COMPUTATIONAL FACE RECOGNITION USING MACHINE LEARNING MODELS

A.A.M. ELMAHMUDI

PHD

UNIVERSITY OF BRADFORD

2021

Computational Face Recognition Using Machine Learning Models

Ali Ahmed M ELMAHMUDI

Submitted for the Degree of

Doctor of Philosophy

Faculty of Engineering and Informatics

University of Bradford

2021

Abstract

Ali Ahmed M Elmahmudi

Computational Face Recognition Using Machine Learning Models

Keywords: Average Face, Partial Face recognition, Age progression and regression.

Faces are among the most complex stimuli that the human visual system processes. Growing commercial interest in face recognition is encouraging, but it also turns out to be a challenging endeavour. These challenges arise when the situations are complex and cause varied facial appearance due to e.g., occlusion, low-resolution, and ageing. The problem of computer-based face recognition using partial facial data is still largely an unexplored area of research and how does computer interpret various parts of the face. Another challenge is age progression and regression, which is considered to be the most revealing topic for understanding the human face changes during life.

In this research, the various computational face recognition models are investigated to overcome the challenges posed by ageing and occlusions/partial faces. For partial face-based face recognition, a pre-trained VGGF model is employed for feature extraction and then followed by popular classifiers such as SVMs and Cosine Similarity CS for classification. In this framework, parts of faces such as eyes, nose, forehead, are used individually for training and testing. The results showing that there is an improvement in recognition in small parts, such as recognition rate in forehead enhanced form about 0% to nearly 35%, eyes from about 22% to approximately 65%. In the second framework, five sub-models

were built based on Convolutional Neural Networks (CNNs) and those models are named Eyes-CNNs, Nose-CNNs, Mouth-CNNs, Forehead-CNNs, and combined EyesNose-CNNs. The experimental results illustrate a high recognition rate when it comes to small parts, for example, eyes increased up to about 90.83% and forehead reached about 44.5%. Furthermore, the challenge of face ageing is also approached by proposing an age-template based framework, generating an age-based face template for enhanced face generation and recognition. The results showing that generated new aged faces are more reliable comparing with state-of-the-art.

Declaration

I declare that this thesis is the result of my own work except where references have been made to the work related to others. This thesis has not been submitted anywhere for the application of another degree, diploma, or other qualification.

Ali Ahmed M Elmahmudi

Dedications

To my Father and Mother,

To My Wife and Children,

To My Brothers & Sisters,

To my friends.

Acknowledgments

All thankfulness to Allah, the most gracious and the most merciful, for providing me with the blessing to undertake research towards this PhD. I could not have done this without his mercy.

I wish to express my sincere appreciation to my supervisor, Professor Hassan Ugail, who has the substance of a genius: he convincingly guided and encouraged me to be professional and do the right thing even when the road got tough to complete this research.

I wish to give my special thanks to my parents for their love, support and encouragement, without whom I would never have enjoyed my study and without them, I would not be here.

I would like to thank my wife for her patience and support through the most challenging period of my PhD research. To my children, the light in my life and my everything, and to my brothers and sisters for their support.

For my friends for their positive influence in my life. To my colleagues and friends; Dr. Irfan Mehmood, Ali Bukar, Shelina Jilani, Ahmad Al-Dahoud and Aliyu Abubakar for their help and kindness.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AAM	Active Appearance Model
AE-CNN	Age Estimation guided Convolutional Neural Network
AI	Artificial Intelligence
AOA	Average of Over All
AT	Affine Transformation
AW	Average weights
BIF	Biologically Inspired Features
BiPCA	Biharmonic Principal Component Analysis
CACD	Cross-Age Celebrity Dataset
CD	Cosine Distance
CDL	Coupled Dictionary Learning
CFE	Composite Face Effect
CNNs	Convolutional Neural Networks
Conv	Convolutional
CPS	Cyber–Physical Systems
CS	Cosine Similarity
DFM	Dynamic Feature Matching
DLSL	Dictionary Learning and Subspace Learning
DT	Delaunay Triangulations
ECC	Error Correcting Codes
ED	Euclidean Distance
EVD	Eigenvalue decomposition
FaceVACS	Face Verification Alarm Compare System
FCN	Fully Convolutional Networks
FFA	Fusiform Face Area
FoL	Floor of Log
GAN	Generative adversarial network
GPA	Generalised Procrustes Analysis
HOG	Histogram of Oriented Gradient
IAAP	Illumination-Aware Age Progression
KNN	k-Nearest Neighbour
LBP	Local Binary Pattern
LFW	Labelled Face in Wild

LRC	Linear Regression Classification
NNAODL	Nuclear Norm based Adapted Occlusion Dictionary Learning
OSPF	Optimized Symmetric Partial Facegraph
OVO	One-vs-One
PCA	Principal Component Analysis
PEM	Probabilistic Elastic Matching
PLDA	Probabilistic Linear Discriminant Analysis
RCLBP	Reinforced Centrosymmetric Local Binary Pattern
ReLUs	Rectified Linear Units
RFA	Recurrent Face Aging
RS-LDA	Random Subspace Linear Discriminant Analysis
RS-LDA	Random Subspace Linear Discriminant Analysis
SIFT	Scale Invariant Feature Transform
SIFT	Scale Invariant Feature Transform
SN	Section No
SP	Sub-Pattern
SPC	Subclass Pooling for Classification
SPDA	Structure Preservation and Distribution Alignment
SRC	Sparse Representation-based Classification
SSIM	Structural Similarity Index
SVMs	Support Vector Machines
TPGM	Topology Preserving Graph Matching
VGG	Visual Geometry Group

List of Publications

Journal papers

1. Elmahmudi A, Ugail H, **Deep face recognition using imperfect facial data**, Future Generation Computer Systems, Volume 99, 2019, pp. 213-225.

2. Elmahmudi A, Ugail H, **The Biharmonic Eigenface**, Signal, Image and Video Processing 13, no. 8 (2019): 1639-1647.

3. Elmahmudi A, Ugail H, **A framework for facial age progression and regression using exemplar face templates**. The Visual Computer 2020, 1-16.

4. Abubakar A, Ugail H, Smith K. M, Bukar A, and Elmahmudi A, **Burns Depth Assessment Using Deep Learning Features**, J. Med. Biol. Eng., 2020.

Conference papers

1. Elmahmudi A, Ugail H, **Experiments on Deep Face Recognition Using Partial Faces**, 2018 International Conference on Cyberworlds (CW), Singapore, 2018, pp. 357-362.

2. Elmahmudi A, Ugail H, **The Effect of Face Ageing in Face Recognition**, 2nd AIERC 2018, Bradford, UK.

3. Elmahmudi A, Ugail H, and Bukar A, **Average Weights of Eigenvalue Decomposition for Efficient Face Image Processing**, in INTECH 2018, London, UK; 2018 (Best paper)

Papers being prepared

1. Ali Elmahmudi and Hassan Ugail *Lightweight CNN models for accurate face recognition using partial face data.*

1 Introduction

Automatic face recognition plays an active role in various practical applications of life such as, biometric authentication and automated immigration clearance, which are usually stationed at airports. Irrespective of the wider applications the process of facial recognition is still riddled with multiple challenges [1]. For example, a difference in facial expression, perceived age, and occlusions of face, which includes partial face data can all affect the mechanism by which a facial recognition framework functions. The majority of security and vision-related systems have actively focussed on developing techniques for biometric-authentication and person identification. In the field of computer vision, there are numerous technologies which use control-assess processes for instance recognition based on the eyes (iris recognition), fingerprint, as well as face recognition. While accuracies associated with iris recognition and fingerprint-based-systems are state-of-the-art, the contactless acquisition of facial images for facial recognition enables the process to be considered a better option [1].

The human face is an expressive, visual organ, which enables an individual to extract information related to gender, race and expression/ psychological state [2]. The ability to recognise a face has significant social implications. The process of familiar face recognition is a proven concept within face perception. Recent publications [3] [4] suggest that faces familiar to a person have a robust representation in memory because the face one encounters is usually within multiple contexts. Hence, when one is asked to describe a face, the human cognitive system attempts to discern the features of the face, which can be matched with what is seen, in order to make a decision. Nonetheless, it is understood that various aspects of the face can play a significant role in face

representation [5]. In face perception literature a categorisation system exists which orders the features of the face into two types: internal and external[2] [6]. The internal features consist of the eyes, nose and mouth, and are known to be involved with the recognition of familiar faces. In contrast, the external features include the head shape, ears, hairstyle, jaw-line and the chin. External features are typically relied upon for unfamiliar face recognition [2] [7].

It is an improbable fact that a general visual framework can describe a human face in the way that humans do, based on a set of features as shown in Figure 1-1. When individual segments of a facial image are cut out and displayed in isolation, it becomes a difficult recognition task for people, even though, faces are the most painted images of the human visual system during an individual's lifetime. In this sense, it is not surprising that humans possess remarkable face recognition ability. Typically, a fleeting glance of an individual's face can be sufficient for the recognition of that individual. Published literature reports on a specialised region within the human brain called the Fusiform Face Area (FFA) which is dedicated to the processing and recognition of faces [6]. It is thought that the brain remembers important details relating to the shapes and colours of crucial features, which correspond to the eyes, nose, forehead, cheeks and the mouth [10]. The human brain is able to cope with significant adaptations in lighting, facial expressions and distance. Contrary to this, computer systems have not shown robust behaviours when presented with faces of varied appearance and lighting conditions. For instance, the variations in illumination, expression, pose and other added physical changes i.e. glasses or facial hair, can have a vast impact on the recognition rates. While this may be considered a limitation, given the sufficient amount of information a computer can cope with, in addition to machines increased data processing power. It is rational to consider

that in certain scenarios a machine algorithm can function on-par, if not outperform a human during face recognition.
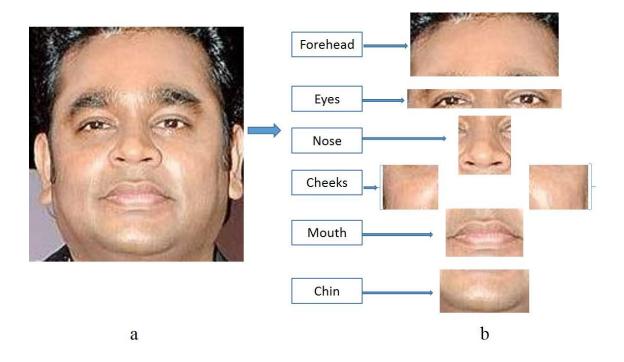
Figure 1-1: a) The full face. b) The important face features after partitions of the face. After dividing the face into parts, it clear that the eyes part contains more information compared to the rest of the parts. The recognition is steel not difficult, contrary to that the cheeks have less information which makes the recognition process hard.

Another aspect of the recognition, which affects face recognition tasks, is the challenges surrounding age progression and regression. Over the past decade, many researchers have been working on face processing systems in order to tackle such challenges of face recognition [11][12]. Currently, one of the most significant topics in computer vision is face analysis and its related applications for recognition or verification. Due to the variability of applications which include finding missing people including children, face recognition in age-invariant has been receiving growing attention. Yet, recognition relating to the concept of age is an area of concern, with limited researcher resource. This is of great research importance especially since some people might appear to look older/ younger facially, compared to other people, despite the fact being that they

are of the same age. Moreover, while the process of ageing is universal, the mechanism by which it occurs is not uniform as people age differently and excessively due to multiple reasons [13].

The variation in the appearance of a human face due to the process of ageing or any other effects can be very different, which subsequently makes the recognition process more challenging for both humans and machines alike. The images presented in a well-known facial image database known as FGNET [14], consists of participant images which are more than 30 years apart, as mentioned in Figure1-2. The face of the individual presented in Figure 1-2 does not appear alike even at the varied ages. The face of the child depicted at 2 years of age looks slightly different when shown at the age of 5. The most visually different representation of the same individual is at the age of 43, which makes the recognition impossible when shown in comparison to the image of the subject at the age of 2.

Age 2   Age 5   Age 8   Age 43

Figure 1-2: Some samples faces from the FGNET face database [14] **at four different ages for the same person**.

Presently, a smart biometric framework is a composite smart system that involves individuals as well as virtual and physical frameworks, which function to accomplish precise security, protection, verification and privacy. This type of smart framework is used for authentication and verification for human-cyber– physical systems (CPS). Moreover, communication, portability of cellular phones

and intelligent devices have different sensors in expansion to enormously upgrade exhibitions and capacity space compared with existing cell phones. Nevertheless, the verification process of these devices and the requirement for a user password when gaining access to personalised directories and services of a social network, for example, prove to be of great vulnerability for smart device security.

With the emerging use of functionally high computational techniques, face analysis has benefited immensely. Face recognition including face verification, human emotion recognition and age synthesis, are some of the prominent application areas of computational techniques. Lifestyle and health-related issues are also known to affect the process of physical ageing [13]. Hence, the process of ageing is complex and is not sufficient to model a computer-based system. As people physically age, the morphology of the face alters and is affected by a spectrum of factors. While it is known that all human faces follow the same pattern of changes, for example, loss of baby fat from a young age to the appearance of prominent wrinkles at old age. The rate of these age-related changes is measurable with ethnicity and specific lifestyle choices. Figure 1-3 shows a typical workflow of a computer algorithm, which is used to create 'ageing' images of an individual, from a single frontal face image.
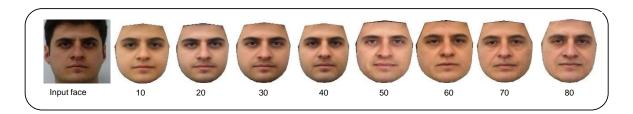
Figure 1-3: Age progression and regression. Given a single frontal view face image of an adult [22]**, a typical algorithm generates faces of varied ages. The input image is used to generate new ages from age 10 to 80.**

Various computational techniques have been presented in the literature, discussing the issues related to ageing, which mostly rely on techniques such as [23] [24]. The Cartoon technique was used to exaggerate age and was reported in research by Burt and Perrett in 1995, who used the method to simulate the effects of ageing in facial images [25]. They computed average faces at various age stops in order to synthesise ages from an input image by producing a new face. On the other hand, Principal Component Analysis (PCA) was used by Changseok with a 3D face shape model for extracting components of age (related) changes in the 3D face to which a test face was added in order to synthesise output faces at different ages [26]. Research by Young et al. addressed changes of faces along with ageing and reported that several parts of the face including the nose, mouth and eyes, in addition to some proportion of differences between such features and wrinkles were extracted [27].

