Hybrid Component-based Face Recognition

BY

Andile Martin Gumede

211513796

Supervisor:
Prof. S. Viriri

Co-supervisor:
Mr M. Gwetu

A dissertation submitted in fulfilment of the requirements for the degree of Master of Science in the

School of Mathematics, Statistics and Computer Science
College of Agriculture, Engineering and Science
Durban 4000

July 18, 2018

Declaration: Authorship

I, Andile M. Gumede, declare that:

- The research reported in this dissertation, except where otherwise indicated, is my original work.
- This dissertation has not been submitted for any degree or examination at any other university.
- This dissertation does not contain other persons data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
- This dissertation does not contain other persons writing, unless specifically acknowledged as being sourced from other researchers.
- Where other persons exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced, a) their words have been re-written but the general information attributed to them has been referenced; b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.
- Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.
- This dissertation does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the Reference sections.

Mr Andile M. Gumede

Declaration: Supervisor

As the candidate's co-supervisor, I agree to the submission of this dissertation.

Supervisor: Prof. S. Viriri: ____________________ Date: ________________

Declaration: Co-supervisor

As the candidate's co-supervisor, I agree to the submission of this dissertation.

Co-supervisor: Mr M. Gwetu: ____________________ Date: ________________

Declaration: Publications

DETAILS OF CONTRIBUTION TO PUBLICATIONS that form part and/or include research presented in this dissertation.

- A. Gumede, S. Viriri, M.V. Gwetu, "Hybrid Component-based Face Recognition", IEEE International Conference on Information Communication Technology and Society (ICTAS), 2017. DOI: 10.1109/ICTAS.2017.7920665.
- A. Gumede, S. Viriri, M.V. Gwetu, "Enhanced Hybrid Component-based Face Recognition", Computational Collective Intelligence, Springer LNCS, vol. 10448, pp. 257-265, 2017. DOI: 10.1007/978-3-319-67074-4_25.
- A. Gumede, S. Viriri, M.V. Gwetu, "Selecting Salient Features from Facial Components for Face Recognition", Advances in Image and Video Technology, Springer LNCS, 2017. (Accepted)

The author's contributions in each of the papers are as follows:

- 1st Author: Giving ideas, writing papers, conducting literature review, designing and implementing of algorithms.
- 2nd Author: Giving ideas, giving advice, discussing models and algorithms, proofreading and editing manuscripts.
- 3rd Author: Giving advice, proofreading and editing manuscripts.

Mr Andile M. Gumede

Abstract

Facial recognition (FR) is the trusted biometric method for authentication. Compared to other biometrics such as signature; which can be compromised, facial recognition is non-intrusive and it can be apprehended at a distance in a concealed manner. It has a significant role in conveying the identity of a person in social interaction and its performance largely depends on a variety of factors such as illumination, facial pose, expression, age span, hair, facial wear, and motion. In the light of these considerations this dissertation proposes a hybrid component-based approach that seeks to utilise any successfully detected components.

This research proposes a facial recognition technique to recognize faces at component level. It employs the texture descriptors Grey-Level Co-occurrence (GLCM), Gabor Filters, Speeded-Up Robust Features (SURF) and Scale Invariant Feature Transforms (SIFT), and the shape descriptor Zernike Moments. The advantage of using the texture attributes is their simplicity. However, they cannot completely characterise the whole face recognition, hence the Zernike Moments descriptor was used to compute the shape properties of the selected facial components. These descriptors are effective facial components feature representations and are robust to illumination and pose changes.

Experiments were performed on four different state of the art facial databases, the FERET, FEI, SCface and CMU and Error-Correcting Output Code (ECOC) was used for classification. The results show that component-based facial recognition is more effective than whole face and the proposed methods achieve 98.75% of recognition accuracy rate. This approach performs well compared to other component-based facial recognition approaches.

Acknowledgements

All praise and glory be to God, the Most Gracious and the Most Merciful, for giving me directions, opportunities and ease in life. I am thankful to the University of KwaZulu-Natal for its good research facilities and working environment provided to me. This enabled me to enrich my knowledge and carry out my research. Similarly, I would like to express gratitude to the Council for Scientific and Industrial Research for funding my research work that helped in expanding the breadth of my knowledge to the next level. My appreciation and heartfelt gratitude goes to my thesis advisors Professor Serestina Viriri and Mr Mandlenkosi Gwetu for their constant support, guidance and the time they devoted for my research work. I have learned so much from them. They will remain a source of inspiration to me in my career. I give big thanks to all my lab mates. I always found them very respectful, friendly and helpful. Finally, I would like mention my father, whom I see as my guide in life. He has always supported me with love, affection and advice, which keeps me going. From the bottom of my heart I appreciate my mother, father, sisters and all the loving family members for their support, patience and encouragement they gave me in pursuing my career.

Contents

List of Tables

List of Figures

List of Abbreviations

AFR	**A**utomatic **F**acial **R**ecognition
AI	**A**rtificial **I**ntelligence
ANN	**A**rtificial **N**eural **N**etwork
CE	**C**ontrast **E**nhancement
CMU	**C**arnegie **M**ellon **U**niversity
CNN	**C**onvolutionary **N**eural **N**etwork
CV	**C**omputer **V**ision
CV	**C**omputer **V**ision **S**ystem **T**oolbox
DOG	**D**istance **O**f **G**aussian
EBGM	**E**lastic **B**uch **G**raph **M**atching
ECOC	**E**rror **C**orrecting **O**utput **C**odes
FD	**F**ace **D**etection
FERET	**F**acial **R**ecognition **T**echnology
FEI	**F**rey **E**evel **I**o-occurrence
FRT	**F**acial **R**ecognition **T**echnology
FR	**F**acial **R**ecognition
FRS	**F**acial **R**ecognition **S**ystem
GLCM	**G**rey **L**evel **C**o-occurrence **M**atrix
HE	**H**istogram **E**qualization
IP	**I**mage **P**rocessing
LBP	**L**ocal **B**inary **P**atterns
NN	**N**eural **N**etwork
PCA	**P**rincipal **C**omponent **A**nalysis
PLS	**P**atial **L**east **S**quare
LDA	**L**inear **D**iscriminant **A**nalysis
SCface	**S**urveillance **C**ameras **F**ace
SIFT	**S**cale **I**nvariant **F**eature **T**ransforms
SURF	**S**peeded-**U**p **R**obust **F**eatures
SSA	**S**equential **S**Selections **A**lgorithms
SFS	**S**equential **F**orward **S**Selection
SVM	**S**upport **V**ector **M**achines

U-SIFT **U**p-right **S**cale **I**nvariant **F**eature **T**ransforms

Chapter 1

Introduction

1.1 Motivation

Currently, modern technology is shifting the focus to biometric traits and the convenience of prevailing low cost embedded systems has created a massive interest in automatic processing of digital images. Facial images have numerous advantages over other biometric modalities such as signature, iris and fingerprint recognition. Besides being natural and non-intrusive, the most important advantage of facial image acquisition is that it can be done at a distance and in a concealed manner.

There are many factors associated with Facial recognition (FR) and these include great variability in head rotation and tilt, lighting intensity and angle, facial expression and ageing. Recent research has shown that the three main factors in facial recognition that have remained unsolved, are illumination, pose and occlusion [1]. There have been numerous approaches proposed by various research studies to overcome these challenges. Among them, holistic, appearance and hybrid approaches appear to be common in the literature. Currently, hybrid approaches are the most prevalent methods in FR research. They combine both holistic and appearance based approaches to overcome the shortcomings of individual methods [2] [3].

This dissertation proposes a hybrid component-based model to recognise faces at component level. The main objective in using facial components is to compensate for pose changes and allow flexible geometrical relations between the components in the classification stage. The proposed model seeks to utilise successfully detected components to recognise and verify the personâĂŹs identity.

1.2 Problem Statement

Recognising faces from the frontal view and controlled illuminating conditions is a reasonably well solved problem [4]. However, challenges arise from high variability in a head tilt, lighting intensity, facial expression, and ageing. In general, pictures of the same person are taken at separate times with variations in pose, lighting, background and with the presence of accessories such as glasses, hairstyle and earrings. In the light of these considerations, can recognising individual facial components, such as forehead, eyes, nose, cheeks, mouth and chin, enhance facial recognition accuracy? If so:

- Which facial landmarks and components are most discriminative for facial recognition?
- What features can adequately represent the identified facial landmarks?
- How useful is the proposed model regarding variations in factors such as pose, illumination, expression and affine transformation?

1.3 Research Objectives

The objectives of this research are:

- To improve facial recognition using a facial component-based approach.
- To compare holistic feature-based and component feature-based techniques for facial recognition.
- To model a framework for component-based face recognition.

1.4 Contributions of the Dissertation

Achieving an accepted degree of accuracy remains a challenge in FR. The research aims to improve accuracy in facial recognition by recognising faces at component level. In the light of these objectives the key contributions of this dissertation are detection and recognising of faces using facial components as primary features. This is accomplished by extending algorithm functionalities such as Viola and Jones Cascade detector to further detect and recognise additional components such as the forehead, cheeks and chin.

The number of features at the disposal of the classification system is usually substantial, hence this dissertation proposes a wrapper-based approach from Sequential Selections Algorithms (SSA) to select salient features from a pool of local features. This approach is inspired by the study developed by Kittler et al. [5]. The model is a bottom-up search technique which starts with an empty feature set S and gradually adds chosen features to an evaluation function, which minimises the classification error. The features are classified with a minimum classification error. At each iteration, the feature to be included in the feature set is selected among the remaining available features of the feature set, which have not yet been added.

1.5 Assumptions and Limitations

This research assumes that all experiments are conducted using images of faces only. Scenarios where there are clustered scenes are ignored and some experiments using these were not the primary focus. Artificial intelligence concepts were used, but they were not profoundly visited because this study aimed to get an idea about the fundamentals of facial recognition technology as a whole. Hence, advanced techniques such as Convolutional Neural Networks (CNN) were ignored although the face research community is shifting towards the era of CNN. The idea is to consider pure artificial intelligence in facial recognition in future work of this research.

1.6 Structure and Scope of the Dissertation

The thesis is organised into the following chapters:

- **Chapter 1** introduces the research and addresses the problem area that this dissertation is centred on.
- **Chapter 2** reviews prior facial recognition methods and evaluates their application by previous studies. It also justifies why the hybrid component-based approach is reliable.
- **Chapter 3** describes methods employed by this dissertation and provides the implementation details of the system.
- **Chapter 4** discusses the main findings of this research.
- **Chapter 5** concludes this work and defines possible future work.

Chapter 2

An Overview of Facial Recognition

2.1 Introduction

This chapter provides an overview of Facial recognition and a survey of prior research; including theories of how FR works; and discusses various approaches to FR and their applications. The rest of the chapter is organised as follows: Section 2.2 discusses the history of FR; Section 2.3 presents the literature review; Section 2.4 discusses the face recognition algorithms and applications; Section 2.5 concludes the chapter.

