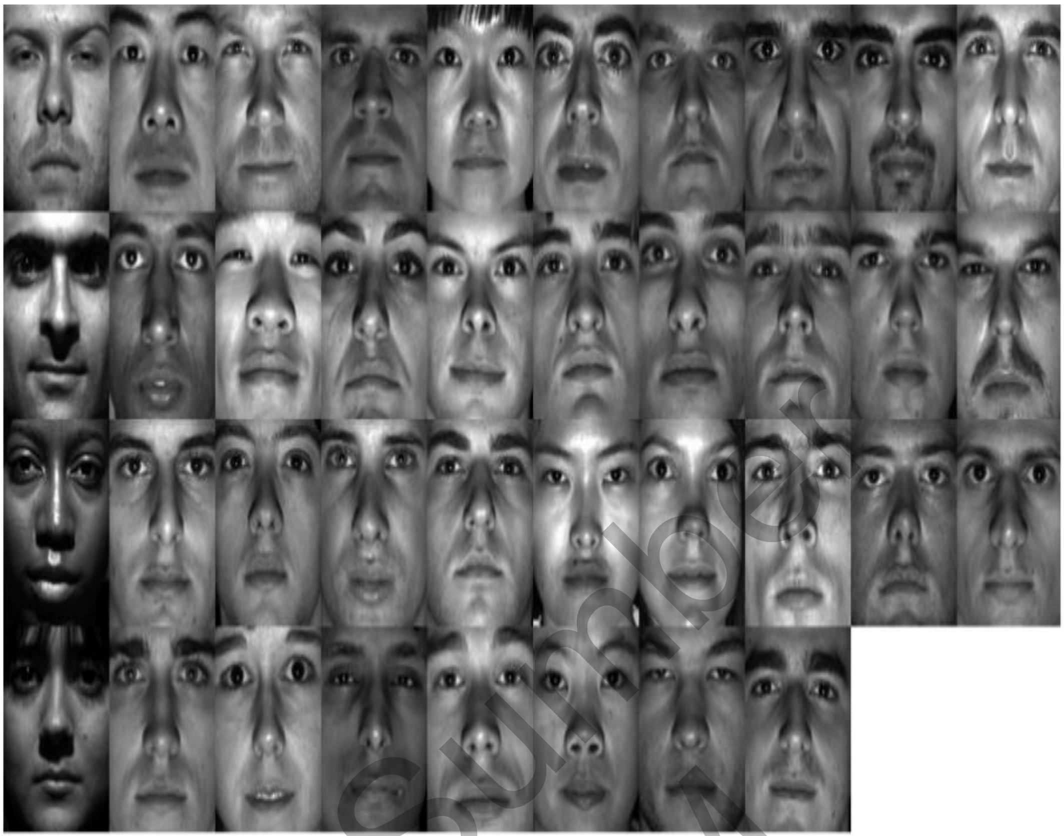


Figure 3. 5 Images sample of Extended Yale-B database from every poses.

Moreover, for these two datasets, description in more detail can be illustrated in the following Table 3.1.

Table 3.1 2D face databases. The image variations are represented by (i) illumination, (p) pose, and (t) delay time.