Computer based face recognition is still riddled with many challenges compared to human face recognition ability. For a human, seeing someone briefly is usually enough for recognition, subject to the person being familiar. This is because the brain memorises important details relating to the person. As a matter of fact, it is thought, that when a familiar face is presented within a different scenario the brain compares the 'before' and 'after' images, without the use of any significant new information [6]. Nevertheless, in general, for a machine, the variability of the appearance of a face has a direct effect on face recognition [29].

There are many machine learning algorithms within computer vision specific to face-related functions. The algorithms are either unsupervised frameworks, which are not explicitly programmed or supervised which are based on the idea that a user can select a part of data with known labels (based on the knowledge of the operator) or feed it to a software as a training set.

Principle Component Analysis, PCA, was introduced in 1905 [30] and is an unsupervised machine learning model widely used for reducing dimensionality and image compression [31]. It is a technique for coding large faces into facespace. In 1991, Eigenface [32] was applied for face recognition algorithms to extract the most important details known as "eigenvectors" from the faces which corresponded to the maximum eigenvalues, representing variation. Essentially, when it comes to face recognition, the concept of an average face is interesting, and in some way appears to "mimic" human face recognition. Previous attempts to use the average face as a tool for computer-based face recognition appears to be plentiful in published literature. In this respect, the works presented in [33] and [34] are noteworthy, since it is shown that combining a number of images of an individual into one reliable measure, is sufficient for face recognition. Moreover, Eigenspace and Eigenvalue decomposition (EVD) techniques for the construction of face recognition algorithms, which use the average face computed in the pixel space, appears to be a relatively common method in literature [35] [36] [37].

Another commonly used technique is Local Binary Patterns LBP, which was introduced by [38], and is a simple yet powerful feature for texture classification. This approach divides an image into different regions and extracts features from each region separately and those features are subsequently used for the purpose of classification.

Machine learning is a supervised approach and sub-category of Artificial Intelligence (AI) [39]. The main objective of machine learning is to understand the data structure and fit it into models that can be understood and used by others. While machine learning is a part of computer science, it varies from conventional computational methods. In conventional computing, the algorithms are groups of

explicitly programmed commands utilized by machines to figure or solve a problem. Machine learning approaches permit training by using input data and utilise statistical analysis to produce values that fall within a particular domain. For that purpose, machine learning simplifies computers to build models from samples of data in order to make an automated decision, based on data inputs. [39]. One of the most popular examples of machine learning is Convolutional Neural Networks (CNNs) [39], which have been utilized recently in an area of computer vision, especially in face-specific applications. CNNs are supervised models that function to gain knowledge by training and can also be used to represent the most discriminatory features, i.e., feature extraction, face recognition, classification, and segmentation. There are several trained models in literature e.g. VGGF, FGG16, VGG19, OverFeat etc [40] and [41]

1.1   Research Objectives

In this research, our aim is to investigate the challenges posed by face recognition and develop efficient and robust algorithms by focusing on two main challenges: occlusions and ageing effects. In order to work towards the development and understanding of face recognition and to identify the various aspects, there are certain challenges that need to be investigated as part of this research. More specifically, our objectives are to:

- Reduce the dimensionality of a large dataset of facial images and build rich and deep feature dataset.

- Explore how a computer system performs face recognition subject to imperfect facial information as recognition cues. More specifically, gearing our framework to explore how various parts of

the face perform for the task of face recognition, taking into account factors such as face orientation and distance, which is achieved with the effect of zooming out and rotation.

- Develop and build a framework which addresses the issue of age progression and regression from facial images based on different ethnic identities, as well as gender. And investigate how such factors affect face verification.

- Create a deep learning-based model for the task of face recognition, for individual aspects of the face, with minimal feature dimensions and convolutional layers.

1.2 The Importance of this Research

Various aspects of face recognition issues are covered within this research. Firstly, the use of average weights (AW) in the Eigenspace is presented for face processing and the importance of this research is the ability to reduce the dimensionality of the computational space and to find an efficient method to represent features of a large dataset of images. This is achieved by using Eigenvalue Decomposition (EVD) with averaging, since in the past, EVD has been applied to faces, whereby averages were computed in the pixel space. A fundamental deficiency in such methods is that a certain degree of manual operations is required to separate the face from the given image as well as to warp it over a template image. Contrary to such a method, the concept of AW in the Eigenspace is introduced for representing the features for face processing. Additionally, an algorithm is investigated and developed by investigating deeply into the various regions of the face, in order to best understand which part plays the most significant role in recognition. This is achieved by studying the separate face parts from a variety of modification such as rotations, zooming and face

inversion. The importance of this research is to apply deep learning algorithms for feature extraction and using them for face recognition based on the multiclass classification method.

Also, a novel approach to address face generation and recognition problems are introduced. Template faces are used based on the formulations of an average face of a given ethnicity and for a given age. For a recognition task between photos taken years apart, a pre-trained VGG-Face model was utilised for feature extraction. For the purpose of classification, four well-known classifiers were applied. Our experimental results suggest that the proposed approach achieves accuracy, efficiency and possess much-needed flexibility when it comes to facial age progression or regression challenges.

Finally, introducing a new face recognition model which is based on five new models for partial face recognition and each model has been trained on a different region of the face. Our results indicate that the region of the eyes yields the highest rate of recognition compared to the other parts of the face.

1.3 Research Contributions

The outcome of this research and the main contributions presented in this thesis are summarised as follows:

1) Development of an efficient approach to compute average faces, which is the concept of EVD to compute average weights over facial images in the Eigenspace. Our approach has a competitive advantage for face processing and analysis compared to the standard average face computed in the pixel space through image

pre-processing and transformations. This research is was presented at the INTECH 2018 conference, London, UK [42].

2) A novel modification to the method of PCA is done which utilises the inherent averaging ability of the discrete Biharmonic operator as a pre-processing step. This mechanism is referred to as the BiPCA. Interestingly, by applying the Biharmonic operator to images, it is possible to generate new images of reduced size while keeping the inherent features them intact. The resulting images of lower dimensionality can significantly reduce the computational complexities while preserving the features of interest. This work has been successfully published in Signal Image and Video Processing Journal 2019 [43].

3) Studying the rate of recognition subject to the various parts of the face such as the eyes, nose, mouth and the forehead. In this study, a Convolutional Neural Network based architecture with the pre-trained VGG-Face model were used to extract features for training. Two classifiers are applied namely the Cosine Similarity CS and the Linear Support Vector Machine to test the recognition rates. This research is published in future generation computer systems journal in 2019 [44] and in International Conference on Cyberworlds (CW) [45].

4) Introducing a novel approach to address face generation and recognition problems. The proposed approach addresses the

problem of face-age progression and regression challenges in a more direct manner by relying on individualised face-age templates. The templates are built based on the formulations of an average face for various ethnicity and gender over a range of ages. This work has recently been published in the visual computer journal [46].

5) Reducing the number of features and trainable parameters in order to decrease a time complexity is done by building a model for each part of the face. Comparing to previous models, our model produces 200 feature dimensions instead of 4096 features with high accuracy.

1.4 Thesis Outline

The rest of the thesis is structured into six chapters. Chapter 2 outlines related work and a literature review for each chapter. A face averaging framework for enhanced face recognition is presented in Chapter 3. Chapter 4 illustrates a framework for deep face recognition using imperfect facial data. Chapter 5 demonstrates a new method for age progression and regression and studies the effect of face-age generation, for the recognition process. Chapter 6 introduces new deep learning models for partial face recognition. And, finally, in Chapter 7, the conclusion of this research and discussions relating to the present work including its limitations are presented alongside with the scope for future work.

2   Literature Review

Computational Face Recognition plays an important role in the field of security and crime prevention, since it functions as a crime deterrent in many real world scenarios. There are a variety of applications for facial analysis, for instance, image interpretation, health examination, forensic medicine, face animation, and facial image compression [47].

In this chapter, studying the various aspects of Computational Face Recognition from several viewpoints. The avenues of study include; (1) the computation of an average face and its effect on the recognition process and (2) the recognition of ageing and partial face information.

This chapter is organised as follows: Section 2.1 presents literature relating to previously conducted work on the concept of average faces. Section 2.2 presents the literature review for work discussed as part of Chapters 4 and Chapter 6, which includes Deep Face Recognition using Imperfect Facial Data. Lastly, Section 2.3 presents literature in relation to Chapter 5, which investigates the effect of face age generation.

2.1   Review of the Concept of Average Faces

The concept of average faces has been widely studied for face analysis purposes. From a computational viewpoint, the work presented as part of [35] and [36] provide a sufficient background on this topic of the average face.

The earliest work on the topic of the average face was presented by Burton et al., [33]. Burton and colleagues used a simple image-averaging technique in order to derive abstract representations of known faces. This process was carried out by applying face recognition on a database of images

collected from the Internet. A total of 1000 images were collected of 50 celebrities, with a set of 20 images representing each celebrity. All the database images were morphed into the same shape using a graphical program; by overlaying an image of a face with a grid to calculate the average. Averages were generated over 3, 6, 9 and 19 images in the pixel space. The research by Burton et al., concluded that by increasing the number of images for face representation to compute an average face, affects the results positively.

Similarly, Jenkins and Burton [34] report that the percentage of automatic face recognition increases up to 100% when using an image averaging technique. They implemented publically available Face Verification Alarm Compare System (FaceVACS) [34]. A database of 31,077 images of 3,628 well-known identities [34] were used and face averaging was carried out over nine different images. This was an online probe database with 500 images of celebrities, where 20 photographs for each of the 25 male celebrities were made publically available. A new image was created by computing an average from 20 images per subject.

Robertson et al., [48], tested average faces by using the face verification system available on a popular Samsung Galaxy Smartphone. They conducted an experiment on the smartphone, to measure the recognition rate between normal and average faces. Two experimental conditions were investigated firstly using the real faces and secondly, computer generated averaged faces. The experimental results showed that recognition based on the average face outperformed real face photographs. A drawback of the experimental system as reported by the researchers, was that facial landmark identification required manual input.

Hwang et al., [29], proposed a novel face recognition method by using the concept of average example image. The approach was to divide the images into different sets by using a k-means clustering [49]. These sets were clustered into several groups according to the image variation approach. The researchers then applied the example faces to compute an average example face for each cluster. It was then compared with the averages to find the most similar cluster. A local distance was calculated between the input image and the examples in the cluster. Hwang and colleagues tested their system by using several facial feature extractions algorithms; Local Binary Pattern, (LBP) [50], the Histogram of Oriented Gradient (HOG) [51], Gabor features [52] and PCA + LDA methods [53] and [54]. The experimental method was tested on the multi-PIE face database [55] and the recognition rates reported were 85.93%, 88.60%, and 90.47% for LBP, HOG and PCA +LDA, respectively.

Similarly, Lee et al., [56], proposed a method called Best Basis Selection by using learnt weights for face recognition. The methodology was based on predicting an error of classification via training stages and detecting the valuable basis faces, which was used as a similarity metric of traditional pattern methods. They applied their experimental model to different databases by selecting approximately 9% of data for testing and the rest for training. The results showed that the negative effect on the misaligned face images was reduced and also that the weights of the valuable basis faces (eigenface) declined with a view of enhancing the accuracy of classification.

Çarıkçı and Özen [57], examined face recognition by using an Eigenfaces approach for extracting features from face images. They used the most popular distance measure known as the Euclidian Distance, to compute a distance between a query image and training images, as a recognition step. The authors

applied their method to the Face 96 database [58] and reported a classification accuracy of 94.74%.

As previously mentioned in the introduction, Turk et al.[32], introduced the novel concept of Eigenfaces, which treats the problem of face recognition in two dimensions where PCA is at the heart of it. An Eigenface essentially can capture important properties relating to the facial features, which go beyond the usual visual features such as the eyes, nose, and the mouth. A prominent advantage of the Eigenface-based face recognition is its capability to recognise new faces of a given individual in an unsupervised way. Further advantages include the ease of implementation, low computational complexity, and low sensitivity to subtle changes to the face of a person of the same identity.

In the past, there has been much work in trying to establish improvements to the Eigenface methodology resulting from the original PCA implementation. For example, a hybrid approach based sub-pattern technique, within the context of face recognition using PCA, was proposed by Hsieh and Tung [59]. Their experimental work was inspired by the method of Independent Component Analysis (ICA) [60][61] and contained the combination of two centred Principal Component Analyses which they referred to as PCA I and PCA II. Hsieh and Tung combined the sub-pattern approach [59] with PCA I and PCA II, which they called SP-PCA I and SP-PCA II, and applied it to the task of face recognition. This method was tested on three datasets, namely the Yale Face Database (A and B) [62] and the Weizmann Face Database [63]. The experimental results report that this approach outperformed traditional PCA with recognition as high as 98%. While the results appear state-of-the-art, one must exercise caution when interpreting the results since Hsieh and Tung used the leave-one-out testing approach, with a relatively small data sample.

Similar work was carried out by Ghinea et al. [64] whose prime aim was to tackle the issue of illumination differences and pose for face recognition. The researchers explored gradient orientated based PCA subspace using the well-known Schur decomposition [65]. They computed Schurvalues and Schurvectors in order to find subspace projections, which they referred to as Schurfaces. By computing similarities for matching, the Hausdorff distance [66] with the Nearest Neighbour classifier [67] was used. Similar to previously published literature two different face datasets were tested; ORL [68] and Yale [62]. The error rates were lower (error rate of 14% for Yale and 15% for ORL) compared to the traditional PCA results, which were 25% and 23%, respectively.

When considering research with the use of PCA, the earlier work of Poon et al. [69] is significant. The researchers proposed a technique called Gradientfaces to images as a pre-processing step prior to PCA. The main idea behind the technique was to seek the orientation of the gradient of the face at pixel level. This enabled the researchers to re-represent an illumination invariant version of the input face. Using the Asian face database, consisting of 40 subjects (10 images per subject) with various expressions and lighting conditions, the method was tested. With the use of Gradientfaces an improvement in the recognition rate was reported increasing by 6.25% to 60.75%, when compared to the traditional PCA method. While the results are promising, a noted limitation of the study is that the researchers only tested one dataset where the images were fairly uniform with a lack of noisy data.