2.2 Background History

There are several reasons for the growing interest in automated face recognition, including rising concerns for public security, the need for identity verification for physical and logistical access, and the need for face analysis and modelling techniques in multimedia data management and digital entertainment. Research in automatic face recognition started in the 1960s and over the years it has made significant progress. Between 1964 and 1965, Bledsoe and others worked on using the computer to recognise human faces [6]. However, they published little academic work.

In 1973, Takeo Kanade developed the first Automatic Facial Recognition (AFR) system to perform face recognition without human intervention [7]. The system used three facial components, the eyes, nose and mouth, to detect and recognise human faces. A decade later, in 1987, Sirovich and Kirby introduced Principal Component Analysis (PCA). This is a statistical method that reduces the dimensionality of data for efficient preprocessing [4] [8] [9].

In 1990, Sirovich and Kirby revised PCA to improve the recognition efficiency and perform face recognition in a small dimensional face representation to reduce computation effort [8]. In 1991, Turk and Pentland introduced the practical applications of PCA, named the Eigenface; to represent pictures efficiently using PCA [8] [9] [10]. They used Eigenface to classify faces from non-facial objects, that is, the background and other non-facial objects. Eigenface pioneered the FR research. In 1996 Etemad and Challepa introduced Linear Discriminant Analysis (LDA), a broader view of the Fisher Linear Discriminant method, to achieve high recognition accuracy [4] [11].

In 2001, Viola and Jones introduced a learning-based face classifier; named Adaboost for real-time face processing, such as face detection and recognition [4] [12]. The classifier worked with the local features, the eyes, nose, and mouth [13]. Table 2.1 provides a summary of the methods discussed above. These methods have pioneered facial recognition, and particularly Eigenface has been the major milestone that reinvigorated FR research. Currently, many FR approaches and techniques use these concepts.

TABLE 2.1: Legacy face recognition methods

Method	**Author**	**Year**
First Automated FR	Kanade	1973
Principal Component Analysis	Sirivich & Kirby	1987
Eigenfaces	Turk & Pentland	1991
Fisherface	Etemand & Challeppa	1996
Adaboost + Haar Cascade	Viola & Jones	2001
Gabor Jets	Naruniec & Skarbek	2007

2.3 Literature Review

Facial recognition algorithms are categorised into (i) appearance-based (holistic), (ii) feature-based, and (iii) hybrid approaches and this taxonomy is widely accepted and also applies to face detection, localisation and verification algorithms [14].

Holistic, or appearance-based approaches, perform well on images with frontal view faces and they are characterised by the use of the whole face image for recognition. However, they are computationally expensive as they require dimensionality reduction algorithms such as Eigenface techniques to reduce the computation effort by representing the matching of faces by the PCA [10].

Feature-based approaches are much faster and robust against face recognition challenges. They use purely geometric methods that extract local features from facial landmarks [15]. Feature-based and hybrid approaches further divide into: (1) generic methods based on face image features, (2) feature-template methods that detect specific facial features and (3) structural matching methods that consider geometrical constraints on the facial features. Feature-based approaches make use of the face components such as nose, eyes or mouth for recognition. Some algorithms of this category include pure geometry, dynamic link architecture, Hidden Markov Model, elastic bunch graph matching and local feature analysis.

Hybrid approaches have been prominent in this rapidly expanding research field and they have the potential to improve recognition rate since they combine different methods to overcome the individual-method shortcomings [1]. They combine holistic and feature-based approaches to overcome the shortcomings of the two methods and give more robust performance. Recently, they have shown promising results in various object detection and recognition tasks such as face detection and face recognition. Also, they compensate for pose changes and allow flexible geometrical relations among the face components in the classification stage. Hybrid approaches use both local regions and the whole face. Modular Eigenface, hybrid local functions, and shape-normalized methods also belong to this category [16] [17].

Dargham et al. [18] proposed a hybrid component-based face recognition system that recognises faces using the three main facial components; eyes, nose and mouth. The system is dysfunctional on faces rotated 45° from frontal view. The face contains rich information, focusing only on these three facial components might not be ideal in low lighting conditions.The proposed approach in this research uses LDA to extract the feature from each component. Our hybrid component-based face recognition approach mainly focuses on utilising any successfully detected component to identify a person from images.

Mandal et al. [19] proposed a hybrid approach that combines the structural features with holistic features. Their approach combines these two strategies to capture every detail of the face. Srinivasa et al. [20] proposed a system considered as a hybrid approach, as it combines the Viola Jones algorithm and SURF method to extracts features from the three main facial components of eyes, nose and mouth, using SURF patented local feature descriptor.

In [21], the authors present a component-based Support Vector Machines (SVM) classification and morphable model that are invariant to pose and illumination. When a model is rendered under varying pose and illumination conditions, it creates a vast number of synthetic face images which are used to train a component-based face recognition system. These synthetic faces are captured in various angles of face poses and lighting conditions. Recent research [22] [23] has combined techniques to improve the rate of face recognition. However, the issue with combining techniques is that the required amount of data for training the algorithms is huge.

In [12], a component-based face recognition system is presented that employs a two-level SVM It employs two-level SVM to detect and validate facial components. In this system, learned face images are automatically extracted from 3-D head models that provide the expected positions of the components. The majority of facial recognition systems are hybrid-based that formulate new strategies to improve on existing FR systems and are all based on PCA decomposition, which reduces higher dimensional training data of face images to preserve the original information about the face. Numerous face recognition algorithms fall into this category. Eigenface, Fisher-faces and SVM are the most popular holistic algorithms.

Several studies [22] [25] [26] have used many well-known methods such as the Artificial Neural Networks (ANN) and Local Binary Patterns (LBP) to develop robust FR models. ANNs is a machine-learning technique that makes decisions after a series of trainings using specific data similar to that of the targeted output. However, the drawback of ANNs is that they require a significant amount of training to obtain good results.

In a study by [22], a method was proposed that built an ANN-based FR system. The method utilised the facial components, the mouth, nose and eye facial components. These were used to train the classifier for later classifications stages. Although the systemâĂŹs models worked successfully, they failed to achieve high-level recognition accuracy on images with significant variations in illumination and occlusion [26]. Many FR systems follow a set of routine steps. Usually, they consist of four phases, as shown in Figure 2.1: face detection (localisation), face preprocessing (face alignment/normalisation, light correction), feature extraction and feature matching.

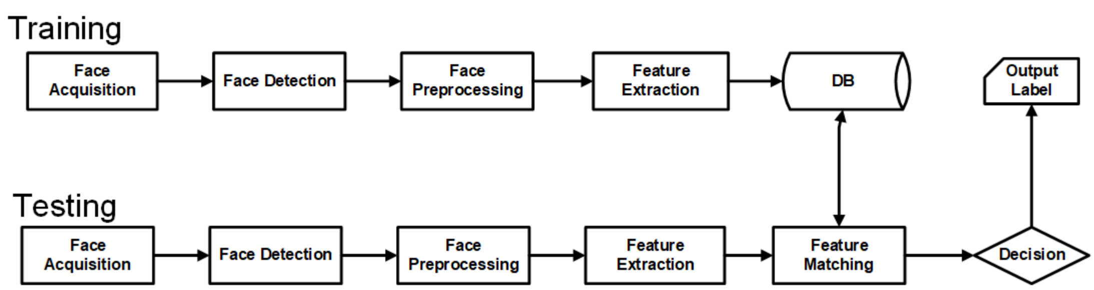


FIGURE 2.1: Face recognition pipeline [27].

Component-based face recognition studies are not frequently found in the literature in the literature. Even methods which compute similarity measures at specific facial landmarks, such as Elastic-Bunch Graph Matching (EBGM) [28], do not operate in a per-component manner. This work focuses on face recognition at component level. The approaches discussed above mainly rely on combining algorithms and methods for facial recognition. It has been established that the face has a lot of information that can be utilised for recognition, and rejecting certain parts from it has an influence on recognition accuracy. In this research a robust hybrid component-based strategy for facial recognition is proposed. The strategy seeks to utilise any successfully detected components to recognise and verify the identity of a person. The next section will look at various applications of techniques in relation to the work of other studies.

2.4 Related Techniques and their Applications

There are different techniques of facial recognition and this section will discuss them in detail and provide an overview of their implementation.

2.4.1 Holistic, Local and Hybrid-based Techniques

Among this class of techniques, the holistic is comprehensive and uses the whole face for identification. An example of holistic methods is PCA; more on PCA in section 2.4.2.1. Local techniques such as LBP use local facial features for face recognition. Whereas hybrid-based techniques combine both comprehensive and local techniques to overcome their shortcomings. Standalone hybrid methods use 3D images to allow the system to regard the curves of the eye sockets, the shapes of the chin and the forehead. Even a face in profile would serve because the system uses depth and an axis of measurement. In that way, the system gets enough information to construct a full facial look.

The 3D system usually processes as follows:: Detect, Locate, Scale and Measure, Represent and Match. (1) Detection captures the face by either scanning a photograph or obtaining the person's face from a feed in real time. (2) Localising determines the location, size and angle of the head. (3) Scaling and Measurements measure each curve of the face to make a template with the particular focus on both the internal and external of the eye and the angle of the nose. (4) Representation converts the model into a code, a numerical illustration of the face and Matching compares the derived data with faces in the existent face dataset. In the case where the 3D image is to be compared with an existing 3D model, it needs to have no alterations. Typically, however, photos are in 2D, and in that case, the 3D image needs a few changes. This practice is delicate and is one of the most significant challenges in the field today.

2.4.2 Appearance-based Techniques

Appearance-based techniques rely on statistical techniques such as Partial Least Square (PLS) to compare the sample image with the stored prototypes in the features space. They operate directly on an image-based representation, and they strictly consider the entire face region as the input into an FR system. Appearance-based approaches rely on statistical methods and machine learning algorithms. Their goal is to find related characteristics from the face image that are in the form of a distribution model [29].

Conventional methods implemented using this approach are PCA and LDA [8] [9]. Below is an overview of the methods. They are described better in a statistical framework. A feature vector derived from the image can be seen as a random variable x that represents the face or a non-face class that defines the conditional density function deifned in equation (2.1) and (2.2). A typical example where this application is valid is the Bayesian classification that uses facial images to classify facial images for the candidate as faces or non-faces.

$$p(x \mid face) \tag{2.1}$$

$$p(x \mid non-face) \tag{2.2}$$

2.4.2.1 Principal Component Analysis

Principal Component Analysis constructs an optimal face subspace to represent only the face object. It transforms images into a small set of attribute feature images, called Eigenface [8] [9]. Eigenface represent the face as weighted vector features in a subspace derived from training images.