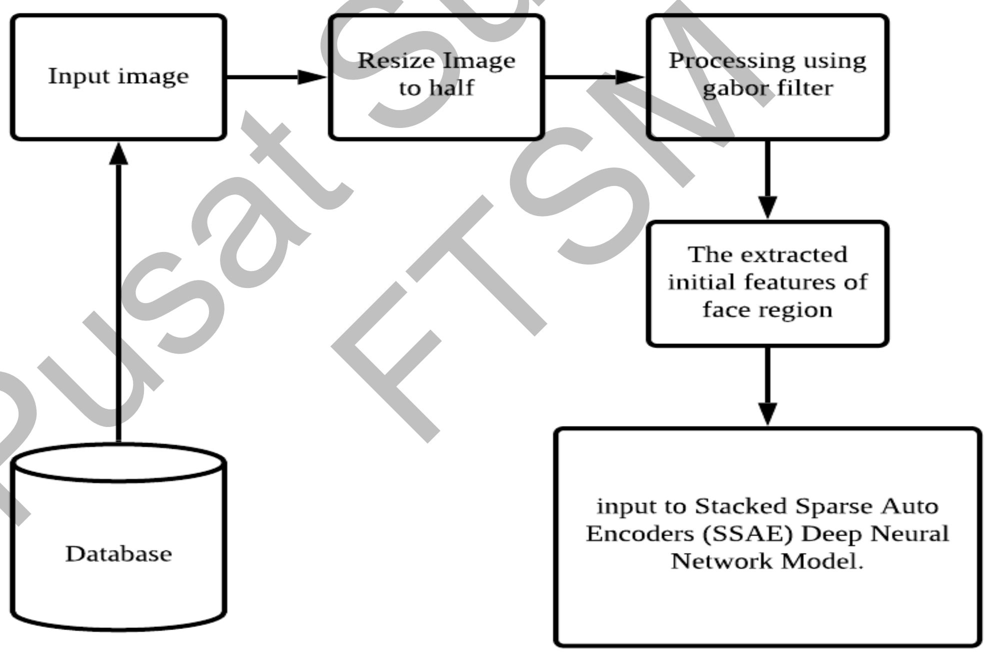
Database	RGB Color/grey	Images Size	No. of Persons	Number of Images/Person	Variation	Description
ORL	gray	92 x 112 pixels	40	10	i,t	-Always with a dark background -A limited number of people -Not consistent lighting conditions from one image to another -Non-annotated images for different facial expressions, head rotation, and lighting
Extended Yale B	gray	168 x 192 pixels	38	64	p,i	-Variations in pose (9 poses) -Illumination conditions (64)

3.4 THE HYBRID PROPOSED MODEL OF FACE RECOGNITION

Hybrid Algorithm meaning is a combination of more than one algorithm for efficiently solving the problem to yield better performance than the individual algorithms. (Tad Gonsalves 2015,).

The hybrid proposed method is to combine two algorithms using Gabor Filter and Stacked Sparse Auto Encoders (SSAE) for face recognition. The first step is to resize the input image to reduce the execution time. Then, initial features of the face region are extracted by implementation the Gabor filter that considers as the input to SSAE deep neural network, such as shown in Figure 3.6. The main contribution of this research is to improve the accuracy of the results after applying this proposed hybrid model is to reduce the time-consuming in face recognition that comes from the effect of several types of deformations and noise.

Figure 3. 6: The proposed hybrid model steps



3.4.1 Gabor Filters

Gabor filters have been widely utilized in photographs processing and analysis of texture due to their excellent characteristics, which include optimum joint localization of spatial-spatial-frequency and the ability to mimic the receptive simple cells fields in the cortex visual (J. G. Daugman 1985; Hamamoto et al. 1998; Jain & Farrokhnia 1991).

Therefore, the theory of Gabor filters and Gabor filters-based feature extraction method will be explained in more detail in the following subsections.

3.4.1.1 Theory of Gabor Filters

Gabor filter of two-dimensional is a complicated modulated sinusoidally function of Gaussian with the response in the field of spatial (Equation 3.1) and field of spatial-frequency (Equation 3.2) as follows (Figure 3.7):

$$\begin{aligned} &h(x, y; \lambda, \phi, \sigma_x, \sigma_y) \\ &= \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left[\frac{R_1^2}{\sigma_x^2} + \frac{R_2^2}{\sigma_y^2}\right]\right\} \\ &\qquad \times \exp\left[\mathrm{i} \cdot \frac{2\pi R_1}{\lambda}\right], \end{aligned} \tag{3.1}$$

Where

$$R_1 = x\cos\phi + y\sin\phi$$
$$R_2 = -x\sin\phi + y\cos\phi$$

$$\begin{aligned} &H(u, v; \lambda, \phi, \sigma_x, \sigma_y) \\ &= C\exp\left\{-2\pi^2\left(\sigma_x^2\left(F_1 - \frac{1}{\lambda}\right)^2 + \sigma_y^2(F_2)^2\right)\right\} \end{aligned} \tag{3.2}$$