Moreover, there are other concepts based on pattern recognition such as Linear Discriminant Analysis (LDA) [70], which has also been tried in conjunction with PCA. A notable piece of work in this area is that of Oh et al. [71], who combined PCA with LDA subsequently referred to as the PCA-LDA method. For

a given set of images, Oh and colleagues used PCA to reduce the dimensionality of a feature space and LDA to improve the separation between the classes. They tested their approach using the ORL [68] and the Yale [62] database, and the results reported a 2.5% improvement in accuracy for the ORL dataset, while an 8.9% improvement was achieved for the Yale dataset when compared to the traditional PCA. The main drawback of the methodology was that almost 80% of the data was used for training and the remaining 20% was used for testing. Moreover, the datasets used were somewhat limiting in that, the Yale database consisted of only 15 subjects with 11 images per subject.

2.2 Deep Face Recognition using Imperfect Facial Data

When it comes to machine-based face recognition, traditional algorithms such as PCA [30] [32] and Linear Discriminant Analysis [72] have been widely employed. However, more recently the implementation of Machine Learning models such as Convolution Neural Networks (CNNs) [39] have consistently proved to be accurate and efficient.

Though face recognition via Machine Learning algorithms is a well-researched problem [73], it is yet to yield promising results, when probed with partial face data, as currently the mechanism is restricted to the utilisation of the full face data. To date, little work appears to have been done on the use of partial face for face recognition [74].

As far as computer aided face recognition based on partial facial images is concerned, the literature centring this topic is relatively sparse and inconsistent. One of the earliest work known to be conducted on partial face was conducted by Savvides et al., [75]. In the study, Savvides and colleagues tested various facial regions to establish quantifiers with discriminative ability. Based on grey

scale images the method of kernel correlation filters was utilised to reduce image dimensionality and feature extraction [76]. Following that, Support Vector Machines (SVMs) were used to discriminate between facial features. Three main regions of the face were investigated namely the eye, nose, and mouth. Experimental results indicated that the region of the eye has a higher verification rate compared to the mouth and nose.

Similarly, He et al., [77] introduced a technique called Dynamic Feature Matching (DFM) for partial face recognition. The study was based on a combination of Fully Convolutional Networks (FCN) [78] with sparse representations. The purpose of FCN is to extract a feature map of images that has the capacity to cater for more discriminative features. The core of this work was the utilisation of the VGG-Face model [41] from which features were transferred to the FCN. This method appears to have produced good classification accuracy of 97% in CASIA-NIR-Distance [79] and 74% for YouTube datasets [77].

Furthermore, several robust face recognition methods have been suggested in order to address the challenges arising for face recognition as a result of facial occlusion. Long et al.,[80], proposed Subclass Pooling for Classification (SCP) to solve the double occlusion problem by using limited data in a training set. A fuzzy max pooling method was applied in addition to average pooling schemes. The results reported that a remarkable margin of performance reached at 95% for the AR (Aleix and Robert) [81] face dataset and approximately for 94% in Labelled Face in Wild (LFW) [82] .

More recently, Lahasan et al., [83] proposed a framework named the Optimized Symmetric Partial Facegraph (OSPF) for face recognition, which

tested multiple conditions, such as face occlusion, varied facial expression and differences in lighting conditions. The experimental results demonstrated that improvements in recognition rates can be achieved by introducing partial facial data.

Duan et al., [84] introduced a technique called Topology Preserving Graph Matching (TPGM), to enhance the recognition process in the case of partial faces. The method is centred on building geometric graphs for probe faces and gallery faces. The TPGM method minimises a geometric and textural cost function. As a result of the experiments conducted upon four face databases (LFW, EYB [85], AR [81], and PubFig [86]) Duan and colleagues demonstrated that their experimental approach outperformed other state-of-the-art methods such as FaceNet [87], Local Binary Pattern (LBP) [88], and Probabilistic Elastic Matching (PEM) [89].

Similarly, research by Cai et al., [90] proposed a facial variation modelling system for sparse representation "to refer specifically to an expression of the input signal as a linear combination of base elements in which many of the coefficients are zero" [91] for face recognition. This can help single sample face recognition algorithms to build facial variation bases to separate neutral, frontal faces from different facial views. The experiment reported significant enhancements could be performed in single-image face recognition problems.

Another piece of work that is notable in the area of partial face recognition is by Li et al., [92], who considered issues relating to frontal view face images with varying illumination, disguise and occlusion. They presented a novel method for face recognition, which extracts a dynamic subspace from images and obtains the distinctive parts in each subject. A characteristic of discriminative components

was represented by those parts in order to give a recognition protocol to classify face images by using the k-nearest neighbour algorithm (K-NN) [93]. Similar to previous research, the choice of database included ORL and Extended Yale B. The results were promising and demonstrated that the recognition rates could be improved using part facial cues.

Comparing to human, Calvo et al., [94] studied the rate of face recognition on partial face subject with regards to facial expressions. For one of the experiments, the researchers tested the face recognition rates for six common facial expressions: happiness, anger, sadness, disgust, fear, surprised. Face images were divided into two segments, one containing the eyes and the other containing the mouth. A considerable finding based on the experimental methodology was that humans have poor recognition skills when the region of the eyes and mouth are shown in isolation. On the other hand, Calvo et al., [94] noted that the expression of smile produced slightly better recognition rates compared to the rest of the expressions.

Research surrounding the mechanisms which underpin human face perception have also addressed how the human visual system perceives parts of the face including faces shown in inversion [95], [96], [97] and [98]. The work of Murphy et al., [99], for example, reports that human faces are difficult to perceive and recognise when presented upside down, see Figure 2-1. Moreover, in their experiments, the researchers tried to test the participant's ability to classify faces presented in whole and region-by-region viewing conditions, using a dynamic aperture which moved incrementally through the facial picture, as illustrated in Figure 2-2. The main idea of this work was to understand the limits of human face perception. The results achieved by the observers were put into categories

separated by identity, gender, age and facial expression, and the results report

detrimental effects of face inversion.

Figure 2-1: Humans possess inherent difficulty in accurately recognising inverted faces. The left image is the correct pose and the right image is upside down. These samples are taken from the FEI data set [22]**.**
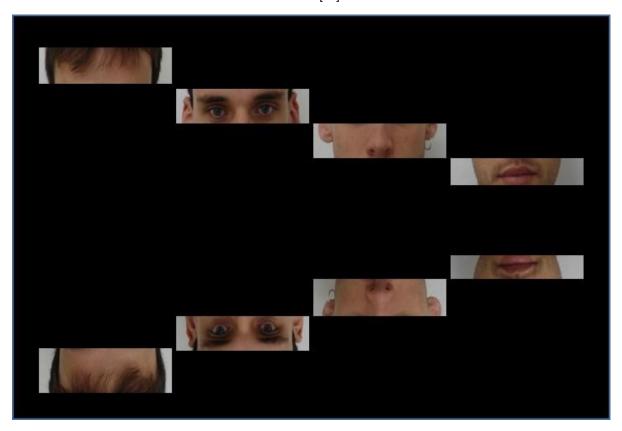
Figure 2-2: An illustrative example of physical facial stimuli utilised to understand human face perception [2]**.**

Young and his colleagues [7], conducted experiments using a face composite technique (Composite Face Effect) [8][9] by dividing a face horizontally into two parts i.e. upper and lower. Whilst the participants may have performed better at the recognition task, if the upper portion of a holistic face was shown in isolation of the lower as illustrates in Figure 2-3. The researchers demonstrated that when the two parts are shown in combination (upper and lower portion), with an incorrect lower face representation, the task of recognition became difficult. It also appears that when a face is misrepresented i.e. features presented in a disjointed manner, recognition is hindered [8]. This Composite Face Effect (CFE) is known to occur when face images are split horizontally and shown holistically. If two parts of the face are both present, but not aligned with one another, then the recognition of each face half is not impacted, i.e., each part is correctly recognised independent to the other. However, it is usually easier to recognise the upper half of the face when misaligned with the lower portion, instead of being smoothly joined. CFE is substantial in showing that the issues with recognition arise as a result of there being a general facial arrangement – separated from this distinction, the misaligned and composite are the equivalent.
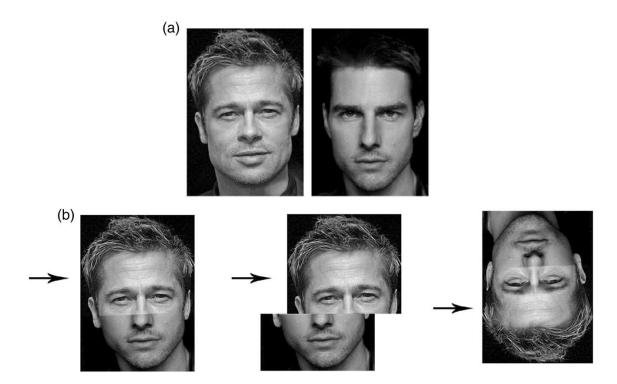
Figure 2-3: Illustration of the composite face task. On the top (a), two original faces (Brad Pitt and Tom Cruise) [9]. b) The first face is an aligned composite face between two halves, the top part belongs to Brad and the bottom half is Tom. The second face is a mis-aligned composite face in the same order. Last face is an aligned and composite face in an upside-down arrangement.

However, when dealing with acute occlusions of the face, the performance of current methods declines remarkably. Many previous studies note that when it comes to human face recognition, familiarity appears to be a key recognition factor [100] [101]. The effect of familiarity changes when the target face image is partially presented, occluded, or is displayed with varied expression and changes in age [99] [94].

Similarly, Du, et al [102] proposed a frame work for face recognition under occlusions and illumination changes, called Nuclear Norm based adapted Occlusion Dictionary Learning (NNAODL). Firstly, a model named a nuclear norm-based error for characterizing was introduced for the corrupted region (a damaged area of an image) and occlusion of a test face. Then, they integrated an error image (i.e. the representation residual matrix) [102] within the training

set; to build a dictionary, in order to rebuild a corrupted and non-corrupted area within test faces. Comparing to present sparse representation approaches, the model proposed by Du, et al reports better representation for samples with noise with a decreased effect of occlusion.

Dictionary learning and subspace learning (DLSL) was suggested in 2019 by MengmengLiao, XiaodongGu [103] for face recognition under significant variations in pose and occlusion. The researchers compared their method with other algorithms tested on different face datasets. The experimental results showed that their method could achieve better recognition rate than other approaches such as Structure preservation and distribution alignment (SPDA) [104] and Appice [105].

The Sparse Representation-based Classification (SRC) face recognition framework was introduced by Wright et al. [91]. In this system, Wright and colleagues coded occluded face images by using l1-norm minimization similar to a sparse linear combination of the widened dictionary. The expanded dictionary included training sets and the occlusion dictionary. For the purpose of classification, they worked to find which category of training set images results in a minimum error of reconstruction with coding coefficients.

Recently, in 2020 Wenbo Zheng, et al [106] presented a novel piece of work called a Novel Approach Inspired by Optic Nerve Characteristics for Few-shot Occluded Face Recognition, to solve the challenge of face recognition using occluded faces. They introduced a powerful method to extract features of local and contextual information for the purpose of recognition. In addition, they introduced an adaptive fusion approach to combine various features, including the suggested structural element feature, Reinforced Centrosymmetric Local

Binary Pattern (RCSLBP), and connected-granule labelling features. Face recognition is inferred by the fusion of all classification results according to the method. The experimental framework was tested on three different databases, Extended Yale B, LFW [107], and AR [81] and the results demonstrated a high performance compared to current methods of face recognition under occlusion, which reached 98.98% in the case of a face with sunglasses and 87.96% in the case of a scarf.

Additionally, more recent work [108] introduced a high-efficiency method which relied on features by using a contemporary process for face recognition implementations called Floor of Log (FoL). This approach has the advantage of reducing the amount of required energy, storage, and preserves precision. To learn the best parameters of the proposed method, the experiments applied the Support Vector Machines (SVMs) and K-Nearest Neighbours algorithms with a cross-validation technique [109]. Three publically available face datasets were used to test the proposed method and the results were comparable to the methods utilizing identical classifiers with more reduction in the size of data. Moreover, their system can adapt in order to achieve significant results and incorporated edge computing schemes.

On the other hand, Machine Learning algorithms can utilise the power of computations to use copious amounts of input data for training and use numerical analysis in order to produce outputs that can challenge the power of human face recognition [110]. Thus, Machine Learning helps a computer to build models from input data, with a view to making a more accurate decision. This is a distinct advantage that Machine Learning algorithms appear to have over human face perception and recognition [111]. Thus, it is also plausible to state that machine learning algorithms can potentially provide better recognition rates on partial

faces or, in the worst case, may aid humans to perform better at face recognition, especially in challenging cases where very limited or partial facial data are presented [44].

Generative adversarial network (GAN) [112] is also another technique which has been applied to the problem of face recognition. YANG LU, et al [113] suggested a novel Wasserstein generative adversarial network-based framework for occluded facial expression recognition. The method consists of one generator (G) to complete occlusion of a facial image and is subsequently followed by two Discriminators, D1 and D2 for distinguishing between simulated and real image. Lu et al., verified the efficiency of the method by using two datasets namely RAF-DB [114] and AffectNet [115]. The experimental results show that the technique has a positive influence on various patterns of occlusion. Moreover, the rate of recognition is enhanced in both cases of occluded and non-occluded faces.

2.3 The Effect of Face Aged Faces in the Recognition Process

Just like many themes in computer vision, automatic age generation and recognition is one of the most challenging yet vital image-based problems, which have multiple real-life applications. Deep Neural Networks such as the use of generative adversarial networks (GAN) [112] for age synthesis has become prominent. Moreover, the focus for the majority of techniques is based on simulation, whereby facial data is utilised for constructing generative models that are then used for age synthesis in cases of either age progression or regression.