The Eigenvector and Eigenvalues occur in pairs, that is, for every Eigenvector, there is a corresponding eigenvalue. The former represents the length and the direction whereas; the latter represents magnitude, that is, a number that measures how much variance is in that particular direction. If an Eigenvector contains a high Eigenvalue, it is regarded as a primary component [23].

However, PCA can be used to reduce dimensions. Presenting discriminative feature vectors amongst the face images is insufficient. Also, though it helps to represent the data in the most compact form, the most characteristic features are not always the best discriminitive features. Therefore, Linear Discriminative Analysis (LDA) solves this problem by providing discriminitive information among faces with the requirement of larger samples during the training process.

2.4.2.2 Linear Discriminant Analysis

Linear Discriminant Analysis is a better substitute for the PCA. Instead of paying attention to the underlying structure it provides discrimination among the classes, unlike the PCA which deals with the input data in their entities [30]. Linear Discriminant Analysis aims to find a base of vectors that provide the best discrimination amongst the classes, while it maximises the between-class differences which tend to minimise the within-class ones. The underlying problem in LDA is the within-class scatter matrix, which is singular most of the time because of the more significant dimension of image pattern compared to the number of training images.

2.4.3 Feature Matching-Based Techniques

Feature-based techniques extract features from eyes, lips, nose and mouth with their locations and geometric shape and inject them into a structural classifier. Kanade developed one of the earliest face recognition algorithms based on automatic facial feature detection [7]. By localising the corner of the eyes, nostrils and other details of features in frontal views, that system compares parameters for each face analysed using Euclidean distance metric against the parameters of known personâĂŹs faces.

Elastic Bunch Graph Matching (EBGM) approach is an example of the feature-based approaches, and it has appeared in many applications [31]. Other well-known approaches in these systems are Hidden Markov Model (HMM) and convolution neural network [17]. A method based on an EGBM approach has been applied to face detection and extraction, pose estimation, gender classification, sketch image-based recognition and general object recognition.

Feature invariant approaches rely on the structural features that barely change on a face regardless of the pose, illumination conditions and facial profile, that is, the viewpoint. Researchers have made numerous assumptions about human being able to detect and recognise the face from various poses and low illuminating conditions. As a result, their conclusions are that there must exist features or properties that are invariant over these circumstances. Although there is much work done on facial detection approaches, there are problems associated with element invariant methods. Such image features become corrupt because of the noise, occlusion and illumination conditions. However, feature invariant approaches have a high capability in face localisation and they achieve significant results [29].

Facial Detection (FD) is a relevant research field that has numerous challenges associated with it. It is also vital for any Facial Recognition System (FRS), as it plays a crucial role in the detection of the face and facial components.

2.4.4 Template Matching and Neural Networks

Template matching uses models, pixels, curves and sometimes texture to describe patterns for recognition and the recognition function is the distance measure. Since Template matching uses statistical methods, it represents models as features and the identification is a discriminant function. For neural networks, pattern representation varies, although there is always network function at some point. Pattern representations are diverse standard patterns of a face to describe the entire look and facial components independently.

Template matching methods compute the correlation between the input face image and the pattern describing the face. The relationship between the input image and the stored model is calculated for both face detection and facial feature detection. Attempts made by using Template Matching Methods have made sub-templates for each facial component such as eyes, mouth, nose and the contours of the face to model the face. Template matching models are easy to implement, although they cannot detect faces efficiently due to variations in pose, scale, and shape. However, they apply to both face localisation and detection phases in FRS [32].

Sakai et al. made the first attempt at face detection using template matching methods to detect faces from images [33]. The idea was to construct sub-templates for eyes, nose, mouth and the temple to model the face. Each sub-template is a line of segments, and each line was compared and matched against the sub-template that corresponded with it. This way allowed them to compute the correlation between the sub-images and contour templates to match with other sub-templates.

2.5 Conclusion

Face recognition technology has made impressive gains, but it is still not able to meet the accuracy requirements of many applications. A sustained collaborative effort is needed to address many of the open problems in face recognition, such as illumination, occlusion and facial pose. The next chapter will outline the methodology used in this study to fulfil the aims of this work to achieve the primary objectives.

Chapter 3

Methods and Techniques

3.1 Introduction

This chapter discusses the high-level representation of the methodology and it is depicted in Figure 3.1. It incorporates several steps followed to achieve a component-based facial recognition model. These are database acquisition and cross validation that is splitting the database into training and testing sets using the hold-out cross validation method; pre-processing using three state of the art techniques to remove noise from images; Gamma Correction, Difference of Gaussian (DOG) and Contrast Enhancement; and convert photo-based features from facial components using texture, shape and size attributes. Five texture and shape descriptors were used to extract distinctive features from various levels of the face. Prominent features from facial components were selected and normalised into homogeneous domain and later classification and recognition of faces was done using the Error-Correcting Output Classifier (ECOC). The rest of this chapter is organised as follows: Section 3.2 gives an outline of the high-level end methodology that defines the hierarchy and the application of methods employed to implement the proposed technique; Section 3.3 concludes the chapter.

3.2 Methodology

The following is a high-level representation of the methodology used in this study to develop the model architecture of this research.

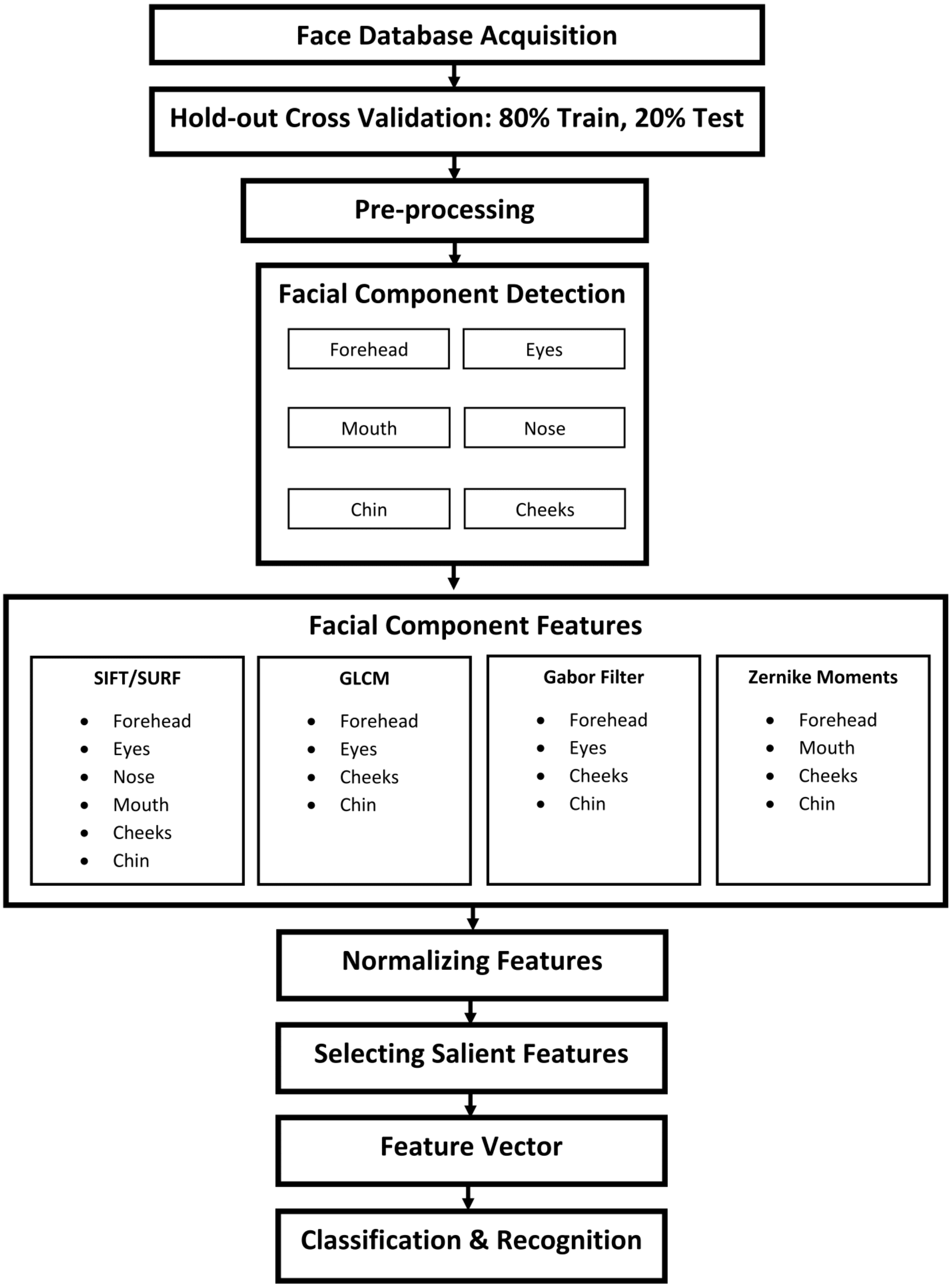


FIGURE 3.1: The representational structure of the methodology

3.2.1 Face Image Database Acquisition

This study considered four state of the art facial databases, FERET, CMU, SCFace and FEI which were acquired from their affiliates for research purposes.

3.2.2 Cross Validation

To select the training and testing sets, the holdout cross-validation technique was employed to split the face datasets into 80% training and 20% testing sets.

3.2.3 Pre-processing

Three processing techniques were employed and combined: Gamma Correction, Difference of Gaussian (DOG) filter, and Contrast equalization [34] [35] [36]. These techniques have a significantly high discriminative level of pre-processing the image by enhancing the image contrast, reducing noise and removing dark patches on the image. The Gamma correction method improves shadowed regions on the face image by scaling the pixel intensities from $[0, 255]$ to $[0, 1.0]$. The filter defined in equation (3.1) filters the image to produce an image with improved brightness called gamma corrected image.

$$I_{output} = I_{input}^{1/\gamma} \begin{cases} \text{if } G < 1, \text{shift towards dark.} \\ \text{if } G > 1, \text{shift towards light.} \\ \text{if } G = 0, \text{no effect.} \end{cases} \tag{3.1}$$

In the filter (3.1), the effect of $G < 1$ shifts the image towards a darker spectrum, while $G > 1$ shifts towards lighter and $G = 1$ has no effect. Gamma correction does not remove all shading effects; hence, DOG is used to overcome this factor. The DOG filter, defined by equation (3.2), removes shadowing effects and suppresses high-frequency spatial information that are present in the image. Figure 3.2 shows a sequence of face images before and after pre-processing using the Gamma filter defined by equation (3.1) and the DOG filter defined as (3.2).

$$DOG_{(x,y)} = \frac{1}{2\pi\sigma_1^2} e^{-\frac{x^2+y^2}{2\pi\sigma_1^2}} - \frac{1}{2\pi\sigma_2^2} e^{-\frac{x^2+y^2}{2\pi\sigma_2^2}}. \tag{3.2}$$

DOG helps to increase the visibility of edges to make it easier to detect and locate the face and the facial components. However, it reduces the overall contrast

(A) (B) (C)

FIGURE 3.2: Images pre-processed with Gamma Correction: (A) Original image, (B) Gamma corrected image, (C) DOG filtered image

of an image; hence, the difference is enhanced in subsequent stages, using the Contrast Enhancement (CE). CE adjusts pixel intensities to standardise the overall intensity variations by normalising the histogram of discrete grey values of the image, which are aligned using the Histogram Equalisation (HE) technique defined by equations (3.3) and (3.4).

$$p_i = \frac{n_i}{n_t}. \tag{3.3}$$

where n is the number of pixels with intensity i and n_t is the total number of pixels. The image I with adjusted intensities is defined by (3.4)

$$I_{(i,j)} = floor((L-1)\sum_{n=0}^{f_{(i,j)}} p_n). \tag{3.4}$$

where $floor$ rounds down to the nearest integer. The Histogram Equalisation has simplified the detection and recognition process in low lighting conditions. Figure 3.3 shows the results after the application of the three pre-processing techniques.