Where

$$F_1 = u\cos\phi + v\sin\phi$$
$$F_2 = -u\sin\phi + v\cos\phi, \quad C = \text{const}$$

Gabor filter spatial localization can be represented via Δx and Δy, which are regular measurements of efficacious spatial widths (J. G. Daugman 1985).

$$(\Delta x)^2 = \frac{\int_{-\infty}^{+\infty} hh^*(R_1)^2 \mathrm{d}(R_1)}{\int_{-\infty}^{+\infty} hh^* \mathrm{d}(R_1)}$$

$$(\Delta y)^2 = \frac{\int_{-\infty}^{+\infty} hh^*(R_2)^2 \mathrm{d}(R_2)}{\int_{-\infty}^{+\infty} hh^* \mathrm{d}(R_2)} \tag{3.3}$$

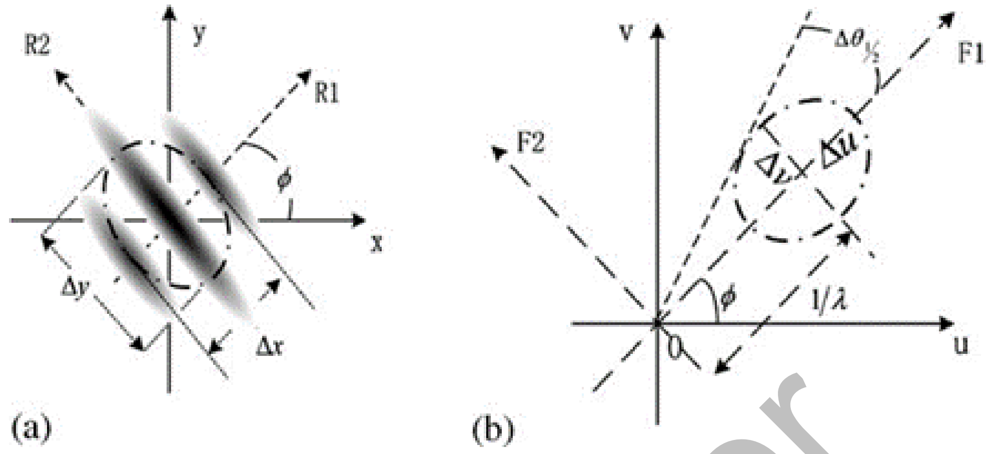


Figure 3. 7 A Gabor filter up-viewpoint in the spatial field (a) and spatial-frequency field (b).

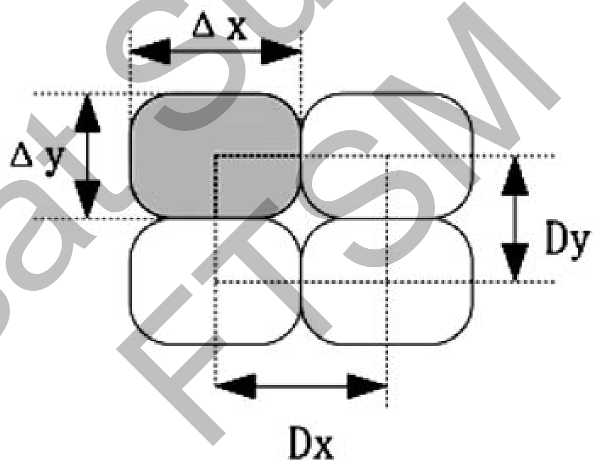


Figure 3. 8 The relationship between intervals of spatial sampling and Gabor filter widths of efficient spatial then get

$$\Delta x = \sigma_x/\sqrt{2}, \Delta y = \sigma_y/\sqrt{2} \tag{3.4}$$

In Gabor filters, the distances among filters neighbouring an image are referred to as intervals of spatial sampling Dx and Dy. To prevent unintended picture information loss, the following relationships among widths of effective spatial and intervals of spatial sampling must be followed, as seen in Figure 3.8:

$$\mathrm{D}x \leqslant \Delta x, \mathrm{D}y \leqslant \Delta y \tag{3.5}$$

Intervals of spatial sampling are critical parameters that require check when designing a Gabor filter. However, in previous studies (Hamamoto et al. 1998), it was overlooked, resulting in low results with many image detail loss.
Gabor filter spatial-frequency localization can also be expressed by Δv and Δu, which are typical efficient bandwidth measures. Like to Equation 3.3. will get

$$\Delta u = 1/\left(2\sqrt{2}\pi\sigma_x\right), \text{ and } \Delta v = 1/\left(2\sqrt{2}\pi\sigma_y\right) \tag{3.6}$$

Depend on the bandwidth of spatial-frequency, also can get another notion named bandwidth of orientation (J. G. Daugman 1985) (Figure 3.9 (b))

$$\begin{aligned}\Delta\theta &\approx 2\text{arcsin}\left((\Delta v_V/2)/(1/\lambda)\right)\\ &= 2\text{arcsin}\left(\lambda/\left(4\sqrt{2}\pi\sigma_y\right)\right)\end{aligned} \tag{3.7}$$

In this thesis, can be expressed localization of spatial-frequency in two-dimensional space in two aspects: selectivity of line orientation and selectivity of line width, which are clarified in Figure 3.9. The Gabor filter is the most sensitive $h(x, y: \lambda, \phi, \sigma_x, \sigma_y)$ to lines with the orientation $\phi + {}^{\pi}/_{2}$ and the width λ /2.

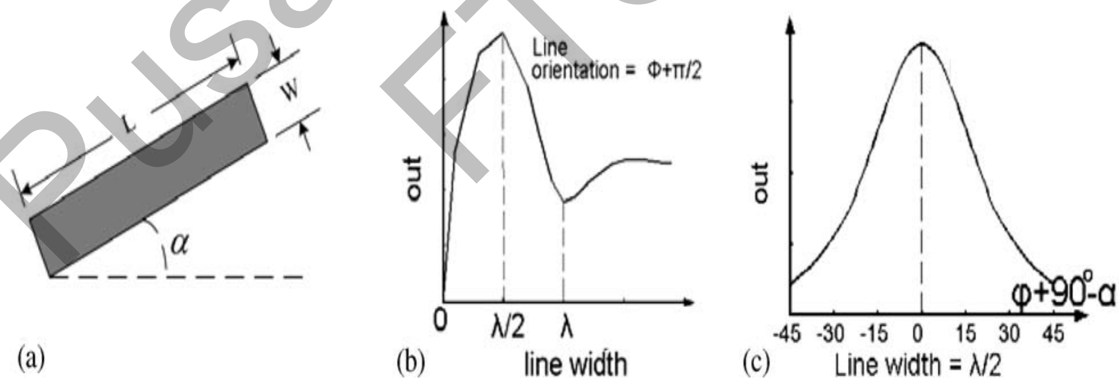


Figure 3. 9 For a line with width W(L≫W) and length L as demonstrated in (a), when its orientation α and width W alteration, the Gabor filter maximum output (only real parts) $h(x, y: \lambda, \phi, \sigma_x, \sigma_y)$ will alteration correspondingly as demonstrated in (b) and (c), That propose selectivity of line orientation and selectivity of the line width of the Gabor filter (X. Wang et al. 2005).

3.4.1.2 Gabor Filters-Based Feature Extraction Method

Method of feature extraction-based Gabor filters (Haghighat et al. 2015) for face picture is utilized to extract and locate initial features from the face region. The most significant merit of Gabor filters is their invariance to translation, rotation, and scale.

Moreover, they are robust against disturbances of photometric, like illumination variations and noise of images (Kamarainen et al. 2006; Liu & Wechsler 2002; Meshgini et al. 2013; Shen et al. 2007).