The majority of age-related work have been introduced for the task of face recognition, to handle age-invariant difficulties and challenges [15] [16]. For instance, the gradient orientation pyramid was introduced by Ling et al [17] as a feature extraction methods and for the recognition step, they applied Support

Vector Machines (SVMs). However, Li [18] used Local Binary Pattern (LBP) [19] and Scale Invariant Feature Transform (SIFT) [20] for feature extraction and Random Subspace Linear Discriminant Analysis (RS-LDA) [21] technique for face recognition. Hence, there are many discriminative algorithms, which have been designed for suitable feature extraction and effective classification.

Recently, Bukar et al. presented an automatic face ageing method [116], which included the development of a person-specific facial ageing system. The methodology consists of face feature extraction by a colour based Active Appearance Model (AMM) and then applying regression to generate a face image of a given age. Experiments were conducted using the HQFaces dataset [117] and the Dartmouth Children's Faces database [118] and the results generated were deemed reliable estimates of the input faces in comparison to ground truth.

In 2017, Convnet features for age estimation were used for facial age estimation by Bukar and Ugail [119]. The method was based on extracting features from an input image using the VGG-Face model [41] and Partial Least Squares Regression (PLS) was applied to reduce the dimensions of extracted features and redundant information [120]. The experimental framework was tested on two different databases, FGNET-AD [14] and Morph II [121] and the results reported were comparable to some previous algorithms, such as the Biologically Inspired Features (BIF) [122] and 3-step [123].

Similarly, Riaz et al. introduced a method based on 3D gender-specific ageing model which automatically produced simulated faces at a specified age from an input face [124]. The model was constructed with different datasets and the results were comparable with other methods. Also, the ground truth faces demonstrated efficiency of their suggested technique.

An age simulation algorithm built on super-resolution in tensor space [125] and active appearance models (AAMs) [126], was proposed by Wang et al. [127]. This method simulated the effects of an ageing adult on the face by reducing blurring effects, which resulted from a normalisation process of a face. The FGNET [14] database was used and the experimental results illustrated that the ageing simulation results were adequate based on the recognition rate between the generated and ground truth faces which was 70%.

A closely related topic to machine-based face ageing is the area of age estimation. Recently, machine learning approaches have been applied to solve age estimation problems by using Convolutional Neural Networks (CNNs). For example, Zheng [28] proposed a method based on the deep face, which is the first attempt at age estimation, to obtain features of age-invariant. The technique is referred to as Age Estimation guided Convolutional Neural Network (AE-CNN). Zheng and colleagues addressed the method of separation for the variations that resulted because of ageing from human-specific features, which are considered stable. Similarly, Liu et al. [8] proposed a method for age estimation called AgeNet. The researchers addressed the problem of age estimation by implementing two types of models as well as a Gaussian model for classification. They used a large-scale 22 layer deep CNNs with a label distribution, while regression relied on the classification to encode apparent age.

Similarly, personalised Age Progression with ageing Dictionary was proposed by Shu, Xiangbo et al [128]. The main goal of the method was to apply rendering ageing faces in a personalised manner. Their approach relied on two phases; (1) online and (2) offline. During the offline stage, short-term ageing image pairs were collected from available datasets and an ageing dictionary was trained. Whereas during the online stage, the researchers rendered an ageing

face for an input face in an age group with a nearest neighbour. Then, the resulted ageing face was used as an input for an ageing face in the next age group. This process was repeated until all aged faces were generated. Cross-Age Celebrity [129] and Morph ageing databases [121] were used as part of the experiments and the results demonstrated an advantage of the proposed method compared with Illumination-Aware Age Progression (IAAP) [130] and ground truth.

Recent work based on Generative Adversarial Network (GAN) was introduced by Zhang et al. [131] for age regression and progression. The approach is referred to as Conditional Adversarial Autoencoder Network (CAAE). The idea behind it is that they use a convolutional encoder in order to map an input face to a latent vector [132] and then to project the resulted vector to a face manifold conditional on age by applying a de-convolution generator [133]. This vector conserves features of a personalized face and an age condition controls regression and progression. The system was trained on a large dataset called the UTKFace dataset [134] and it was evaluated by different databases such as Morph and Cross-Age Celebrity Dataset (CACD) [129]. The results indicated that the system could generate faces in a more realistic manner and encompasses flexibility.

Regarding face recognition on cross generated faces, a number of studies have been conducted [127] as a result of the lack of algorithms, which can be used to generate new faces from original input face images. As mentioned above, Wang et al [127] created a method for age progression based on super-resolution in tensor space. They adopted Eigenfaces [32] to extract features from the samples of the new face groups. Then the nearest neighbour method K-NN [93] was used for classification with Euclidean Distances. Their

results show that there were improvements in recognition across computer generated-faces.

In recent years, most age related work has been introduced on face recognition to handle age-invariant difficulties and challenges [15] [16]. For instance, the gradient orientation pyramid was introduced by Ling et al [17] as a feature and for the recognition step, they applied Support Vector Machine (SVMs). However, Li [18] used Local Binary Pattern (LBP) [19] and Scale Invariant Feature Transform (SIFT) [135] for feature extraction and Random Subspace Linear Discriminant Analysis (RS-LDA) [21] for face recognition. Thus, there are numerous discriminative algorithms, which endeavour to design a suitable feature and an effective classification model.

Xu et al [136], proposed the model of a neural network named Coupled Auto Encoders (CAE) to deal with age-invariant face identification problem. Their experiments were applied to three public databases with FGNET [14] and the results demonstrated that the proposed method had an efficiency on recognition.

Recently, Zhou et al [13], introduced a model called Identity-Inference, which depends on an age subspace learnt from appearance-age labels. Firstly, Probabilistic Linear Discriminant Analysis (PLDA) was used, then the ageing subspace is learnt with the apparent age labels and an identity subspace was located by an Expectation Maximization (EM) algorithm [137]. The experimenters used FGNET [14] and the results illustrated that their technique generated higher performance accuracy compared to other studies. For example, in the FGENT data set the percentage of recognition reached

88.23% comparing with 76.2% and 71.3% respectively when using Maximum entropy model [15] and Feature-aging model [138].

Wen et al [132], proposed a novel approach to learn age invariant features from a face based on Convolutional Neural Networks, called a deep face recognition framework. Again, experiments were conducted on FGNET database [14] in addition to other ageing databases. Comparable to Zhou et al., the results demonstrated a significant performance.

Gong et al [15], presented a method to address the problem of representation and classification of face recognition in age invariance. They extracted information from an image by using maximum entropy feature descriptor and used a system called Identity Factor Analysis IFA for the process of matching. Similar to previous literature experiments were conducted on the FGNET and Morph face databases and good results were achieved; 76.2% using FGNET and 92.26% in Morph compared to 69.0% in FGNET and 91.14% for Morph by Gong et al. [139].

2.4 Summary

As per the concept of average faces and based on all of the cited literature, there is a strong indication that there is merit in investigating the use of the average face as a powerful tool for face processing. However, most of the work previously done relies on computing the average face directly in the pixel space of the images. One of the main drawbacks for such a method is the degree of manual work required to prepare the images, e.g. removal of the background and warping of the face image to a template image. Aside from this, the speed of computation as well as noise (present in the image) are issues that need to be considered. These issues are tackled by performing the computations within

the Eigenspace as opposed to directly on the image space. Our work, therefore, presents the possibility of using the average face as a far more efficient and reliable processing tool, compared to the previously presented techniques.

Apart from some isolated pockets of work mentioned within the literature review, to our knowledge, there has not been any rigorous study on Machine Learning to test the recognition rates for various facial parts. The main aim of this work hence is to try and close that gap. Thus, an efficient Machine Learning framework has been proposed to look deeply into the effects of partial face for recognition. In particular, trying to understand the specific recognition rates for various parts of the face. To do this, our framework has been implemented using Convolutional Neural Network (CNN) architecture for facial feature extraction, and utilising both Cosine Similarity (CS) and SVM for classification and testing as mentioned in Chapter 4 and 5.

What is apparent from the literature within the discipline of facial ageing progression, regression and recognition, is that some of the face age generation techniques develop ages in a way that depends on the features of the face by ignoring the forehead [116]. Moreover, GANs-based approaches [112] generate new faces at different age stages and concludes good results if the test faces are part of the training database, however, the results were unsatisfactory when images are used outside of the training set. Table 2-1 illustrates a snapshot summary of the most relevant state-of-the-art and more details about of comparison of our work with the most relevant works are found at the end of each chapter.

Field	Descriptions	
Average Faces	**Features**	**Limitation**
Burton et al., [33]	A simple image-averaging technique in order to derive abstract representations of known faces.	Computing the average face directly in the pixel space of the images. This requires manual work to prepare the images, e.g., removal of the background and warping of the face image to a template image
Robertson et al., [48]	Average faces by using the face verification system available	A drawback of this system as reported by the researchers, was that facial landmark identification required manual input.
Ageing	**Features**	**Limitation**
Zhang et al. [131]	Generate new ages in different groups using Generative Adversarial Network (GAN). The system was trained on a large dataset called the UTKFace dataset.	It does not work with other face datasets.
Bukar et al. [116]	Generate ages using linear equations. Face feature extraction by a Colour based Active Appearance Model (AMM) and then applying regression to generate a face image of a given age	This work does not include a forehead, which can hold more details about ageing including wrinkles and lines.
Xiangbo et al [128]	Personalised Age Progression with ageing Dictionary. Applying rendering ageing faces in a personalised manner, which relied on two phases; (1) online and (2) offline.	Quality and resolution of the new generated faces are not clear enough (contains blurring).
Imperfect Facial Data	**Features**	**Limitation**
Zheng, et al [106]	Introducing a powerful method to extract features of local and contextual information for the purpose of recognition.	The experimental framework was tested on face datasets including sunglasses and scarf not parts of a face. The presented experimental results were lower than our results as mentioned in table 6-2.
Du, et al [102]	face recognition under occlusions using Nuclear Norm based adapted Occlusion Dictionary Learning (NNAODL)	They used images with a degree of noise.
He et al., [77]	Dynamic Feature Matching (DFM) for partial face recognition using combination of Fully Convolutional Networks (FCN) [78] with sparse representations	The method was applied to the recognition task for the largest parts of the face, such as upper, down, right, and left half of a face and the results were low comparing to our method as illustrated in table 4-1.

Table 2-1: Illustrates a snapshot summary of the most relevant state-of-the-art.

3   Averaging for Enhancing Face Recognition

In this chapter, two contributions relating to Eigenface are presented. Firstly, introducing the concept of average weights in the Eigenspace for efficient face recognition using Eigen Value Decomposition (EVD). Secondly, a novel modification to Principle Components Analysis (PCA) using discrete Biharmonic equations for the recognition system is introduced. This system is called Biharmonic Principle Components Analysis (Bi-PCA).

This chapter is divided into two sections. In the first section, presenting and explaining the effect of applying average weights of eigenvalue decomposition to facial image processing. Secondly, explain the methodological framework of the Biharmonic Eigenface the face recognition.

3.1   Average Weights of Eigenvalue Decomposition for Efficient Face Image Processing

In this section, the use of averaging over faces as an efficient computational approach is investigated for face processing. More specifically, utilising the concept of Eigenvalue Decomposition to compute average weights over facial images in the Eigenspace. Our approach has a competitive advantage for face processing and analysis especially when considering the standard approach to face averaging in the pixel space, which is computed through image pre-processing and transformations. The offered approach uses a small set of training images for computing the average weights in the Eigenspace for faces, which results in a reduction in the level of complexity.

As part of our approach for a given individual, computing the average weights from a chosen of known images avoiding the need to manually separate

the face and the background. To demonstrate the competence of our approach, face recognition is used as an example, i.e., recognition is performed by comparing the average weights of a probe facial image to that of pre-computed average weights of the query image. A varied image-selection technique (as mentioned in section 3.1.2.2) has been used to compute the average weights, which had enabled us to measure the effect of image numbers on recognition rates. Showing that, beyond a certain small number, in our experiments (four images), increasing the number of images to compute the average weights has no significant impact on the recognition rate. In fact, saturation occurs with the composition of as low as four images. Hence, the average weights formed using four images are sufficient to produce an optimum result.

Based on those approaches, a method of EVD was developed in order to compute average weights for faces in the Eigenspace. To test our methodology, as well as to measure rates of face recognition, several standard datasets have been used namely, the ORL [68], the Grimace and the face96 [58]. First, our method was tested against standard average faces computed over the pixel space i.e., created using averages of cropped and warped faces as previously done by others. Then, ran a number of experiments was ran to test the efficiency of our method. The obtained results through our experiments are interesting, i.e., average weights are a more reliable and are an efficient measure for face recognition when compared with simple averages (without needs to correct a pose and adjust images to the same shape). Moreover, a sample size of as little as four images is enough to compute reliable Average Weights (AW) from an image set for a given individual.

It is important not to confuse or directly compare the proposed AW approach for face processing, with some of the state-of-the-art methods such as

Convolutional Neural Network (CNN) based deep learning [41] and [140]. Whilst deep learning for example, provides excellent results in face recognition, such methods are based on supervised learning and are high-dimensional, and computationally intensive, during both the training and testing phases. On the other hand, the EVD-based face image processing technique is computationally very less intensive. Even with some of the low dimensional CNN approaches, the dimension space can be as high as 128 dimensions (number of extracted features per image) whereas the AW approach utilises as few as 45 dimensions, with virtually no training time as opposed to the CNN approaches. Thus, with reduced image feature space the AW approach may be useful in cases where face recognition in computationally less intensive environments. In this work, three important questions are answered, where are;

1. Are average weights in the Eigenspace a reliable measure for face processing?

2. Is it possible to use the average weights obtained directly from images, without any complex transformations or manual operations, as a reliable method to undertake efficient face recognition?

3. What are the optimum number of facial images that are required to find a reliable set of average weights?

3.1.1 Weights of Eigenvalue Decomposition

The main idea is presented in this work hinges on the computation of an average weight space for a given number of images in the Eigenspace, instead of computing the average in the pixel space of a manually prepared original

image. Our approach does not require additional manual intervention or transformation and therefore is more efficient and more reliable.

A modified version of Eigenvalue decomposition (EVD) [37] [141] is employed, whereby the concept of average weights is introduced. EVD is a well-known method that can be used to extract features from original images that can be embedded in the Eigenspace. Below, a description of our proposed method of EVD with average weights in detail.