(A) (B) (C)

FIGURE 3.3: Post-processing on face images using CE and HE: (A) DOG filtered image, (B) Standardised intensities using CE, (C) Image post-processing using HE

3.2.4 Facial Components Detection

In component-based facial recognition, the most challenging task is to locate the components from the face. The Viola and Jones [12] [37] algorithm is one of the powerful algorithms that performs this. Although it does not cover all the components. A cascade detector developed to use this algorithm is capable of detecting the eyes, nose and the mouth. However, we have further trained the cascade detector to detect the cheeks, chin and forehead per-component. These three additional components are considered to be considered to be distinguishing components for our facial recognition model. Figure 3.4 depicts eight detected components from the face.

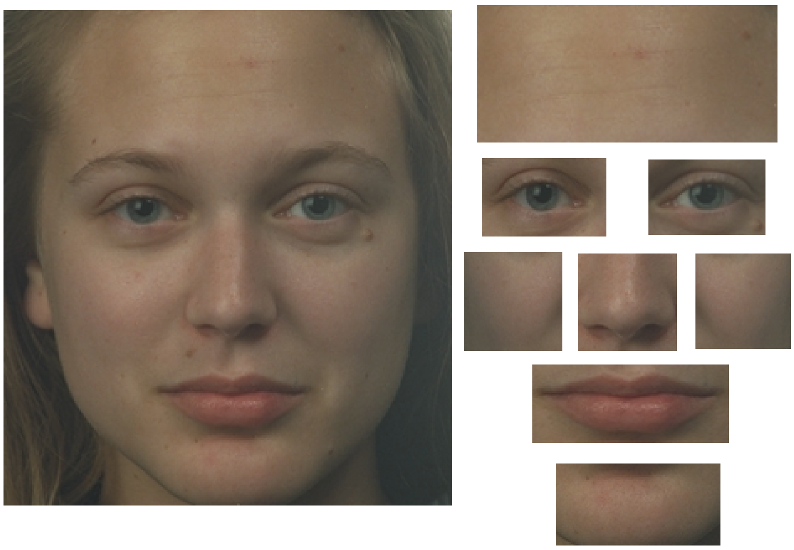


FIGURE 3.4: Eight facial components detected individually

3.2.5 Facial Component Features

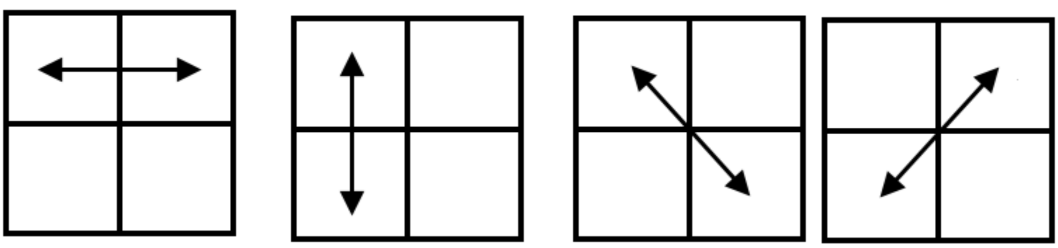
Using texture-based analysis, together with an appropriate filter, each photo in the training set is converted into a feature set. An exploration of previous literature has identified several successful filters that could be used to generate features from the image automatically. An investigation to find the most appropriate filter for the hybrid component-based face recognition was conducted. Five feature descriptors were explored and utilised for their discriminative power to compute salient features from face images. A detailed analysis of these descriptors is provided below.

3.2.5.1 Grey-Level Co-occurrence Matrix (GLCM)

GLCM, also known as the grey-level spatial dependence matrix is a statistical method of examining texture that considers the spatial relationship of pixels. It has several functions to characterize the texture of an image [38] [39] [40]. These functions calculate how often the pair of pixels with specific values and in a specified spatial relationship occur in an image. The basis for these features computed by is the GLCM G (3.5). This matrix G is a square matrix of dimension N_g, where N_g is the number of grey levels in the image. Element $[i, j]$ of the matrix is generated by counting the number of times a pixel with value i is adjacent to a pixel with value j from four directions 0°, 90°,45°,135° and then dividing the entire matrix by the total number of such comparisons made. Each entry is therefore considered to be the probability that a pixel with value i will be found adjacent to a pixel value j.

$$G = \begin{bmatrix} P(1,1) & P(1,2) & \dots & P(1,N_g) \\ P(2,1) & P(2,2) & \dots & P(2,N_g) \\ \vdots & \vdots & \dots & \vdots \\ P(N_g,1) & P(N_g,2) & \dots & P(N_g,N_g) \end{bmatrix} \tag{3.5}$$

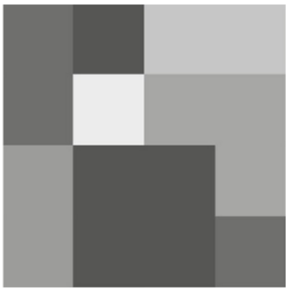
To utilise texture features to their fullest using the GLCM method, a single matrix is not enough to describe the texture. Hence, each facial component has multiple GLCMs, consisting an array of offsets defining the relationship between the pixels from different directions representing, horizontal, vertical, and two diagonals which correspond to 0°, 90°,45°,135°. The offsets are a p-by-2 array of integers, with each row describing a two-element vector, $[\text{row-}\vec{\text{offset}}, \text{col-}\vec{\text{offset}}]$, that specifies one offset. Row-offset are rows between the pixel of interest and its neighbour pixels, and col-offset are columns between the pixel and its neighbour pixels. One neighbouring pixel in the possible four directions is defined as $[0, 1], [-1, 1], [-1, 0], [-1, -1]$, as shown in Figure 3.5.



(A) (B) (C) (D)

FIGURE 3.5: Four directions of adjacency matrix: (A) 0°, (B) 90°, (C) 45°, (D) 135° defined for calculating texture features

The GLCM can reveal specific properties about the spatial distribution of the grey levels in the texture image. For example, if most of the entries in the GLCM are concentrated along the diagonal, the texture is coarse concerning the specified offset. Figure 3.6 shows the calculation of the values of the GLCM.



Grey Scale Image

3	1	6	6
3	8	5	5
5	1	1	5
5	1	1	3

Pixel Intensities

	1	2	3	4	5	6	7	8
1	2	0	1	0	1	1	0	0
2	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0
5	2	0	0	0	1	0	0	0
6	0	0	0	0	0	1	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0

Co-occerence Matrix

FIGURE 3.6: The spatial co-occurrence calculation

In the output GLCM, element $(1,1)$ contains the value 2 because there are two instances in the input image where two horizontally adjacent pixels have the values 1 and 1. Respectively, $(1,2)$ in the GLCM has the value 0 because there are no instances of two horizontally adjacent pixels with the values 1 and 2. Element $(1,3)$ contains the value 1 because there is one instance where two horizontally adjacent pixels have the values 1 and 3. There are eight functions of the GLCM and this work employs only three of them: Energy (3.6), Contrast (3.7) and Entropy (3.8) to extract the texture feature from the facial components. Energy provides the sum of squared elements in the GLCM; Contrast known as Inertia, measures the local variations in the grey-level co-occurrence matrix; Correlation is the joint probability occurrence of the specified pixel pairs and Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$Energy, E = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [p(x,y)^2]. \tag{3.6}$$

$$Contrast, I = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i-j)^2 p(i,j). \tag{3.7}$$

$$Entropy, S = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i,j) \log[p(x,y)]. \tag{3.8}$$

Since rotation invariance is a primary criterion for any features used with these images, a kind of invariance was achieved for each of these statistics by averaging them over the four directional co-occurrence matrices. The maximal correlation coefficient was not calculated due to computational instability, so there were 13 texture features of each image.

3.2.5.2 Zernike Moments

Zernike Moments are a sequence of polynomials that are orthogonal on the unit disk [41] [15] [19]. The two-dimensional Zernike Moments, $A_{n,m}$ of order n with m repetition of image $I(p,\theta)$ are given by:

$$A_{pq} = \frac{n+1}{\pi} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(p,\theta) \times V_{pq}(p,\theta), p \leq 1. \tag{3.9}$$

where:

- (p,θ) is a polar coordinate,
- V_{pq} is a complex conjugate,
- $p = \sqrt{x^2+y^2}$ and $\theta = \arctan(y/x)$,
- V_{pq} is a complex polynomial defined inside a unit circle with the formula:

$$V_{pq}(p,\theta) = R_{pq}(p)\exp(jm\theta). \tag{3.10}$$

where:

- $p \leq 1$ and $j = \sqrt{-1}$ are imaginary units.

$R_{pq}(p)$ is a radial polynomial, which can be generated using:

$$R_{pq}(p) = \sum_{s=0}^{n-|m|/2} (-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-2)!} p^n - 2s. \tag{3.11}$$

where:

- n is a positive integer,
- $n-|m|$ is even, $|m| \leq n$.

To reconstruct the original image $I(i,j)$ from the calculated Zernike Moments features, this function is employed:

$$I\prime(i,j) = \sum_{p=0}^{M} \sum_{q=0}^{N} A_{pq} V_{pq}(p,\theta). \tag{3.12}$$

Zernike Moments themselves are complex numbers and are sensitive to rotation of the image, hence their magnitudes are used as features. Figure 3.7 shows the nose and the mouth rotated by five different angles and Table 3.1 lists the set of Zernike Moments features for these components shown in Figure 3.7 with order 8.