The features of Gabor filter are straight extracted from photographs of grayscale. A 2D Gabor filter is a function of Gaussian kernel modulated via a complicated wave of sinusoidal plane in the spatial domain using Algorithm (1) such as follows:

Algorithm (1): Gabor filter for initial feature extraction

1: Input Image (after resize to half)

2: Input values: f, π , γ , σ , and ϕ

// where f is the sinusoid frequency, γ is the ratio of spatial aspect that

// determines the ellipticity of the Gabor function's support,

// σ is the Gaussian envelope standard deviation and,

// ϕ is offset of the phase.

3: Compute the following equation:

$$G(x,y) = \frac{f^2}{\pi\gamma\eta} \exp\left(-\frac{x^2+\gamma^2y^2}{2\sigma^2}\right) \exp\left(j2\pi f x' + \phi\right)$$

4: Compute the following equations:

$$x' = x\cos\ \theta + y\sin\ \theta$$

$y' = -x\sin\ \theta + y\cos\ \theta$ // where θ exemplify the normal orientation to a Gabor

// function for the parallel stripes

The size of the face zone images utilized in our experiments after reducing to half are 56 x 46 pixels for OLR database while 96×84 pixels for Extended Yale B Face database.

In this thesis employs forty Gabor filters in eight orientations and five scales, as shown in Figure. 3.10. Using forty Gabor filters, the dimension of the feature vector is

56 x 46 x 40 = 103,040 for OLR database, while the dimension of the feature vector is 96×84 x 40 = 322,560 for Extended Yale B Face database. Since in an image the adjacent pixels are commonly highly correlated.

Furthermore, the feature picture generated by Gabor filters can reduce this information redundancy (Liu & Wechsler 2002; Shen et al. 2007). Through a factor of sixteen feature pictures are down sampled, which means the vector of the feature will have a size of 1680 for OLR database while 5280 for Extended Yale B Face database. Also, have then normalized these vectors to unit variance and zero mean.

The extracted initial features of the face region then input to Stacked Sparse Auto Encoders (SSAE) Deep Neural Network Model.