Let M be the number of images in a training set of the same size. The first step is to convert all the images from 2D into 1D. Secondly, in order to take the background into account in the training set, a normalising step is applied. This is done by computing an average and subtracting the mean image from each image vector Γ. Given that Ψ represents the mean image as described in Equation 3-1 and shown in Figure 3-1.

$$\Psi = \frac{1}{M}\sum_{i=1}^{M}\Gamma_i, \qquad i = 1,2,3,\dots,M, \qquad\qquad (3\text{-}1)$$

where M is number of images.

Figure 3-1: An Illustration of the method used to compute the mean face representation from a given number of images [68].

The ϕ_i is used to define a mean centred image as described in Equation 3-2 below and as illustrated in Figure 3-2.

$$\phi_i = \mathbb{T}_i - \Psi, \qquad i = 1,2,\dots,M. \tag{3-2}$$

Figure 3-2: An Illustrative example of how to remove the mean from an original image [68].

Note that Equation 3-2 leads to a matrix A of the form such that $A = [\phi 1, \phi 2, \dots, \phi_M]$. To compute the corresponding eigenvectors, the covariance matrix C is calculated whereby,

$$C = AA^T. \tag{3-3}$$

If the size of C is larger then the computation of C is,

$$C = A^T A. \tag{3-4}$$

To find the eigenvectors v_i of C, we consider,

$$A^T A v_i = \lambda_i v_i, \tag{3-5}$$

where λ_i are the corresponding eigenvalues.

Now, by multiplying Equation 3-5 by A can obtain,

$$AA^T A v_i = \lambda_i A v_i. \tag{3-6}$$

Then, projecting the eigenvectors v_i on A can obtain the corresponding Eigenspace U_i, such that,

$$U = \sum_{i=0}^{m'} Av_i. \tag{3-7}$$

where $m' \leq M$ is the number of eigenvectors.

The vector corresponding to the weights $\Omega = [\omega_1, \omega_2, \dots, \omega_{m'}]$ represents the contribution of each Eigenface in the input face image space and where $\omega_1, \omega_2, \dots, \omega_{m'}$ are weights for each image and $\omega = U'A$. The average weights, AW for a given subject is computed by,

$$\text{AW} = \frac{1}{c}\sum_{i=1}^{c} \Omega_i, \tag{3-8}$$

where c is the number of weights that belong to the same subject.

For recognition, the test image $\mathbb{T}_e$ is projected into the face space to compute the corresponding weights by,

$$\Omega_e = U^T(\mathbb{T}_e - \Psi). \tag{3-9}$$

A number of measures can be used to compute the similarity in recognition. In this work, the Cosine Distance (CD) similarity measure has been used. CD is a measure of similarity between two non-zero vectors of the inner product family [142]. In order to determine the distances between a test face and the training faces, using CD, the following Equation is applied,

$$D(\Omega_e - \text{AW}_i) = 1 - \frac{\Omega_e \text{AW}_i^T}{\sqrt{(\Omega_e \Omega_e^T)(\text{AW}_i \text{AW}_i^T)}}. \tag{3-10}$$

3.1.2 Experiments and Results
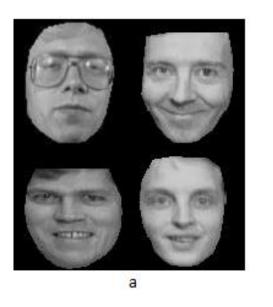
This section presents details of the experiments, which have been undertaken using the average weights on faces in the Eigenspace. Also presenting the results of our experiments showing how the average face within the Eigenspace can be an efficient and convenient way to enhance face recognition. To start with, details of the datasets are discussed and utilised to carry forward our experiments.

For the experiments which have been undertaken, three separate datasets are used, namely ORL [68], Faces96, and Grimace [58]. The ORL database contains 40 subjects, each with ten different facial images in various poses. The images are all an equal size of 112 pixels by 92 pixels. The facial images are all frontal with a slight tilt of the head, different illumination, facial expression, and in some cases, the subjects wore glasses. A sample of faces from the ORL database is shown as part of Figure 3-3.

Figure 3-3: A sample of faces from the ORL database [68]**.**

One of the experiments, which undertook in this chapter, was to compare our method with that of the state-of-the-art, i.e., comparing the average weights

computed in the Eigenspace against average faces computed in the pixel space. To do this, all faces were cropped to remove irrelevant objects [143], and then warped them to a standard template face [144]. This is the standard method of computing the average face in the pixel space as reported by [33] [48]. Figure 3-4 shows examples of cropped and warped faces that have been produced from the data available within the ORL dataset.
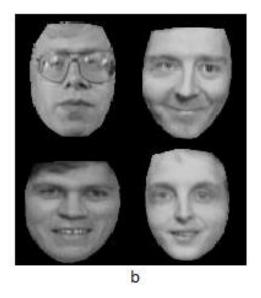
Figure 3-4: A sample of processed images taken from the ORL database. (a) Cropped and (b) warped face images [68].

In order to compute the averages in the pixel space, ten images per subject were used to get it and the remaining cropped images were used for testing. This procedure is illustrated in Figure 3-5.
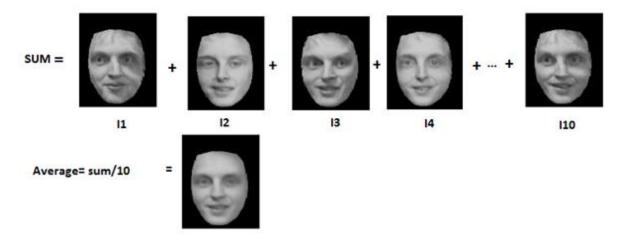
Figure 3-5: An example of an average warped face generated using the ORL dataset [68].

Face96 is a database of coloured facial images and is composed of 152 individuals with 20 images per individual, totalling to 3040 images [58]. The image resolution for the images within the database is 196 pixels by 196 pixels. The database consists of both male and female subject images, with complex backgrounds as well as variations in lighting and facial expressions. Figure 3-6 shows a sample set of images taken from this database.

Figure 3-6: An example of face images taken from the Face96 database [58].

The Grimace database [58], contains 18 individuals with 20 images per subject (male and female). The resolution of the images are 180 pixels by 200 pixels and the images contain a uniform background, with variations in the size of the headshot and facial expression, with little variation to illumination. An example of the images is shown as part of Figure 3-7.

Figure 3-7: A sample dataset from the Grimace database [58]**.**

To test our proposed methodology, as well as to compare ours with the state-of-the-art methods, a number of experiments were conducted. First, an experiment was run to test out the performance of our proposed method of averages in the Eigenspace against the standard method of computing averages in the pixels space. The results show the superiority of our proposed method. Then, testing how varying the number of images used to compute averages in the Eigenspace affects recognition rates. As mentioned earlier, three different datasets were used, namely the ORL, Face96 and Grimace in our experiments. More details of our experiments and results are provided below.

3.1.2.1 Average Weight vs Average Face

Using images prepared from the ORL database, the averages over 10 faces in an incremental manner were computed, both for our method and in the pixel domain. Furthermore, these averages were computed by using image sets, whereby images have been cropped and warped into a standard template (as used in the state-of-the-art methods) and image sets in which no pre-processing or transformation has been applied.

Figures 3-8 and 3-9 show the results from which it is clear that our method outperforms the rest in both scenarios.
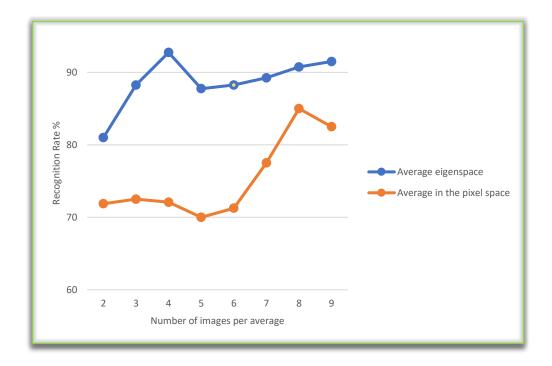
Figure 3-8: Percentage rates (%) of recognition using the average face computed in the pixel space and in the Eigenspace (without cropping and warping).
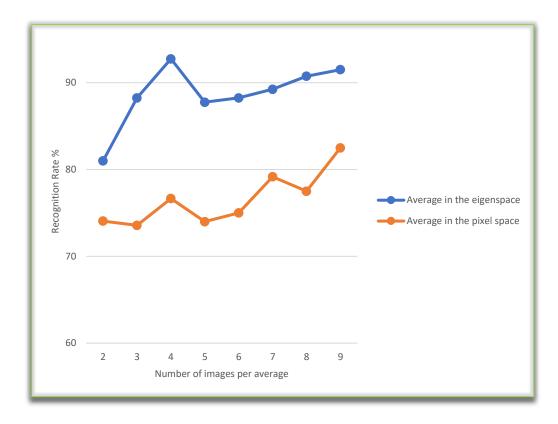
Figure 3-9: Percentage rates (%) of recognition using the average face computed in the pixel space and in the Eigenspace (with cropping and warping).

3.1.2.2 Number of Images Chosen vs Average Weights in Eigenspace for Recognition

In the next set of experiments, test the effectiveness of using average weights of faces in the Eigenspace without any pre-processing and transformation. In particular, setting out to test whether increasing the number of images to form the averages has a positive effect on recognition and also if there is an optimal number.

For the conducted experiments images from three datasets have been used. The average weights were computed in the Eigenspace by varying the image numbers from 2 to 10, as shown in Figure 3-10. Our results demonstrate that the best recognition rate is 92.75% at AW4 meaning that the highest recognition rate is obtained by using an average of only 4 images.
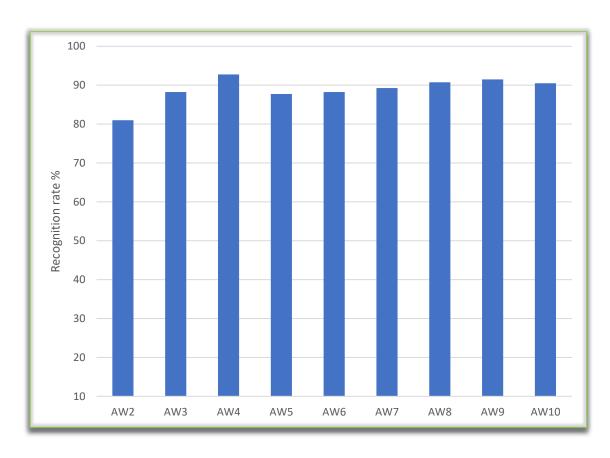
Figure 3-10: Percentage rates (%) of recognition for the ORL database using the average face in the Eigenspace – AW results show for up to 10 images.

Later, the number of used images was increased to form the averages in the Eigenspace even further, i.e. up to 20 and this done for 151 subjects in the database. Again, the results (as shown in Figure 3-11), indicate that increasing the number of images does not directly affect the recognition, as observed, for AW2 the recognition rate is 86.23%, for AW11 it is 78.45%, with the best rate being 90.29% for AW19. Moreover, when AW4 was tested on the full database, by selecting images randomly, the results showed that the rate of recognition is optimal at 92.62%. Thus, our results clearly demonstrate that computing the average weights in the Eigenspace over four images is optimal and beyond that, the rate of recognition declines, and the task becomes computationally more expensive.
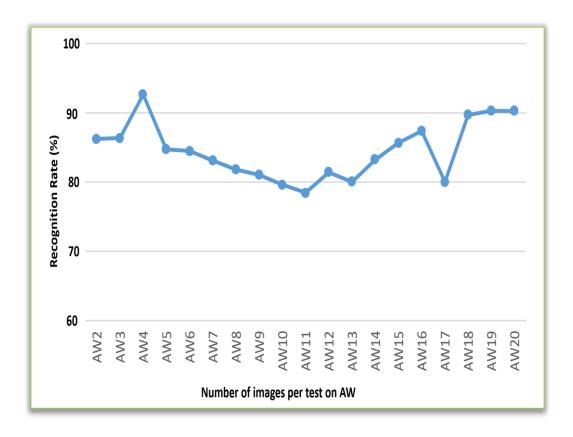
Figure 3-11: Percentage rates (%) of recognition for Face96 and using the average face in the Eigenspace – AW results shown up to 20 images.

With regards to the Grimace database, the experiments were carried out with AW4 and the results report that AW4 achieves 97.78% recognition.

3.1.2.3 Comparison with PCA and PCA+LDA using ORL Database

To further clarify the effect of AW, in the upcoming experiment, combined PCA+LDA [71] was applied to the ORL database, in order to compare the implementation of AW using EVD and PCA+LDA for face recognition. By using different samples of images, training and testing started with the use of four images concluding at nine. The highest result was 92.5 % when applying PCA+LDA and using 8 images for training and 2 for testing (see table 3-1). On the other hand, when implementing AW, the highest result was 92.75% when using 4 images for training and 8 images for testing. It is feasible to suggest that by increasing the number of images per subject during the training phase does

affect the rate of recognition. With regards to the computation process, in PCA+LDA there are two main operations, which take added computations. The first being the computing of a normal PCA (all steps) in order to find PCA subspace and secondly, when the output is passed to LDA as the input data to compute two scatter matrixes and Eigenfaces.

Method	The Number of Images Per Subject					
	4	5	6	7	8	9
PCA	83.75	78	82.5	80.83	85	80
PCA+LDA [71]	91.25	91.5	91.88	91.67	92.5	90
EVD-AW	**92.75**	87.75	88.25	89.25	90.75	91.5

Table 3-1: Comparative recognition rates (%) of our method and other recognition methods on the ORL Database using a different number of images in training sample per person.

3.1.2.4 Discussion

The purpose of this study was to measure face recognition using averages formed over a number of images. In particular, we set out to establish where the average-face formed within the computational Eigenspace would provide any advantage over the present state-of-the-art methods of computing the averages in the pixel space. The results show that our method has clear advantages over other methods. Firstly, in order to compute the average weights in the Eigenspace does not require the pre-processing on images (e.g. separating the face from the image background) and complex transformations (e.g. warping the image to a standard template). Secondly, recognition is undertaken by comparing a handful of weights in the Eigenspace enabling fast and efficient computation.