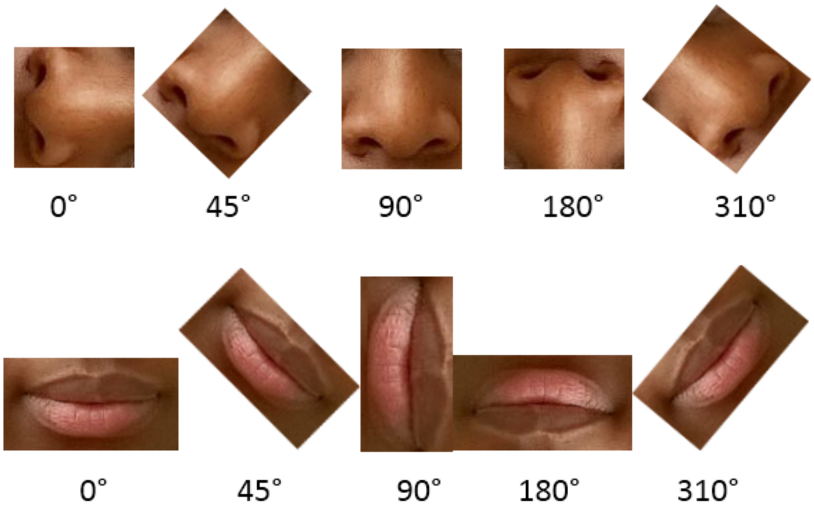


FIGURE 3.7: The nose and the mouth rotated in five different orientations

3.2.5.3 Gabor Filter

Gabor filter is a linear filter and texture descriptor to analyse whether there is any specific frequency content in the image in specific directions in a localised region around the point. Gabor features are computed by convolving the extracted components of interest $\epsilon(x, y)$, that is, forehead, eyes, cheeks and the chin, with the filter in equation (3.13) and produces corresponding response images r_ξ. The response images are computed for a bank of filters tuned onto various frequencies and orientations. The resultant Gabor feature thus consists of the convolution results of an input image $\xi(x, y)$.

$$\psi = e^{-\alpha^2(t-t_0)^2} e^j 2\pi f_0 t + \varphi. \tag{3.13}$$

A feature matrix G is defined by

$$G_{m,n}(x, y) = \frac{1}{M \times N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(x - m, y - n)\xi(m, n). \tag{3.14}$$

TABLE 3.1: Zernike Moments of order 8^{th} of five facial components

Orders		Orientations				
n	**m**	**0°**	**45°**	**90°**	**180°**	**310°**
0	0	33.2286	33.2611	33.2369	33.3263	32.6404
1	1	1.56949	1.48822	1.67472	1.84001	2.28814
2	0	77.0796	77.1345	77.0792	77.1650	76.0767
2	2	2.82715	2.86862	2.87631	2.99295	2.62875
3	1	4.64687	4.32835	4.79714	5.19414	6.42867
3	3	0.26162	0.25764	0.29876	0.33251	0.40200
4	0	71.9322	71.9486	71.8903	71.7254	71.8438
4	2	10.1105	10.2382	10.2614	10.6439	9.41514
4	4	0.47304	0.46720	0.45730	0.45247	0.43086
5	1	5.94619	5.27685	5.66541	5.95075	7.26442
5	3	1.02323	1.01829	1.18405	1.31612	1.60699
5	5	0.08196	0.04754	0.04454	0.03829	0.03251
6	0	29.9166	29.9283	29.8891	29.6276	31.0563
6	2	16.2318	16.3676	16.3866	16.8999	15.1456
6	4	1.99486	1.97107	1.93717	1.91590	1.87229
6	6	0.24819	0.24992	0.24386	0.25439	0.21876
7	1	3.61672	2.98167	3.02594	3.09192	3.56448
7	3	1.76426	1.72502	2.01277	2.22251	2.76925
7	5	0.39047	0.24773	0.24401	0.21233	0.19461
7	7	0.02459	0.02769	0.03276	0.04053	0.05486

where ξ is the filter mask of $m \times n$ and G_{mn} the matrix of Gabor coefficients of the same size as the image $\xi(x, y)$. The image is convoluted with the filter to produce a response image, and the Gabor features are formed by combining responses of several filters from a single to multiple spatial location. Figure 3.8 shows the response image produced by the filter after convolving the face component and Figure 3.9 shows the response image for both eyes. In equation (3.14), α is the sharpness (time duration and bandwidth) of the Gaussian, t_0 is the time shift defining the time location of the Gaussian, f_0 is the frequency of the harmonic oscillations (frequency location), and ψ denotes the phase shift of the oscillation.

$$\psi(f) = \sqrt{\frac{\pi}{\alpha^2}} e^{-\left(\frac{\pi}{\alpha}\right)^2 (f-f_0)^2} e^{-j2\pi t_0 (f-f_0)+\phi}. \tag{3.15}$$

From (3.14) and (3.15) it can be seen that the Garbo function, or more precisely its magnitude, has the Gaussian form in the time domain and frequency domain. The Gaussian is located at t_0 in time and f_0 in frequency. If you increase the bandwidth α, the function will shrink in time, but stretch in frequency. In this research, we have

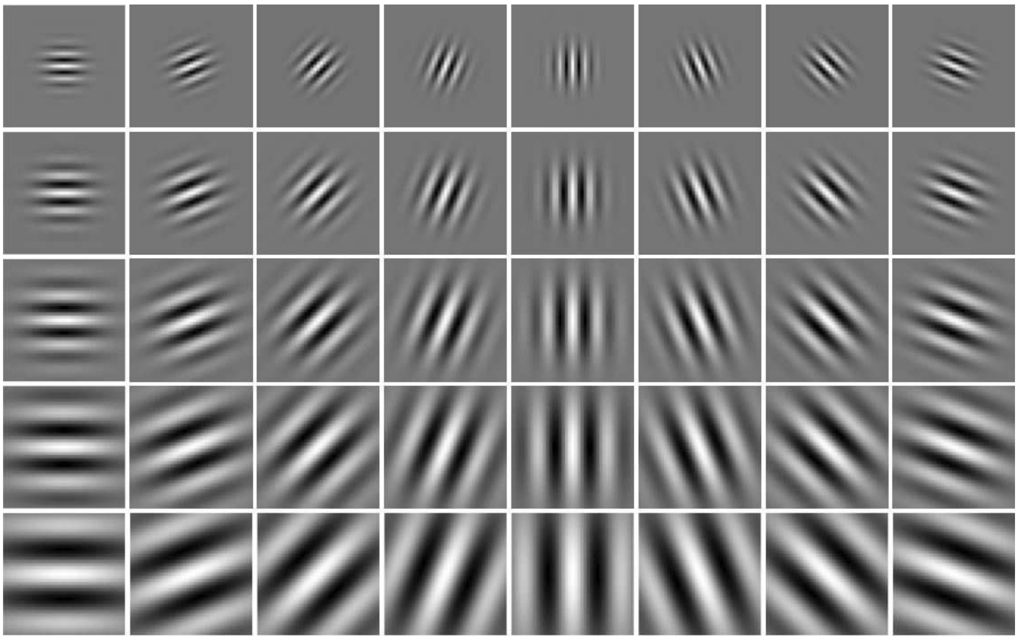


FIGURE 3.8: Response image $\xi(x, y)$, after convolving the filter in the component

used a 2D Gabor function defined as equation (3.16).

$$\psi(x, y) = e^{-\left(\alpha^2 x\prime^2, \beta^2 y\prime^2\right)} e^{j} 2\pi f_0 x\prime,$$
$$x\prime = x\cos\theta + y\sin\theta,$$
$$y\prime = -x\sin\theta + y\cos\theta \tag{3.16}$$

where the new parameters are β for sharpness and θ for its orientation. The 2D Gabor filter in the spatial domain as (3.17).

$$\psi(x, y) = f\frac{f^2}{\pi\gamma\eta} e^{-\left(\frac{f^2}{\gamma^2} x\prime^2 + \frac{f^2}{\eta^2} y\prime^2\right)} e^{j2\pi f x\prime} \tag{3.17}$$

The normalised 2D Gabor filter function has an analytical form in the frequency domain as defined in equations (3.18) and (3.19).

$$\psi(x, y) = e^{\frac{\pi^2}{f^2}\left(\gamma^2 (u\prime - f)^2 + \eta^2 v\prime^2\right)}$$
$$u\prime = u\cos\theta + v\sin\theta,$$
$$v\prime = -u\sin\theta + v\cos\theta. \tag{3.18}$$

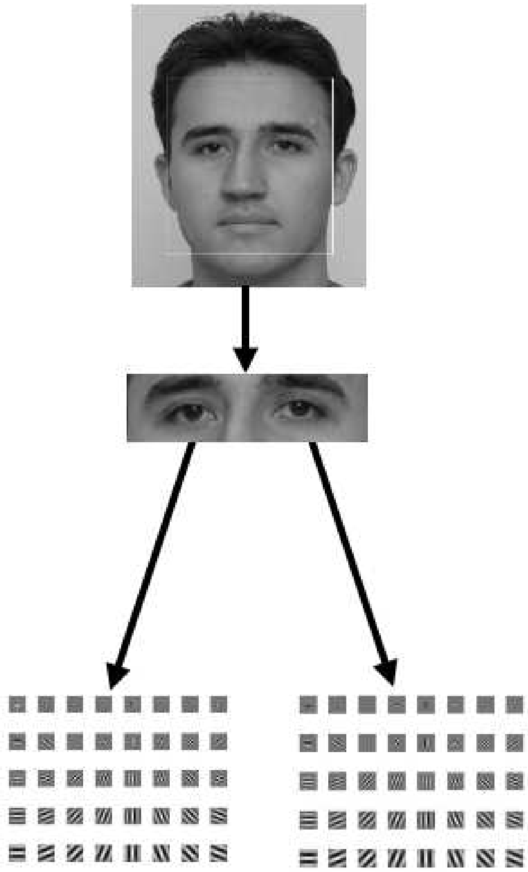


FIGURE 3.9: An example of a response image on both eyes (left and right).

$$f_k = c^{-k} f_{max}, \quad for \quad k = 0, \ldots, m-1 \tag{3.19}$$

where f_{max} is the maximum frequency and c is the frequency scaling factor. The filter orientations are spaced uniformly with scaling defined in equations (3.20) and (3.21).

$$\theta_k = \frac{k2\pi}{n}, \quad k = 0, \ldots, n-1. \tag{3.20}$$

For real signals the responses on $[\pi, 2\pi]$ are complex conjugates of responses on $[0, \pi]$ and therefore only the responses for the half plane are needed:

$$\theta_k = \frac{k\pi}{n}, \quad k = 0, \ldots, n-1. \tag{3.21}$$

3.2.5.4 Scale Invariant Feature Transform (SIFT) Texture Features

SIFT is popular in visual object categorisation and baseline matching [42] [43]. It can generate a vector of 128-dimensions that stores the gradients of pixel locations of size 4×4. It has been used in face recognition as a rotation invariant descriptor. However, some instances where the face is rotated at some angle greater than 15°, the descriptor results in false matching [44]. To overcome this problem an upright version of this descriptor (U-SIFT) is used [45]. SIFT works well in pose and expression invariant and since some images from the databases used in this work have different face poses SIFT captures visual points from the components. Table 3.2 lists the feature scores computed by 128-SIFT descriptor for all seven facial components with the corresponding total number of features.