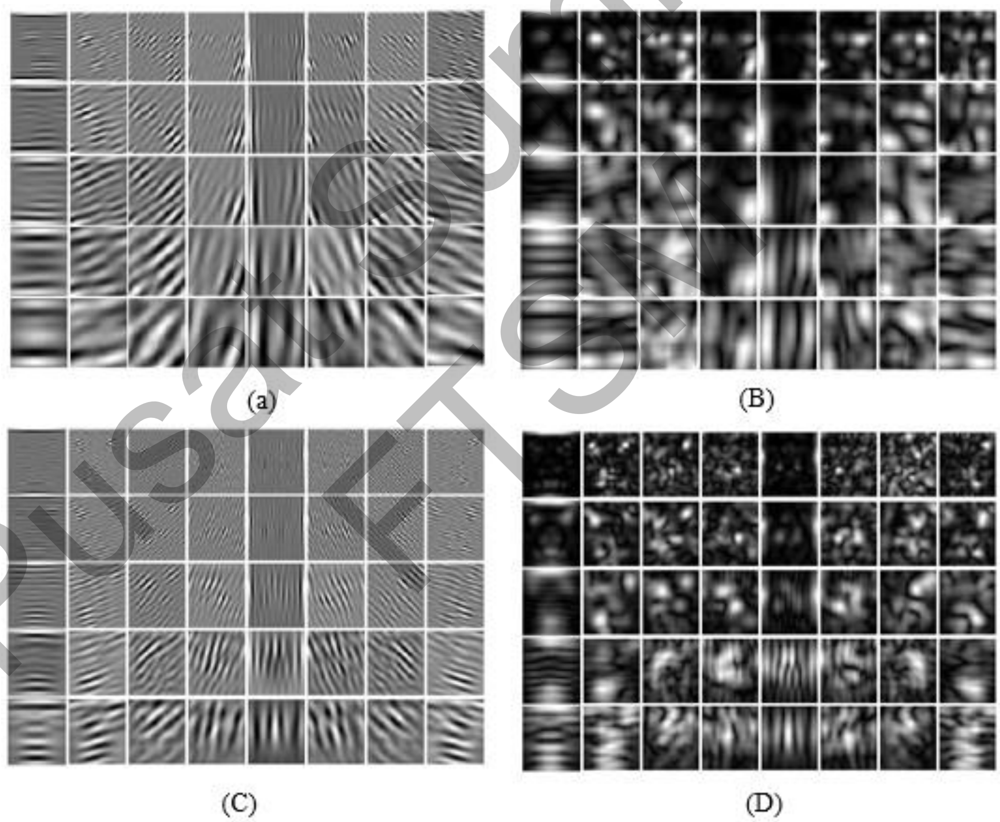


Figure 3. 10 Gabor wavelets in eight orientations and five scales (a) Real Parts Gabor for OLR database; (b) Magnitudes Gabor for OLR database; (c) Real Parts Gabor for Extended Yale B Face database; (d) Magnitudes Gabor for Extended Yale B Face database.

3.4.2 STACKED SPARSE AUTO ENCODERS (SSAE) DEEP NEURAL NETWORK MODEL

Autoencoder is a feature learning algorithm of an unsupervised that seeks to enhance better feature exemplification of high-dimensional input data via determining the correlation between the data. An auto-encoder is essentially a neural network of multi-layer feed-forward that has been learned to exemplify the input utilizing back-propagation. The autoencoder uses back-propagation to reduce the difference among input and reconstruction as far as possible via learning a decoder and an encoder (See Figure 3.11), which results in a collection of biases b and weights W.