Our results show that by using a combined dataset taken from 3 standard databases, AW4 (the computing average weights over four images as explained

in section 3.1.2.2 by selecting four weights per subject using equation 3-8) enhances the percentage of recognition significantly, with a result of 92.75%. Further, increasing the number of weights (images) only had a marginal effect to recognition. In our experiments, the average weights were computed by using two images initially sequentially increasing with one image at a time. The experimental results show that when AW2 is used, the recognition rate is 81%, similar to the results obtained with AW3 and AW6. When comparing between other averages, (from AW6 to AW10) there was a slight difference in recognition rate: 87.75%, 88.25%, 89.25%, 90.75%, 91.5%, and 90.5%, respectively. However, the optimal recognition rate achieved is 92.75% were using AW4.

In another experiment, the order of images was changed, to compute the average weights. The results show that the recognition rate could be affected by the quality or type of images used such as variations in the pose of the face or a difference in illumination. For instance, with AW20, in one experiment, the percentage of recognition was 90.26%, but by using a different set of images, the rate increased to 92.61%. Thus, the type of images has an effect on the recognition rate as well as the number of images used to compute an average. Nevertheless, it is clear that the average weight with four images gave the optimal results, instead of using ten images to compute an average face for example.

The work discussed in this section demonstrates that there is great potential in using Eigenspace for computing averages, especially for images. It is possible to use the average weights obtained directly from images, without any complex transformations or manual operations, as a reliable method to undertake efficient face recognition. In addition, it has shown the potential of this method with a demonstrated application to the area of face recognition.

3.2   The Biharmonic Eigenface

Since 1991, when the well-known Eigenface approach was introduced by Turk et al. [32], the method of PCA has played a prominent role in face processing, especially applicable to face recognition systems. Since then, a variety of face recognition algorithms have been proposed centred on PCA [145]. PCA is a powerful mathematical technique for data analysis the most significant task of it is to remove redundant data and to reduce the dimensionality of the data, which is the first step in many image compression applications [146]. Though in its own right PCA is well designed for data reduction in order to reveal interesting patterns in the data, novel modifications or enhancements are likely to make PCA more powerful. Existing literature points to this direction where PCA based subpattern (SpPCA) [147] and modular PCA (mPCA) [148] are considered examples.

Our aim in this work is to add a novel modification to the traditional PCA to further enhance its functioning abilities. To do this, the elliptic Biharmonic partial differential operator is introduced as a pre-processing step into the PCA pipeline. The Biharmonic operator is well-known for its smoothing properties [149], [150] [151]. Thus, postulating that the inherent averaging ability of the discrete Biharmonic operator on images can help reduce the size of images without losing the characteristic features of them. And it is this property of the Biharmonic operator, which is proposed to integrate into the PCA pipeline for creating efficient Eigenfaces, which is referred to as the BiPCA.

There are multiple proposed methods to improve upon for the Eigenfaces of traditional PCA. A question that might spring to the mind is, "why yet another technique to improve PCA?". Our motivation for this work is to develop an

intelligent, computationally less extensive and an easy to implement method, which improves traditional PCA. Though the mathematical formulation of the elliptic partial differential operator proposed in this work appears to be complex, we demonstrate that the discrete Biharmonic operator is intuitive and easy to implement. In addition to this, also showing that our technique of BiPCA is superior in comparison to the existing techniques that can be used to improve traditional PCA.

3.2.1 The Biharmonic Function

Given a closed domain $\gamma \subset R^2$, we seek a 'smooth', differentiable and well-behaved function X such that, $X = (x(u,v), y(u,v))$, where $x(u,v)$ and $y(u,v)$ are the usual Cartesian forms in R^2 spanned by the two parameters u and v. There are various choices for the form of the function X which can be required. In this particular case, X is required to be satisfying a boundary value problem whereby X is the solution to the Biharmonic equation,

$$\left(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial u^2} \right)^2 X(u,v) = 0 \tag{3-11}$$

Prescribing the condition on the boundaries representing the function values and the normal $\frac{\partial x}{\partial n}$, at the boundary $\partial \gamma$. This particular choice for which X is required to be a function of Biharmonic type, is inspired by many interesting properties of the Biharmonic function of which smoothing is of particular interest in the present context [151].

3.2.2 The Smoothing Properties of the Discrete Biharmonic Operator

To understand the smoothing properties of the Biharmonic operator presented in Equations 3-11, let us consider its following variation:

$$\mathrm{F} = \int L \, du \, dv, \tag{3-12}$$

where the integrand L takes the form,

$$\mathrm{L} = \mathrm{L}(X, \; X_u, X_v, \; X_{uu}, \; X_{vv}, \; X_{uv}, \dots, \; X_{uuuu}, \; X_{vvvv}) \tag{3-13}$$

Now, taking the Euler–Lagrange form of Equation 3-12 and requiring the first variation of L to tend to zero, it can consider,

$$\mathrm{L} = \left| \frac{\partial^n X(u,v)}{\partial u^n} \right| + \left| \frac{\partial^n X(u,v)}{\partial v^n} \right| \tag{3-14}$$

where $n = 2$ implies the Biharmonic case.

To consider the smoothing properties associated with the various elliptic operators arising from (3-14), one can consider $n = 1$ and 2. In the case of $n = 2$, the resulting operator is simply $\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial u^2}$ which is nothing but the Harmonic or the Laplace operator. If we now look at the discrete finite differences associated with this operator, it can be expressed as,

$$X(i,j) = \frac{1}{4} \left[X_{i+1,j}, X_{i-1,j}, \; X_{i,j+1}, \; X_{i,j-1} \right], \tag{3-15}$$

where $X(i,j)$ is the resulting finite difference approximation of the point on (i,j) on a finite difference grid [149] [59].

As one can infer from Equation 3-15, and Figure 3-12(a), applying the Laplace operator results in the arithmetic mean of the immediate neighbouring points corresponding to the north, south, east and west.

Similarly, considering the fourth-order operator, $\frac{\partial^4}{\partial u^4} + \frac{\partial^4}{\partial u^4}$, the resulting finite difference scheme can be represented as shown in Figure 3-12(b), where the averaging of the point under consideration is obtained from a more dispersed

region, though the averaging still occurs along with the parametric directions (north, south, east and west).
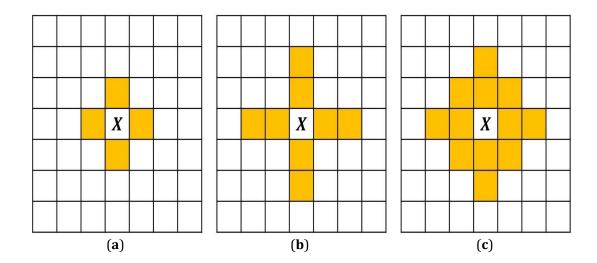
Figure 3-12: Smoothing masks arising from the discrete form of the elliptic operators: (a) The Laplacian, (b) the 4th order PDE operator and (c) the Biharmonic operator.

On the other hand, as considering the Biharmonic operator as shown in Equation 3-11, the average can be obtained from a more uniform region surrounding a given point, as shown in Figure 3-12(c). This is due to the fact that the Biharmonic operator not only has the 4th order derivative information but also contains the corresponding mixed derivative information.

3.2.3 Eigenfaces of Biharmonic Type

As shown in the previous section, the discrete Biharmonic operator provides a very useful averaging process whose exact form is shown in Equation 3-16. This operation is applied to a given set of images as a pre-processing step before PCA. Within the pixel space, the BiPCA can be considered as a linear pixel operation process and can be considered to be equivalent to a pooling layer within a standard Convolutional Neural Network [152]. Moreover, the BiPCA

process can be regarded as a type of down sampling process in order to reduce an image representation, using a filter bank without any complex weights and no associated coefficients. For the purpose of this work, a mask of 3x3 is used without an overlap between any two regions of the image space for our BiPCA. Thus, the discrete Biharmonic mask $B(i,j)$ can be described as,

$$B(i,j) = (20(X_{i,j}) - 8(X_{i+1,j} + X_{i-1,j} + X_{i,j+1} + X_{i,j-1}) + 2(X_{i+1,j+1} +$$ (3-16)

$$X_{i-1,j+1} + X_{i+1,j-1} + X_{i-+1,j-1}) + (X_{i+2,j} + X_{i-2,j} + X_{i,j+2} +$$

$$X_{i,j-2}))$$

$$K = B(i,j) \oplus I_{resized}$$ (3-17)

where $\oplus$ is a mathematical operator for convolution between two arrays.

Suppose there is an image I with a size of 5x6 pixels, as shown in Figure 3-13(a). The first step, in this case is to resize the image so it is equivalent to a square image of 6x6 ($I_{resized}$) as shown in Figure 3-13(b). This operation becomes especially useful when handling the boundary of an image. Assuming the mask that is used in the convolution is B, with a size of 3x3, the convolution of the Biharmonic operation can be defined using Equation 3-17.
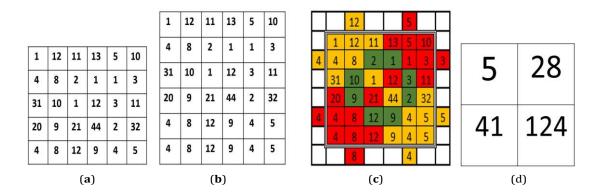
Figure 3-13: Operational procedure to produce a square image: (a) The original (non-square) image and, (b) the resulting (square) image after resizing. The application of the discrete Biharmonic to an image; (c) Representation of how the Biharmonic mask is applied to the image and (d) the representation of the resulting image.

Consider the example shown in Figure 3-13(c), if taking the Biharmonic mask B in Equation (3-16), the resulting image K is the reduced representation as shown in Figure 3-13(d). Now, for a given set of images in a dataset, the above procedure can be performed and PCA can be applied as described in the previous section 3.1.1.

3.2.4 Experiments and Results

This section presents the experimental setup to measure the capacity of the BiPCA approach in face recognition. In particular, looking at standard face recognition as an example, whereby comparing the BiPCA approach against the traditional PCA approach. BiPCA method is also compared with the Shurfaces method as discussed in [64]. For the purpose of experimentation, three publicly available datasets are utlized; the ORL [68], and Face95 and Faces96 [58], more details in section 3.1.2.

The experimental approach comprised of taking images for the purpose of training beginning with 1 image and increasing it to 5 images per individual. The remainder of the images for each individual were used in the testing phase. The Euclidean Distance (*ED*) value was computed using the formulation defined in Equation 3-18, which is taken as a similarity measure. *ED* is a measure of similarity between two non-zero vectors of the Minkowski family [142]. For each experiment, a ten-fold cross validation have been utilised, whereby the average percentage error rates from the cross validations are reported. Also noting the computational time required in each case. A detailed discussion relating to the conducted experiments is presented in the upcoming section.

$$ED(\Omega_e, \Omega_i) = \sqrt{(\Omega_e - \Omega_i)^2} \tag{3-18}$$

where Ω_e The corresponding BiPCA weights and Ω_i are the training weights.

3.2.4.1 BiPCA vs PCA and Shurfaces using the ORL Dataset

In this experiment, BiPCA, Shurfaces and the traditional PCA approach for face recognition is applied on images taken from the ORL dataset. Training sequences were run by beginning with 1 image and ending with 5 images per subject. Table 3-2 lists the computational time and error rate of PCA, Surfaces and BiPCA for face matching based on 10-fold cross validation for each of the training image sets. The error rates (using equation 3-19) in all the experimental cases are much lower for the BiPCA framework indicating that the BiPCA method outperforms the traditional PCA method.

$$Error\ rate = \frac{number\ of\ mismatching\ samples}{total\ number\ of\ samples} \times 100 \tag{3-19}$$

Faces	PCA		Shur		BiPCA	
	Error	Time	Error	Time	Error	Time
1	37.5	0.09	37.0	0.09	35.3	0.04
2	29.1	0.14	20.0	0.15	22.2	0.07
3	23.6	0.17	23.0	0.19	13.6	0.09
4	20.8	0.21	20.8	0.21	14.2	0.10
5	15.0	0.22	15.0	0.21	11.0	0.11

Table 3-2: A Table of Comparison for the face matching experiments using PCA, Shurfaces and BiPCA on the ORL dataset. The error rates are shown as a percentage (%) and the computational time is shown in seconds.

3.2.4.2 BiPCA vs PCA and Shurfaces using the Face95 Dataset

In contrast to the first experiment, in this case, two tests were run whereby in the first case the PCA, Schurfaces and the BiPCA were applied on the uncropped Face95 images, and in the second case, the same was done on

cropped images. Again, in both cases, the training routines were run on images of individuals starting from 1 image to 5 images per subject. The remaining images in each case were used for testing and the results of the 10-fold cross validation are reported, see Table 3-3 and 3-4. Again, as previously demonstrated the BiPCA approach provides a functional advantage above the traditional PCA for both cropped and uncropped images.

Faces	PCA		Shur		BiPCA	
	Error	Time	Error	Time	Error	Time
1	58.7	0.47	58.6	0.47	55.6	0.24
2	44.9	0.89	45.0	0.89	35.7	0.45
3	36.3	1.26	36.3	1.26	37.1	0.67
4	34.1	1.45	34.1	1.41	24.5	0.88
5	29.8	1.57	29.7	1.52	21.7	1.00

Table 3-3: A table of comparison for face matching. The percentage error rates and computational time (in seconds) between PCA, Shurfaces and BiPCA on the full (uncropped) faces in the Face95 dataset.

Faces	PCA		Shur		BiPCA	
	Error	Time	Error	Time	Error	Time
1	59.9	0.47	60.0	0.47	50.0	0.24
2	29.7	0.90	29.7	0.89	22.7	0.48
3	22.3	1.31	22.2	1.30	15.9	0.70
4	19.5	1.44	19.4	1.47	12.9	0.85
5	18.0	1.54	17.8	1.53	11.3	1.02

Table 3-4: A table of comparison for face matching. The percentage error rates and computational time (in seconds) for face matching between PCA, Schurfaces and BiPCA on the cropped faces in the Face95 dataset.