TABLE 3.2: Feature scores of the SIFT descriptor for each component

Face Components	SIFT-128	Total Features
Forehead	$[-0.0081, \ldots, -0.0094]$	395
Eyes	$[-0.0019, \ldots, -0.0065]$	393
Nose	$[-0.0008, \ldots, -0.0100]$	387
Left cheek	$[-0.0064, \ldots, -0.0014]$	402
Right cheek	$[-0.0035, \ldots, -0.0036]$	386
Mouth	$[-0.0004, \ldots, -0.0121]$	391
Chin	$[-0.0004, \ldots, -0.0121]$	345

3.2.5.5 Speeded-up Robust Features (SURF) Texture Features

The SURF descriptor is a scale and rotation invariant detector and descriptor to detect key points. The 64-dimensional SURF focuses on the spatial distribution of gradient information within the interest point neighbourhoods. According to [45], when SURF is applied in face recognition, invariance rotation is often not necessary. This work therefore uses the upright version of the SURF descriptor to compute distinct interest point features. SURF stays more robust to various image perturbations than the more locally operating SIFT descriptor. In [45], they analysed this effect to observe if the SURF are useful in recognising faces under multiple illuminations and their analysis was positive in that SURF is helpful in facial recognition. The features scores represent the number of features from a complete set of the facial components from a single face. They show the number of potential features a unique face may have, and they are susceptible to noise and feature selection which is crucial in this case. Table 3.3 lists seven facial components with the corresponding set of facial features and the number of features computed by the SURF descriptor.

TABLE 3.3: Feature scores of the SURF descriptor for each component

Face Components	SURF-64	Total Features
Forehead	$[-0.0004, \ldots, -0.0017]$	502
Eyes	$[-0.0013, \ldots, -0.0151]$	483
Nose	$[-0.0018, \ldots, -0.0111]$	378
Left cheek	$[-0.1434, \ldots, -0.0139]$	502
Right cheek	$[-0.0117, \ldots, -0.0111]$	483
Mouth	$[-0.0004, \ldots, -0.0018]$	401
Chin	$[-0.0004, \ldots, -0.0317]$	435

3.2.6 Normalising Features into a Homogeneous Domain and Feature Vector

The descriptors compute heterogeneous features and the problem. To convert them into a homogeneous domain, they were normalised. Feature normalisation modifies the location and scales the parameters of feature values to transform them into a standard domain. From among the various feature normalisation schemes, this dissertation employed the Min-max normalisation scheme defined in (3.22), due to its robustness to outliers, to normalise the partial feature vector into a global feature vector $\bar{f}$ in (3.23). In other words given, a feature vector defined as

$$\bar{x} = \frac{x_i - \min_i}{\max_i - \min_i}. \tag{3.22}$$

where $\bar{x}_i$ denotes the normalised i^{th} component, min_i and max_i represent the minimum and maximum of the i^{th} component values in the features space, respectively. We will then obtain the following normalised feature vector

$$f = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p\}. \tag{3.23}$$

The normalization procedure is defined in Algorithm 1, in which f^k is the K^{th}

3.2.7 Selecting Salient Features from Facial Components

Any feature is most likely to be corrupted by noise, which may cause the learning algorithm to over-fit. Hence, selecting only features that are relevant is essential to

Algorithm 1: Feature Vector Normalization

function NORMALIZE **(F)**
input : Set of feature vectors space **F** of dimension p.
output: Normalised feature vectors space $\bar{F}$.
$\bar{F} \leftarrow \{\}$
for *each* $f = \left\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p\right\}$ F **do**
  for $i \leftarrow 1$ *to* p **do**
    $\bar{x}_i = \frac{x_i - \min_i}{\max_i - \min_i}$
  end
  $\bar{f} \leftarrow \left\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p\right\} x_{norm}$
  $\bar{F} \leftarrow \bar{F} \bigcup \bar{f}$
end
return $\bar{F}$

minimise performance biases. Among feature selection algorithms, Sequential Selections Algorithms (SSA) have shown positive results when applied to various domain of Pattern Recognition [46]. This dissertation used a wrapper-based approach from SSA, i.e. Sequential Forward Selection (SFS) to select salient features from a pool of local features. This SFS algorithm is a bottom-up search approach. It starts with an empty feature set S and gradually adds chosen features some evaluation function to minimise the classification error [5]. At each iteration, the feature to be included in the feature set is selected among the remaining available features. The features can be shown better in a graphical form, Figure 3.10.

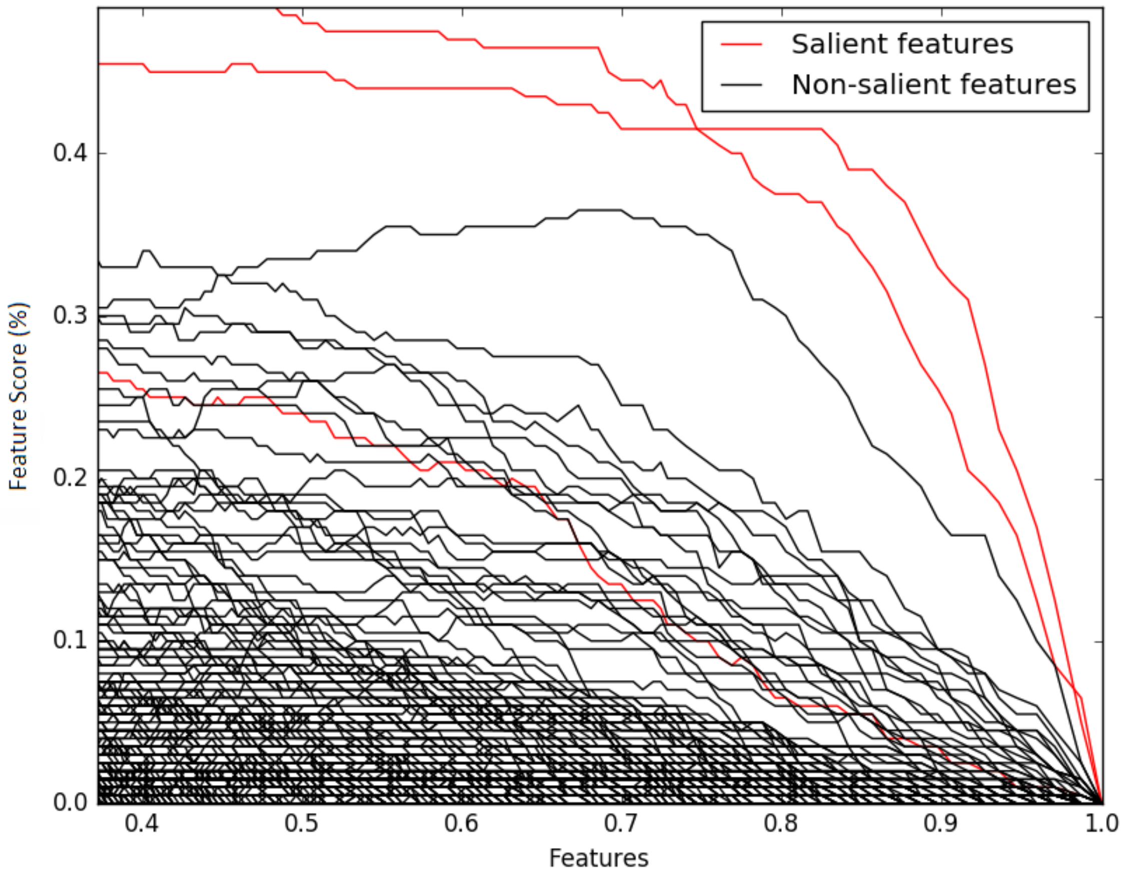


FIGURE 3.10: Salient features vs non-salience features

The area under the curve indicates the relevance of features, that is, their scores. The black lines represent features that are not relevant, some of which are regarded as noise. The red lines represent the prominent features which contain rich attributes in texture form. It can be seen that many features are clustered at a lower level, this implies that they are not abundant in the sense that they may reduce the performance of an algorithm.

3.2.8 Classification

To perform classification of the proposed model, this dissertation employed the error-correcting output code (ECOC). The ECOC is an extended version of the SVM; it can classify multiple subjects; as a result, our chosen data sets contain more than two subjects with various classes per subject. Hence ECOC is a well-suited classification method to handle this issue. The recognition procedure used in this work is defined below, and it is applied to each database:

- Split databases into training and testing using the 80% training and 20% testing split.

- Pre-process images from both training and testing datasets to reduce noise with the aim to have a better detection and recognition.
- Locate the face and the components using an extended cascade detector on each image in the database.
- Extract the successfully detected components for feature extraction.
- Compute features per component using the shape and texture descriptors.
- Use these features to train the classifier ECOC.
- Conduct the tests using test images and the trained database.
- Repeat the above steps with other chosen facial databases.

3.3 Conclusion

This chapter provided an in-depth analysis of the methods and techniques used in this research to support the proposed component-based architecture. The next chapter discusses the results achieved through vigorous testing of the proposed model using state of the art face databases.

Chapter 4

Results and Discussion

4.1 Introduction

This chapter presents the results obtained with the methodology described in the previous section. The chapter is organised as follows: Section 4.2 provides an overview of the instruments used in this research; Section 4.3 discusses the results obtained through simulations using the proposed techniques and presents the performance analysis of the proposed technique and Section 4.4 concludes the chapter.

4.2 Research Instruments

The set of instruments used in this research are not new; have been used before, and they are available as open source resources, mostly for research purposes. Each instrument is discussed, giving its history, previous usage and its sensitivity.

4.2.1 Computer System Configuration

To test the degree of efficiency of the proposed component based facial recognition model architecture, experiments and simulations were performed on different settings of the state of the art facial databases using two computer vision libraries; Computer Vision System Toolbox (CVST) and Open Computer Vision (Open CV). These tools allowed the researcher to extend algorithm functionalities to object detection, content retrieval and classification.

The criterion for choosing these libraries was the successes in many other different computer vision studies, hence, instead of reinventing the basic functionality from scratch, new implementations were built on top of what came before. Here, an in-depth analysis of each of the tools and the merits for their selection in this study is given, and each instrument is discussed, together with its history, previous usages and sensitivity.

The configurations of the tools and simulations were performed using an Intel Core i7 3.10 GHz PC. Experiments and simulations were conducted on a system configured with the Computer Vision System Toolbox (CVST) and Open Source Computer Vision (Open CV). These tools were used to extend algorithm functionalities such as object detection, content retrieval and classification.

- **Computer Vision System Toolbox (CVST)** provides algorithm functionalities and applications for simulating computer vision and video processing systems. The toolbox was employed to extend algorithm functionalities such as feature detection, extraction, and matching.