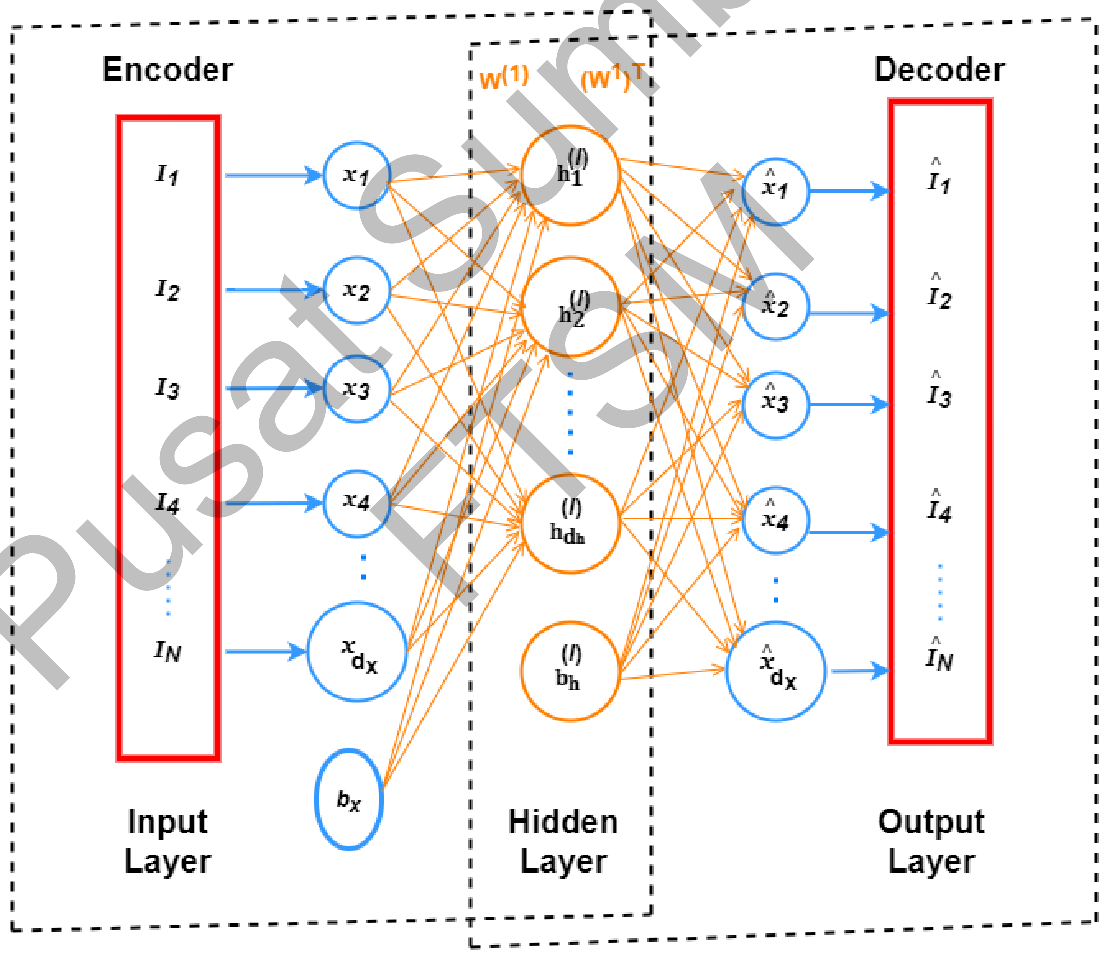


Figure 3. 11 The simple Sparse Auto-encoder architecture.

3.4.2.1 **The Basic Sparse Autoencoder** (Ng 2011)

Suppose that $X = (x(1), x(2), \dots , x(N))^T$ be the whole training (unlabeled) initial face features, where $x(k) \in R^{d_x}$, N and d_x are the training initial face features number and the number of the pixels in each initial face feature, respectively. $h^l(k) = (h_1^l(k), h_2^l(k), \dots , h_{d_h}^l(k))^T$ Indicates the feature of high-level learned at layer l for the $k-th$ initial feature of the face, where in current layer l, the hidden number units are d_h.

Throughout this thesis, we utilize a notation subscript and superscript to describe a unit of hidden layer and the hidden layer, respectively. In Figure 3.10, for example, $h_i^{(1)}$ exemplifies the $i-th$ unit in the 1st hidden layer. For easily, we indicate x and $h^{(l)}$ as an enter feature of the initial face and its exemplification at hidden layer l, respectively.

Figure 3.11 depicts the basic Sparse Autoencoder architecture. In common, the autoencoder's input layer is made up of an encoder that converts input x into the corresponding exemplification h, and the hidden layer h can be thought of as a new feature exemplification of input data. The layer of output is basically a decoder efficiently that has been learned to reconstruct an estimate of the input $\hat{x}$ from the hidden exemplification h.

Essentially, an autoencoder training aims to locate optimal parameters via reducing the difference among the enter x and its reconstruction $\hat{x}$. A cost function is used to explain this disparity. The Sparse Autoencoder (SAE) cost function is composed of three idioms like follows (Bengio et al. 2013):

$$\mathfrak{L}_{SAE}(\theta) = \left[\frac{1}{N}\sum_{k=1}^{N}\left(L\left(x(k), d_{\hat{\theta}}\left(e_{\tilde{\theta}}\left(x(k)\right)\right)\right)\right)\right] + \left[\alpha \sum_{j=1}^{n} KL(\rho||\hat{\rho}_j)\right] + \left[\beta \|W\|_2^2\right] \tag{3.9}$$

The first idiom is a mean of error squares sum idiom that defines the contradiction among entering $x(k)$ and reconstruction $\hat{x}(k)$ overhead the whole data.

Encoder $e_{\tilde{\theta}}(\cdot)$ maps enter $x \in R^{d_x}$ to the hidden exemplification $h \in R^{h_x}$, that is determined via $h = e_{\tilde{\theta}}(x) = s(Wx + b_h)$, where $b_h \in R^{d_h}$ is a vector of a bias, and W is a $d_h \times d_x$ matrix of weight. The encoder is parameterized via $\tilde{\theta} = (W, b_h)$. Decoder $d_{\hat{\theta}}(\cdot)$ maps outcoming hidden exemplification h back into entering space $\hat{x}$. $\hat{x} = d_{\hat{\theta}} =$