3.2.4.3 BiPCA vs PCA & Shurfaces using the Face96 Dataset

Contrary to the face-matching results obtained from the ORL and Face95 datasets, the results for the original (uncropped) face images taken from Face96, demonstrate that the BiPCA method does not provide favourable results, see Table 3-5. However, the results shown in Table 3-6 indicate that the BiPCA

method outclasses the traditional PCA approach when applied to cropped images.

Faces	PCA		Shur		BiPCA	
	Error	Time	Error	Time	Error	Time
1	30.4	2.55	30.4	2.09	37.4	1.03
2	17.5	3.43	17.5	3.43	22.2	2.62
3	14.8	4.44	14.7	4.45	19.2	3.30
4	12.3	5.87	12.3	6.02	17.4	4.56
5	11.4	11.4	11.4	7.22	16.0	5.65

Table 3-5: A table of comparison for the face matching experiments using PCA, Shurfaces and BiPCA on full-face (uncropped) images, taken from the Face96 dataset. The error rates are shown as a percentage (%) and the computational time is shown in seconds.

Faces	PCA		Shur		BiPCA	
	Error	Time	Error	Time	Error	Time
1	57.9	2.06	57.5	2.12	54.3	1.07
2	35.7	3.48	35.7	3.46	29.0	2.15
3	30.2	4.47	30.0	4.47	24.0	3.26
4	26.5	5.54	26.5	5.56	21.7	4.44
5	24.0	6.62	23.7	6.62	19.4	5.62

Table 3-6: A table of comparison for face matching. The percentage error rates and computational time (in seconds) between PCA, Shurfaces and BiPCA for cropped faces in the Face96 dataset.

3.2.5 Discussion

The primary aim of the experiments presented within this section was to investigate whether the addition of a discrete Biharmonic operator as a pre-processing step, would demonstrate an advantage, over the use of traditional PCA. From the results above, it is apparent that in the case where images are taken from the ORL dataset, the BiPCA method gives a lower error rate consistently. For the experimental scenario where 5 images have been used per individual during training, the percentage error rate declined to 11% for the BiPCA

approach, in comparison to 15% for PCA. Further, when using images from the Face95 dataset, the results report a low error rate when using the BiPCA approach, for both cropped and uncropped face images. Though the error rates are much lower in the case of cropped face images.

The reported results for images from the Face96 database (as shown in Tables 3-5 and 3-6) appear to be mixed. Noteworthy, for the original (uncropped) images the BiPCA method performs poorly compared to the traditional PCA. This result, along with the results for the uncropped images taken from Face96, (though in this case the BiPCA still outperforms PCA) suggest that the BiPCA approach is still very sensitive to background noise. This is particularly obvious for uncropped images from Face96 since the faces in these images are relatively small compared to the background. Hence, discrete Biharmonic averaging must be separating some of the prominent facial features across the images resulting in the loss of sufficient detail, which must be available for accurate face matching.

As previously mentioned, an important advantage of using BiPCA is the ability to reduce image dimensionality an inherent ability to preserve features. To demonstrate this further a sub experiment based on face recognition using state-of-the-art Convolutional Neural Network, was conducted. The well-established VGGF [41] face model was used for features extraction and a Linear Support Vector Machines [153] for classification.

In this experiment, the FEI database [22] was used, which contains 200 subjects with 14 face images per individual. Four frontal images per subject were selected and artificially enhanced the data sample by the means of augmentation. The average Biharmonic face was generated using 8 images. Then, the resulting Biharmonic average images when used for training gave a recognition rate of 87.2%. To compare this result, one frontal image was chosen and augmented it

with 20 parts. These facial parts were added along with the average Biharmonic face to the training set. The recognition rate in this experimental condition case was 85.2%, lower than the result for one average Biharmonic face. Furthermore, the same 20 partial images were used with one original image in the training and the recognition rate concluded was 81%, see Table 3-7. It can be inferred that Biharmonic pre-processing provides a distinct advantage in terms of reducing the number of images at the image pre-processing stage.

Method	Training	Test	Rate (%)
1 BiPCA	200	600	87.20
1 image + 20 parts	4200	600	81.00
1 BiPCA + 20 parts	4200	600	85.20

Table 3-7: A table of comparison for the task of face matching using CNN-based recognition using different three training datasets: (1) one average Biharmonic face, (2) one original and 20 facial parts and (3) one average Biharmonic with 20 facial parts.

3.2.6 Conclusion

In this chapter, two approaches have introduced to use the average concept. In the first method, the concept of average weights in the Eigenspace have been introduced for efficient face processing. It has shown that the average weights in the Eigenspace, for a face, are a reliable measure of face recognition. Also, the use of average weights obtained directly from images, without any complex transformations or manual operations, is a trustworthy technique to undertake efficient face recognition. Further, shown that as low as four images can be classed the optimum number of facial images that are required for computing averages in the Eigenspace. Especially for fast and efficient face recognition when compared to the state-of-the-art methods which use averages in the pixel space.

In the second method in this chapter, a novel modification to the method of PCA is introduced which functions to enhance its ability for image dimensionality reduction. Our proposed approach hinges on the utilisation of the discrete Biharmonic operator as a pre-processing step for PCA, which referred to as BiPCA. The Biharmonic operator is among the family of elliptic operators with intriguing mathematical properties, one of which is smoothing / averaging. In this work, this characteristic of the Biharmonic operator has explored and applied to images to efficiently process them. The resulting images are of lower dimensionality where subsequently reduce computational complexities while preserving features of interest.

To demonstrate the capacity of the proposed BiPCA approach, standard face recognition was used as an example. For experimentation, publicly available datasets were utilised, namely, the ORL, Face95 and Face96. The obtained results have demonstrated that the BiPCA outperforms traditional PCA functions. In fact, our experiments suggest that the BiPCA method has at least a 25% average improvement in the error rates when considering face recognition.

While the BiPCA approach as a pre-processing step for reducing image dimensionality reports good results, there are acknowledged drawbacks. For example, when the BiPCA approach applied to face images with very complex backgrounds, it seems to provide poorer results. This is not surprising as the BiPCA performs complex averaging through the elliptic Biharmonic operator. However, in practical terms, at least in the case of face recognition, this should not pose a major issue since in most cases the faces must be cropped to separate them from any complex background. And in such cases the BiPCA approach clearly outperforms.

To show the ability of the BiPCA and the potential use of elliptic operators in image-processing and enhancement, the Eigenface and face recognition have been used as examples. The application domain of the proposed BiPCA approach is not limited and can be used in any image-processing pipeline where some form of dimensionality reduction is required as a pre-processing step. For example, in the context of machine learning (as discussed in the Section 3.2.5) the BiPCA approach can be of used to reduce the number of images in the training set of a Convolutional Neural Network.

While the BiPCA approach has been shown to perform superiorly over traditional PCA, it must be stressed that the elliptic operator associated with the Biharmonic function gives rise to a complex averaging process. Therefore, its use must be carried out with background knowledge of the problem domain and with an understanding into the context of the problem.

4 Deep Face Recognition using Partial Faces

Face recognition is an inherently powerful human ability. In this respect, there are special areas in the human brain dedicated to face processing and face recognition [6]. However, there are known limitations to the human brain with regards to face recognition, especially when tackling the problem of identifying individuals from large datasets, containing unfamiliar faces. On the other hand, more recently, machine based algorithms are proving to be far better placed in successfully addressing this problem.

Face recognition using partial facial data is a particularly challenging problem. This is evident (clear) from human based face stimuli experiments. For example, apart from familiar faces, humans struggle to discriminate between similar faces from individual parts of a whole face such as the eyes, mouth and nose [10], see Figure 4-1.

Figure 4-1: An example (random Images taken from internet) of how partial faces are presented as input probe images for face recognition tasks.

Figure 4-1 illustrates typical examples of how face recognition can be applied for partial facial data analysis. As the example illustrates, full face data is not available and only regions of the face are provided subject to availability, as an input probe.

The practical application of this work is borne out by the increasing need to undertake automated practical face recognition tasks in everyday scenarios. Similar to other biometric authentication tools (e.g. fingerprints), face recognition and face perception have become common practice [154] [155]. Reliable automated face processing [156], [157] and recognition [95]–[98] tools can utilise images from CCTV cameras, and this is readily becoming of paramount importance. Figure 4-2 illustrates the overview of our face recognition framework, which uses partial faces as probes.
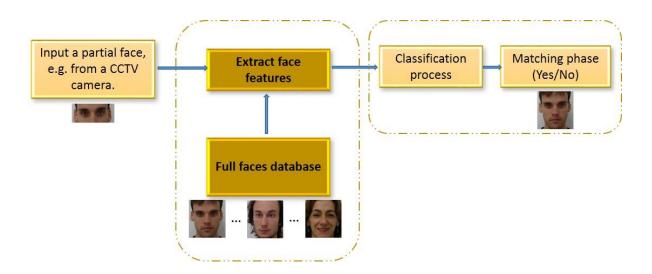
Figure 4-2: An overview of our computational framework for face recognition using partial faces as probes. There are two main stages in this diagram, in the first step all features are extracted from faces in both saved data and an input image. Then the input image is classified in the matching stage.

Based on the work previously conducted using machine-based face recognition, (Chapter 2), a gap in the research field has been identified, since published studies have not investigated how Machine Learning performs in face recognition, using partial face data. Our aim in this study is to close that gap and address the challenge of partial face recognition. Investigating how different regions of the face perform during recognition, and also studying the effects of rotations and zooming at various levels, for the task of recognition within a Machine Learning scenario. Convolutional Neural Network (CNN) based architecture is used along with the pre-trained VGG-Face model to extract features. Then, two classifiers namely the Cosine Similarity (CS) and the Linear SVMs are used to test the recognition rates.

The rest of the chapter is organised as follows. Section 4.1 explains the CNN architecture we have utilised along with a brief description of the VGG-Face model used for feature extraction and the CS and SVMs models, for classification. Section 4.2 presents the experimental framework, which tests partial face data, and reporting the obtained results. Section 4.3 discusses the results. Lastly, providing a conclusion of the chapter.

4.1 Methodology

One of the most popular examples of Machine Learning in recent times are based on deep learning, otherwise known as Convolutional Neural Networks (CNNs). CNNs are supervised Machine Learning algorithms that can extract "deep" knowledge from a dataset through rigorous example based training. Such a Machine Learning approach mimics the human brain as it learns, and has been successfully applied to feature extraction, face recognition, classification, and

segmentation tasks. The increased use of CNNs in recent times is due to their ability to learn complex features using nonlinear multi-layered architectures [158]. While the origin of CNN dates back to the early 1990s, the predominant scepticism for using CNN was based on the assumption that feature extraction using gradient descent will always overfit. The main argument for this has been that gradient based optimisation methods are notorious for getting stuck in the local minima. However, in recent times, these assumptions have been overturned due to the promising results CNNs have produced across multiple domains of research. Thus, today, state-of-the-art deep learning models, based on CNN architectures are being used in almost all visual computing related domains. Examples include image perception [159], recognition [160], classification [161], and information retrieval [162].

Generally, there are three ways of deploying CNNs; (1) training a network from scratch, (2) fine-tuning an existing model, or (3) using off-the-shelf CNN features. The latter two approaches are referred to as transfer learning [163]. It is important to highlight that training a CNN from scratch requires an enormous amount of data, which is often a challenging task [164]. On the other hand, fine-tuning involves transferring the weights of the first few layers learned from a base network to a target network. The target network can then be trained using a new dataset.

For face perception work by CNN, there are several pre-trained models which can readily be utilised for feature extraction, e.g. VGGF, FGG16, VGG19 and OverFeat [41]. For the purpose of our study, the VGGF pre-trained model has been utilised for feature extraction. The adopted methodology uses the pre-trained VGGF model for feature extraction which is then followed by CS [165] or

Linear SVM for classification. Figure 4-3 illustrates an overview of our feature extraction process.
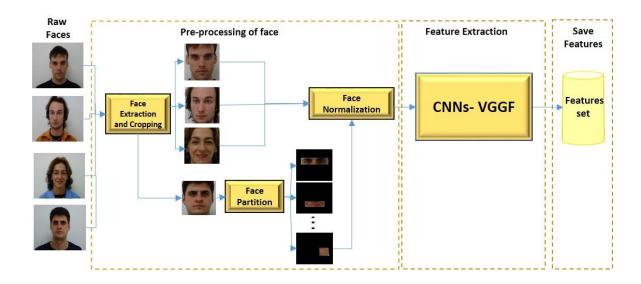
Figure 4-3: An illustrated example of our experimental framework which uses Convolutional Neural Network (VGG-Face) for feature extraction and facial recognition using partial face data.

4.1.1 VGG-Face Model

One of the most popular and widely used pre-trained CNN models in face recognition is the VGGF model, which was developed by the Oxford Visual Geometry Group [41]. The model was trained on a large dataset of 2.6 million face images of more than 2.6 thousand individuals. The architecture of VGG-F consists of 38 layers starting from the input layer up to the output layer. As a fixed criterion, the input should be a colour image of 224 by 224 dimensions, and as the pre-processing step, an average is normally computed from the input image.

In general, the VGG-F contains 13 convolutional layers, each layer having a special set of hybrid parameters. Each group of convolutional layers contains 5 Max-Pooling layers and 15 Rectified Linear Units (ReLUs). After which there are three Fully Connected layers (FC) namely FC6, FC7 and FC8. The initial two have 4096 channels, while FC8 has 2622 channels, and are used to classify the

2622 identities. The last layer is the classifier is a softmax layer which functions to classify an image. The basic architecture of VGG-Face in general is represented in Figure 4-4.
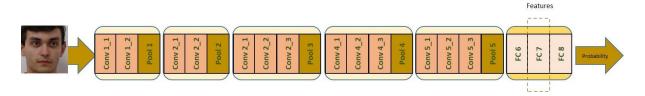
Figure 4-4: An illustration of the basic architecture of VGG-Face model.

4.1.2 Feature Extraction

Given an input image, X_0, it can be represented as a tensor $X_0 \in R^{HWD}$, where H is the image height, W is the width and D represents the colour channels. A pre-trained layer L of the CNN can be expressed as a series of functions, $g_L = f_1 \rightarrow f_2 \rightarrow \cdots \rightarrow f_L$.