- **Open Computer Vision (Open CV)** is designed for real-time applications and it takes advantage of both multi-core processing and hardware acceleration. The library was employed to develop an application prototype that simulates real-time facial recognition.

4.2.2 Face Databases

The experiments were performed using four state of the art facial databases: FERET, CMU-PIE, SCFace and FEI. Images of these databases mimic real-world conditions and allow face recognition algorithm testing. These facial databases are available for non-commercial purposes, and each is discussed in-depth below:

- **FERET** contains 15 sessions collected between August 1993 and July 1996 and it is updated from time to time since it is used for many purposes, including research. It has 1564 sets of images with a total of 14,126 images from 1199 individuals and 365 duplicates [47].

- **SCFace** contains static facial images.These were taken in uncontrolled indoor environment and overall there are 4160 static images from 130 subjects [48].

- **CMU-PIE** contains 41,368 images from 68 subjects. Each individual face configures under 13 various poses and 43 different illumination conditions [49].

- **FIE** has 14 images per subject. There are 200 subjects and every image is captured against a similar white background in an upright frontal position [50].

4.3 Results and Performance Analysis

To verify the proposed hybrid component-based face recognition model, it was vigorously tested, using real-world datasets. Four states of the art face databases, CMU-PIE, FEI, SCFace and FERET, were used to perform experiments. Images of these databases mimic real-world scenarios and enable robust testing on various domains in face conditions. The experiments were based on the general overview of face recognition using the actual facial components. This was done to test which facial landmarks are most discriminative for facial recognition on the various configurations of the face. Each experiment was conducted on the four distinctive databases.

4.3.1 Face Recognition under Occluding Objects

Components compensate for various issues related to facial recognition, and in this research the focus is on the illumination, pose and occlusion. Hence using the CMU-PIE face database allowed us to test the proposed model in different face pose and illumination changes. Change in pose and light degrades the accuracy of face recognition systems (meaning generally), but mainly the holistic-based face recognition systems. The CMU-PIE face database contains images that vary in in facial expression, camera viewpoint and illumination. Each person is configured under 13 different poses, 43 different illumination conditions, and with four different expressions. All images went through preprocessing for scaling and converting to the same colour domain. Table 4.1 lists the face recognition results on this database with a gallery view of permuted components.

TABLE 4.1: CMU-PIE face database recognition results

Gallery View	**Alarm Rate(%)**		**Time(ms)**
	Accuracy	**Error**	
Eyes, Mouth, Nose	96	0.0416	0.32
Eyes, Mouth, Forehead	95.4	0.0482	0.58
Eyes, Mouth, Chin	96.8	0.0331	0.74
Eyes, Mouth, Cheeks	96	0.0416	0.59
Eyes, Mouth, Nose, Forehead	97	0.0309	0.88
Eyes, Mouth, Nose, Chin	96	0.0416	0.94
Eyes, Mouth, Nose, Chin, Cheeks	97.4	0.0261	0.98
Eyes, Mouth, Nose, Chin, Cheeks, Forehead	98.4	0.0163	0.63

The results show that a combination of all face components achieves 98.4% accuracy. Although the accuracy varies on the different configuration of components, vigorous training of the classifier using various samples of facial components was done to ensure that the classifier could correctly classify facial components from other occluding objects that might result is misclassification. The eyes, nose and the forehead achieve the least recognition accuracy compared to different facial component permutations. Because some facial images of this particular database were mainly affected by hairstyle as a result, not many details were obtained from the forehead. However, other components such as the cheeks, and the chin have a lot of rich features. As a result, they achieved significantly better results compared to the other components.

4.3.2 Face Recognition under Pose Changes

The FEI face database contains images taken against a similar white background in an upright frontal position with profile rotation of up to about 180 degrees. Experiments on this database mainly focus on testing the pose changes with a large scale variation of up to 180° which is a complete full side view of the face. Table 4.2 list the face recognition results of this database and it has achieved an acceptable degree of accuracy of 95.8% in an upright view with all the components combined with an error rate of 0.0163. It can be seen from the permutation of other components that all combinations achieved a recognition accuracy that is above 90%. This indicates components which are effective for facial recognition in various illumination conditions.

TABLE 4.2: FEI face database recognition results

Gallery View	**Alarm Rate(%)**		**Time(ms)**
	Accuracy	**Error**	
Eyes, Mouth, Nose	93.4	0.1990	0.36
Eyes, Mouth, Forehead	93	0.0752	0.39
Eyes, Mouth, Chin	90	0.1000	0.33
Eyes, Mouth, Cheeks	93	0.0752	0.42
Eyes, Mouth, Nose, Forehead	93.4	0.0701	0.58
Eyes, Mouth, Nose, Chin	90.3	0.1074	0.63
Eyes, Mouth, Nose, Chin, Cheeks	90	0.1000	0.80
Eyes, Mouth, Nose, Chin, Cheeks, Forehead	95.8	0.0438	0.92

4.3.3 Face Recognition under Illumination Changes

Although illumination degrades the performance of the face recognition system, the contrast and brightness were achieved by preprocessing the face images using the proposed preprocessing technique discussed in section 3.2.3. In this research, the SCFace database was used to test the proposed facial recognition model under varying illumination conditions. This database contains static face images taken in the uncontrolled indoor environment causing images to have various qualities. Table 4.3 lists the recognition results of the SCFace database. Using the hybrid pre-processing technique proposed in this research, the proposed model was able to overcome the low light conditions and achieved an acceptable degree of accuracy of 97.8% in all components combined.

TABLE 4.3: SCFace database recognition results

Gallery View	**Alarm Rate(%)**		**Time(ms)**
	Accuracy	**Error**	
Eyes, Mouth, Nose	83	0.2048	0.46
Eyes, Mouth, Forehead	96.2	0.0395	0.51
Eyes, Mouth, Chin	98.3	0.0173	0.64
Eyes, Mouth, Cheeks	89	0.1236	0.49
Eyes, Mouth, Nose, Forehead	98.8	0.0121	0.33
Eyes, Mouth, Nose, Chin	97.8	0.0225	0.94
Eyes, Mouth, Nose, Chin, Cheeks	79.5	0.2579	0.98
Eyes, Mouth, Nose, Chin, Cheeks, Forehead	97.8	0.0225	0.63

4.3.4 Face Recognition under Clustered Scenes

The FERET face database was collected in 15 sessions between August 1993 and July 1996. It is updated from time to time, since it is used for various purposes such as research. This dissertation used the database to test the proposed model in the clustered scene. Table 4.4 lists the recognition results obtained from using this database.

TABLE 4.4: FERET database recognition results

Gallery View	**Alarm Rate(%)**		**Time(ms)**
	Accuracy	**Error**	
Eyes, Mouth, Nose	88.4	0.1312	0.46
Eyes, Mouth, Forehead	80	0.2500	0.58
Eyes, Mouth, Chin	77.4	0.2912	0.74
Eyes, Mouth, Cheeks	83	0.2048	0.59
Eyes, Mouth, Nose, Forehead	83	0.2048	0.88
Eyes, Mouth, Nose, Chin	80	0.2500	0.94
Eyes, Mouth, Nose, Chin, Cheeks	95	0.0526	0.98
Eyes, Mouth, Nose, Chin, Cheeks, Forehead	95	0.0526	0.63

The overall performance of the proposed model is evaluated using the ECOC, is given in Table 4.5. The system achieves an accuracy rate of 98.75% on the ECOC classifier. Figure 4.1 shows the Receiver Operator Characteristic (ROC) curve of the ECOC classifier which is the accuracy of the correctly recognised face images.

TABLE 4.5: ECOC classifier performance analysis

Classifier	**Alarm Rate(%)**		**Error**	**Time(ms)**
	Sensitivity	**Fall-out**		
ECOC	98.75	1.25	0.0125	0.87

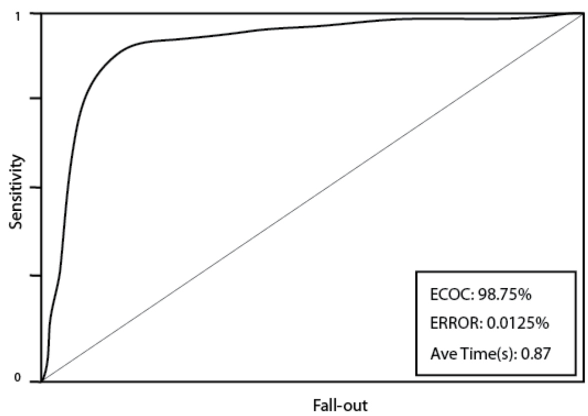


FIGURE 4.1: ROC curve of the ECOC classsifier

Sources on component-based facial recognition systems are scarce in the literature, however, in this study, researchers managed to find a few approaches that have been proposed by previous studies. Table 4.6 compares results of three studies that have used the proposed techniques to those of this research. Although these methods employ state of the art techniques, the proposed hybrid component-based method outperformed them and achieved an acceptable degree of accuracy of 98.75%. This signifies that components are effective for facial recognition and they allow flexible geometric face rebuild in uncontrolled conditions.

TABLE 4.6: Comparison of the proposed method with the state of the art methods

Author and Year	**Technique**	**Database**	**Accuracy** (%)
Proposed Approach	**Hybrid Component-based**	**CMU, FEI, SCFace, FERET**	**98.75**
Bonnen [51]	Component-based	AR Face, FERET	96.67
Yi [52]	PCA, LFW, PAF	FERET, PIE	87.77
Shyam [53]	Eigenfaces with BCD	ORL	95.45

4.4 Conclusion

This chapter discussed the results obtained by the methods and techniques defined in Chapter 3. Observing the performance of the system, it can be seen that the components individually have a high discriminative power compared to holistic-based approaches. The next chapter concludes this research work and discusses possible extensions and future work.

Chapter 5

Conclusion

5.1 Concluding Remarks

This research proposes a hybrid component-based approach that seeks to utilise any successfully detected components. The system synthesises the various facial views using shape and texture descriptors. To improve the fitting accuracy, we initialised the ECOC and SVM using estimates of the facial landmark locations obtained by a method based on a mixture of components.

Experimental results performed on FERET, CMU-PIE, SCface and FEI databases demonstrated the effectiveness of our proposed method, which showed significantly better results than other the state of the art algorithms. The proposed approach works very well in normalising the near-frontal poses; however, its primary weakness is in standardising facial images with large pose variations. In semi-profile poses, half of the face is usually occluded, which results in a distorted normalised face. This distortion has an adverse impact on the recognition accuracy. A further limitation of the proposed method is that it does not handle the normalisation of facial expressions.

5.1.1 Improving facial recognition using a facial component-based approach

The eyes, nose and mouth remain the primary facial components for facial recognition; however, incorporating other facial components such as the forehead, cheek and the chin has made a significant improvement on the accuracy of the proposed facial recognition model. Other than the primary components of eyes, nose and mouth, the forehead and the cheeks have a high discriminative power in providing adequate unique features for face recognition.

5.1.2 What features can adequately represent the identified facial components?

In this research, texture and size are primary sources of features. With aid of powerful descriptors, we simplified the computation of features and made it easier to represent them in a spacial domain. That being said, texture is the most adequate attribute that has rich enough details about each component of the face.

5.1.3 How useful is the proposed model regarding variations in pose, illumination, expression and affine transforms?

The major factors are illumination and pose. Nevertheless, the issue of illumination was largely handled by the proposed hybrid preprocessing scheme that combines three preprocessing techniques is: Gamma Correction, Difference of Gaussian (DOG) filter, and Contrast equalization [34] [35] [36]. These three preprocessing techniques have a significant high discriminative level of preprocessing the image by enhancing the image contrast, reducing noise and removing dark patches on the image. The issue with pose is that the proposed model could only handle faces with tilts of up to 45°, beyond that, the proposed approach fails to conduct facial component detection.

5.1.4 Comparing holistic feature-based and component feature-based techniques for facial recognition.

The proposed method overcomes the shortcomings of individual methods such as the holistic-based and geometric-based approaches to compensating for illumination, pose, occlusion and facial expression and allows flexible geometrical relation between the face components. Its primary focus is to conduct face recognition at the component level and evaluate the effectiveness of facial components on face recognition under abnormal conditions.

5.2 Future Work

In future, we will study the use of features that are less invariant to ageing variations. This will make the system more reliable in recognising people from images that have been taken with significant time gaps. Moreover, we plan to design an intelligent system that can integrate multiple sources of biometric information. We plan to add a method that not only increases the accuracy of the scheme but also is computationally efficient. This research also serves as a basic introduction to the face recognition research community, after which, in future researchers aim to utilise Convolutional Neural Network (CNN) and extend the thought towards Deep Learning.

Appendix A

Formulae of the Alarm Rate

A.1 Calculating the Correct and Incorrect Identifications

- **P** is the number of actually positive samples
- **N** is the number of actually negative samples
- **TP** is the number of actually negative samples
- **TN** is the number of true negatives
- **FP** is the number false positives
- **FN** is the number false negatives
- **TPR** is the true positive rate, samples correctly classified as true
- **TNR** is the true negative rate, sample classified as true but they are false

A.1.1 Number of correct identification cases

$$P = TP + FN \tag{A.1}$$

A.1.2 Number incorrect identification cases

$$N = FP + TN \tag{A.2}$$

A.2 Calculating the Sensitivity and Specificity

A.2.1 Sensitivity

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{A.3}$$

A.2.2 Specificity

$$TNR = \frac{TN}{N} = \frac{TP}{FP + TN} \tag{A.4}$$

Bibliography

[1] M. P. Beham and S. M. M. Roomi, "Face recognition using appearance based approach: A literature survey", in *Proceedings of International Conference & Workshop on Recent Trends in Technology, Mumbai, Maharashtra, India*, vol. 2425, 2012, p. 1621.

[2] B. Heisele, P. Ho, J. Wu, and T. Poggio, "Face recognition: Component-based versus global approaches", 1, vol. 91, Elsevier, 2003, pp. 6–21.

[3] J Ravi and K. Raja, "Hybrid domain based face recognition system", 6, vol. 3, Eswar Publications, 2012, p. 1402.

[4] A. K. Jain and S. Z. Li, "Handbook of face recognition", Springer, 2011.

[5] J. Kittler, "Feature selection and extraction", 1986, pp. 59–83.

[6] W. W. Bledsoe, "The model method in facial recognition", vol. 15, 1966, p. 47.

[7] T. Kanade, "Computer recognition of human faces", Birkhäuser, 1977.

[8] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces", 1, vol. 12, IEEE, 1990, pp. 103–108.

[9] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces", 3, vol. 4, Optical Society of America, 1987, pp. 519–524.

[10] M. Turk and A. Pentland, "Eigenfaces for recognition", 1, vol. 3, MIT Press, 1991, pp. 71–86.

[11] K. Etemad and R. Chellappa, "Face recognition using discriminant eigenvectors", in *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, IEEE, vol. 4, 1996, pp. 2148–2151.

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

[13] J. Naruniec and W. Skarbek, "Face detection by discrete gabor jets and reference graph of fiducial points", in *International Conference on Rough Sets and Knowledge Technology*, Springer, 2007, pp. 187–194.

[14] B. Heisele, T. Serre, and T. Poggio, "A component-based framework for face detection and identification", 2, vol. 74, Springer, 2007, pp. 167–181.

[15] H. Hse and A. R. Newton, "Sketched symbol recognition using zernike moments", in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, IEEE, vol. 1, 2004, pp. 367–370.

[16] H. Kim, "Survey paper: Face detection and face recognition", 2007.

[17] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach", 1, vol. 8, IEEE, 1997, pp. 98–113.

[18] J. Dargham, A. Chekima, and M. Hamdan, "Hybrid component-based face recognition system", Springer, 2012, pp. 573–580.

[19] S. Mandal and B. C. Dhara, "A hybrid face recognition method based on structural and holistic features", in *Advances in Pattern Recognition, 2009. ICAPR'09. Seventh International Conference on*, IEEE, 2009, pp. 441–444.

[20] B. C. D. Srinivas Nagamalla, "A novel face recognition using facial landmarks", 2009.

[21] J. Huang, B. Heisele, and V. Blanz, "Component-based face recognition with 3d morphable models", in *International conference on audio-and video-based biometric person authentication*, Springer, 2003, pp. 27–34.

[22] K. Yesu, H. Jyoti Chakravorty, P. Bhuyan, R. Hussain, and K. Bhattacharyya, "Hybrid features based face recognition method using artificial neural network", in *Proceedings - 2012 2nd National Conference on Computational Intelligence and Signal Processing, CISP 2012*, vol. 8, Mar. 2012.

[23] T. M. Kodinariya, "Hybrid approach to face recognition system using principle component and independent component with score based fusion process", 2014.

[24] H. Cho, R. Roberts, B. Jung, O. Choi, and S. Moon, "An efficient hybrid face recognition algorithm using pca and gabor wavelets", 4, vol. 11, SAGE Publications Sage UK: London, England, 2014, p. 59.

[25] G. Karthik, "A novel approach for face recognition using PCA and artificial neural network", 2014.

[26] A. Mahajan and P. Kaur, "Face recognition system using ebgm and ann", Citeseer, pp. 2277–3878.

[27] W. Wu, M. Kan, X. Liu, Y. Yang, S. Shan, and X. Chen, "Recursive spatial transformer (rest) for alignment-free face recognition", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3772–3780.

[28] D. S. Bolme, “Elastic bunch graph matching”, 2003.

[29] R. B. Qaum Mehdi Bal Gopal Agarwal, “A review on face detection methods”, 2, vol. 74, Springer, 2011, pp. 167–181.

[30] D. K. P. Belhumeur J. Hespanha, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection”, IEEE, 1997, pp. 711–720.

[31] L. Wiskott, N. Krüger, N Kuiger, and C. Von Der Malsburg, “Face recognition by elastic bunch graph matching”, 7, vol. 19, IEEE, 1997, pp. 775–779.

[32] H. Yang, “Human face detection in complex backgrounds”, 1, vol. 27, 1994, pp. 53–63.

[33] F. Sakai Nagao, “Line extraction and pattern detection in a photograph”, 4, vol. 1, 1969, pp. 1735–1740.

[34] H. Guo, H. He, and M. Chen, “Gamma correction for digital fringe projection profilometry”, 14, vol. 43, Optical Society of America, 2004, pp. 2906–2914.

[35] T. K. Kim, J. K. Paik, and B. S. Kang, “Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering”, 1, vol. 44, IEEE, 1998, pp. 82–87.

[36] S. Wang, W. Li, Y. Wang, Y. Jiang, S. Jiang, and R. Zhao, “An improved difference of gaussian filter in face recognition.”, 6, vol. 7, Citeseer, 2012, pp. 429–433.

[37] P. Viola and M. J. Jones, “Robust real-time face detection”, 2, vol. 57, Springer, 2004, pp. 137–154.

[38] G. Beliakov, S. James, and L. Troiano, “Texture recognition by using glcm and various aggregation functions”, in *Fuzzy Systems, 2008. FUZZ-IEEE 2008.(IEEE World Congress on Computational Intelligence). IEEE International Conference on*, IEEE, 2008, pp. 1472–1476.

[39] A. Eleyan and H. Demirel, “Co-occurrence based statistical approach for face recognition”, in *Computer and Information Sciences, 2009. ISCIS 2009. 24th International Symposium on*, IEEE, 2009, pp. 611–615.

[40] *Glcm texture tutorial. (n.d.).* `http://www.fp.ucalgary.ca/mhallbey/tutorial.htm`, Accessed: Feb 24, 2017.

[41] A. E. Hamyani, A Ghanmi, and A Intissar, “Generalized zernike polynomials: Integral representation and cauchy transform”, 2016.

[42] C. Geng and X. Jiang, “Face recognition using sift features”, in *Proceedings of the 16th IEEE International Conference on Image Processing*, ser. ICIP’09, Cairo, Egypt: IEEE Press, 2009, pp. 3277–3280, ISBN: 978-1-4244-5653-6.

[43] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", 2, vol. 60, Springer, 2004, pp. 91–110.

[44] C. Geng and X. Jiang, "Face recognition using sift features", in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, IEEE, 2009, pp. 3313–3316.

[45] P. Dreuw, P. Steingrube, H. Hanselmann, H. Ney, and G Aachen, "Surf-face: Face recognition under viewpoint consistency constraints.", in *BMVC*, 2009, pp. 1–11.

[46] Z. Zhu, Y.-S. Ong, and M. Dash, "Wrapper–filter feature selection algorithm using a memetic framework", 1, vol. 37, IEEE, 2007, pp. 70–76.

[47] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The feret evaluation methodology for face-recognition algorithms", 10, vol. 22, IEEE, 2000, pp. 1090–1104.

[48] M. Grgic, K. Delac, and S. Grgic, "Scface surveillance cameras face database", 3, vol. 51, 2011, pp. 863–879.

[49] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie", 5, vol. 28, Elsevier, 2010, pp. 807–813.

[50] C. E. Thomaz, "Fei face database", 2012.

[51] K. Bonnen, B. F. Klare, and A. K. Jain, "Component-based representation in automated face recognition", 1, vol. 8, IEEE, 2013, pp. 239–253.

[52] D. Yi, Z. Lei, and S. Z. Li, "Towards pose robust face recognition", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3539–3545.

[53] R. Shyam and Y. N. Singh, "Evaluation of eigenfaces and fisherfaces using bray curtis dissimilarity metric", in *Industrial and Information Systems (ICIIS), 2014 9th International Conference on*, IEEE, 2014, pp. 1–6.