Let $X_1, X_1, \ldots, X_L$ be the outputs of each layer in the network. Then, the output of the ith intermediate layer can be computed from the function f_i and the learned weights w_i such that $X_i = f_i(X_{i-1}: w_i)$.

As known, CNNs learn features through the training stage and use such features to classify images. Each convolutional (Conv) layer learns different features. For example, one layer may learn about entities such as edges and colours of an image while further complex features may be learnt in the deeper layers. For example, a result of conv layer involves numerous 2D arrays which are called channels. In VGG-F, there are 37 layers, 13 of which are convolutions and the remaining layers are mixed between ReLU, Pooling, Fully Connected and the Softmax. However, after applying the conv5_3 layer to an input image, which has 512 filters with size 3x3, the features can be extracted for a

classification purpose. By examining the activations of that layer, one can obtain the main features as shown in Figure 4-5, where a sample of the features are presented.
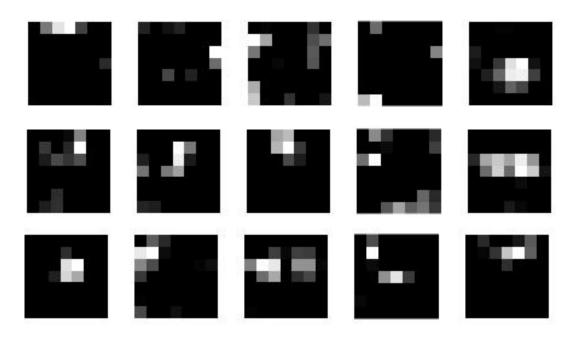
Figure 4-5: An example of the features retrieved from the conv5_3 layer of the VGG-Face model.

In order to decide the best layer within the VGGF model to utilise for facial feature extractions, one must conduct a number of trial and error experiments by testing layers 34 (FC7) through to 37. In our experiments, other layers were tried, however the best results derived from layer 34. It is noteworthy that this layer is the fully connected layer and is placed at the end of a Neural Network, which means the extracted features represent the whole face. Moreover, Layer 34 (FC7) (the last layer before the SoftMax) was used for feature extraction because this layer is a part of the last block in the VGGFace model, and this layer contains the most important details about an image [166] [41]. Based on this layer, VGGFace model can compute a probability for each class in a training set. Many works in literature [167] [168] [45] were used this layer for the same purpose.

The features from layer 34 are the results that arise from the fully connected layer FC7 after applying 'ReLU6', which gives a vector of 4096 dimensions. The reason for suggesting that layer 34 (FC7) is the best layer, is inferred as a result of undertaking a number of face recognition tests where we used the full frontal-face image for both (training and testing), and obtained a recognition rate of recognition 100%. The entire process of training and testing through feature extraction is described further in Algorithm 1.

Algorithm 1: Feature Extraction – from the Face dataset

Input: Training set M, with m classes
n_j = number of images in a given class
for $i = 1\ to\ m$ **do**
   for $j = 1\ to\ n_j$ **do**
      $im \rightarrow read\ an\ image$
      $im \rightarrow resize(im)$
      $im \rightarrow normalize(im)$
      $im\ features \rightarrow ExtractFeatures(CNNs(im))$
   end for
end for

4.1.3 Feature Classification

Classification in a supervised Machine Learning function, which assigns new observations to a known target category. In other words, the objective of classification is to build a brief model of the distribution of class labels in terms of predicted features. There are several techniques for the classification; decision trees [169], K-NN [93], and SVM [170] to name a few.

As part of the current work, all extracted features in both the training and testing phases were used for the purpose of classification. In our experiments, the Cosine Similarity (CS) [165] and Linear SVMs have been utilised for classification [170]. There are two reasons for selecting two classifiers, firstly, other classifiers were tested and the best results were achieved when using CS

and SVMs [45] [171] [172]. Secondly, through our experiments and analysis, we discovered that these two classifiers have an ability to separate data more accurately (see Figure 4-22).

The Cosine Similarity is a measure between two non-zero vectors. It uses the inner product space to measure the cosine of the angle between those two vectors. The Euclidean dot product formula see in Equation 4-1 can be used to compute the cosine similarity such that,

$$a.b = ||a|| \, ||b|| \, cos \, \theta, \qquad (4\text{-}1)$$

where, $a \, and \, b$ are two vectors and θ is an angle between them.

By using the magnitude or length, which is the same as the Euclidean norm or the Euclidean length of vector $x = [x_1, x_2, x_3, ..., x_n]$ as in Equation 4-2, the similarity S is computed using the formulation given in Equation 4-3 such that,

$$\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2}, \qquad (4\text{-}2)$$

$$S = \cos \theta = \frac{A.B}{\| A \| \| B \|},$$

$$= \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}, \qquad (4\text{-}3)$$

where, $A \, and \, B$ are two vectors.

For classification in all our experiments the CS is computed to find the minimum "distance" between the test image, $test_{im}$ and training images, $training_{im}^n$ by using Equations 4-4 and 4-5. The procedure for this classification is further illustrated in Figure 4-6.

$$M_{CS} = min \left(CS(test_{im}, training_{im}^n) \right), \qquad (4\text{-}4)$$

where, im is an image number and n is total images in the training set and,

$$CS(test_{im}, training_{im}^{n}) = \frac{\sum_{j=1}^{m} training_{jm}^{i} \ test_{im}}{\sqrt{\sum_{j=1}^{m} training_{jm}^{i2}} \sqrt{\sum_{j=1}^{m} test_{im}^{2}}},$$

(4-5)

where, m is a length of vector.
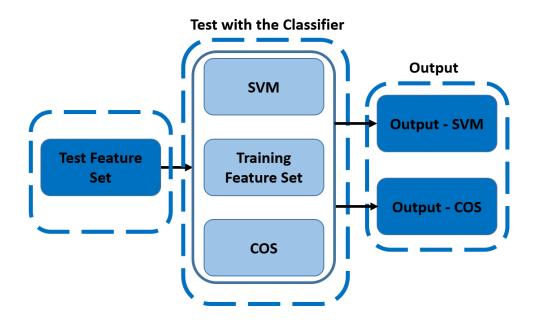
Figure 4-6: A procedure adopted for the purpose of classification.

SVMs are a supervised Machine Learning algorithm, which can be used for both binary classification and multi classification problems. The SVMs focus on identifying the "margins" via hyperplanes to separate the data into classes. Maximising the margin reduces the upper bound on the expected generalisation error by creating the largest possible distance between the separating hyperplanes. It is clear that the SVMs are geared to solve binary classification problems. In all our experiments, the Linear SVMs are used to solve the multi-classes classification problem based on One-vs-One (OVO) approach [173]. This is also known as pairwise classification [174]. The OVO decomposition constructs $n(n-1)/2$ binary classifiers for a given n number of classes. Then, for a final

decision, the Error Correcting Codes (ECC) combination approach [175], [176] decides how the various classifiers can be combined.

Suppose there is a training dataset (x_i, y_i), the Linear SVMs can be used as in Equation 4-6 such that,

$$\min_{w \in R^d} \frac{1}{2} \|w\|^2 + C \sum_i^N \max(0, 1 - y_i w^T x_i),$$

(4-6)

where, w is a weight vector, N is a number of classes and C is the parameter to control the trade-off between the slack variable penalty and the size of the margin.

In equation 4-6, there are two parts, the first part is an optimization problem, which uses to find an optimal separating hyperplane (creates the maximum margin) by solving $\min_{w \in R^d} \frac{1}{2} \|w\|^2$ [177] [178]. The second part of equation 4-6 is Linear SVMs $\sum_i^N \max(0, 1 - y_i w^T x_i)$, which aims to apply linearly separable hyperplane by maximizing the margin between classes [177] [178].

As for the type of SVMs (between a Linear and Non-Linear kernel), we have opted with the Linear SVMs. The reason being that prior to running the main experiments, some preliminary experiments were conducted to test the accuracy of recognition results between Linear and Non-linear SVMs (with radial basis function kernels). In general, it is clear that the Linear SVMs work well when small parts of the face are used as probes. For instance, in one trial experiment involving faces from 60 subjects, it is clear that for the right cheek, the Linear SVM produced a recognition rate of 24.44%, while the kernel SVMs only produced a rate of 2.77% by radial basis functions. In addition to this, in general, the Linear SVMs were computationally more efficient in all the conducted trial experiments. Thus, as a conclusion, the kernel SVMs with their marginal gains (specific only to larger portions of the face such as half or 3/4 face) does not lend

overall additional advantages. Hence, the Linear SVMs is used throughout the rest of the experiments.

4.2   Experiments and Results

Here, presenting a comprehensive series of experiments that have been conducted for face recognition using different parts of the face. To undertake this work, face images from two popular face datasets have been utilised, namely, the FEI [22] and LFW [107]. All images from both of the databases were cropped to remove the background as much as possible using a cascade object detector in order to extract face and the internal facial features [179]. However, for some images with very complex backgrounds, (as in case of the LFW database), the faces from those images were cropped manually. In this work, numerous settings of occlusion have been carried out in order to demonstrate that our proposed methodology is robust for the normal and occluded face recognition tasks. For this purpose, two main types of experiments were conducted; (1) full-face images free of rotation and zoom effects as part of the training face data and (2) partial face images with effects of rotation and zoom. In each case, 14 sub experiments were undertaken, which involving partial, rotated and zoomed-out faces using both classifiers (CS & SVM). For the purpose of training, 70% of the images per subject were utilised and augmented through operations such as padding and flipping. The remaining 30% of the images were used for testing in each case.

In all experiments, the proposed method was evaluated by finding a recognition rate. This was done by counting a number of corrected matched images as shown in equation 4-7 below;

$$Recognition\ rate = \frac{number\ of\ matching\ samples}{total\ number\ of\ samples} \times 100 \tag{4-7}$$

4.2.1 The FEI Dataset

This database contains 200 subjects of Brazilian students and staff of an equal number for males and females. For each subject, there are 14 pictures and in total 2800 images (200 subjects X 14 images per subject). The resolution of the images is 640 pixels by 480 pixels and all the images are in colour taken against a white homogeneous background. The subjects are between the ages of 19 to 40 years old. And the dataset contains images displaying variations in facial expressions and pose. Figure 4-7 shows some sample images.

Figure 4-7: Sample face data from the FEI dataset [22]**.**

4.2.1.1 Experiments on Parts of the Face using the FEI Dataset

In our experiments twelve test sets were generated with each test corresponding to a specific part of the face. The parts of the face that were tested included the eyes, nose, right cheek, mouth and the forehead. Also, faces were generated using just the eyes and nose, the bottom half of the face, the top half

of the face, right half and a three quarter of the face, in addition to the full face.

Figure 4-8 shows the facial parts that were used for testing the recognition rates.
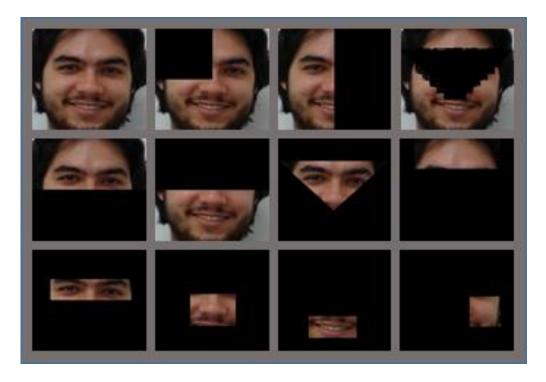
Figure 4-8: Parts of the face were used for testing the recognition rates on the FEI dataset [22]**. This image is a smile expression. The first row and the half of the second row include big parts of the face. Whereas the rest rows represent the small parts.**

After extracting features from the VGG-Face model, the CS without parts (CS-Wo) and the Linear SVMs (with 19900 binary classifiers) without parts (SVM-Wo) were applied in order to investigate the rate of recognition for each facial part separately. The results of these experiments are summarised in Figure 4-9. By visually analysing the figure, the highest recognition rate is achieved when using the full face and the three-quarter view of the face since a recognition rate of 100% is achieved when using both classifiers. However, the recognition rate starts to slightly decrease when using the right half and the top half of the face respectively with SVM-Wo, but in the case of CS-Wo, the rate remains at 100%. As approaching the bottom half of the face, the rate of recognition decreases

further to approximately 50% in the case of SVM-Wo and about 60% for CS-Wo. This decline continues until nearly 0.5% for the right cheek.

To measure the rate of recognition using parts of the face, the previous procedure was repeated, but this time by adding the individual parts of the face into the training set also. As shown in Figure 4-9, the recognition rates significantly improve. For instance, while the results from the right cheek previously were nearly 0%, the recognition rate increased to 15% when using both the classifiers. Also, in the case of the combined eyes and nose features, recognition was previously at 22% for SVM-W and 40% for CS-W. However, in this case the recognition improves to approximately 57% for SVM-W and 90% when using CS-W. However it is noticeable that not all recognition rates steadily increase. In some cases, the results were slightly worse for SVM-W. For example, a slight decrease in the recognition rate was observed when the bottom half of face was tested, a recognition rate of 53% was achieved but decreased to 51%. In contrast, the CS-W approach has produced a significant improvement, for example the recognition rate for combined eyes and nose increased from 40% to 90%.
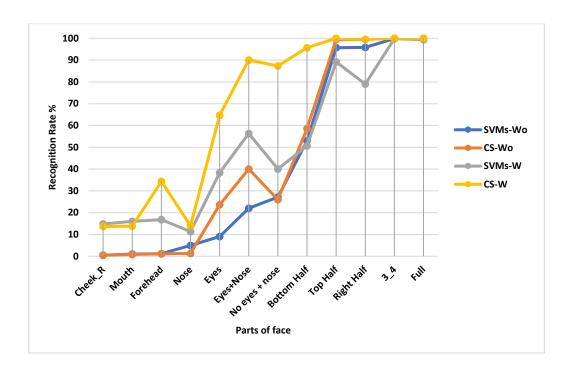
Figure 4-9: Face recognition rates (%) for partial face (with and without rotation) data taken from the FEI database using SVM and CS classifiers

4.2.1.2 Experiments on Rotated Faces using the FEI Dataset

In this experimental section, all the faces in the test sets were rotated in eighteen degree increments, starting from $10°$ to $180°$. Figure 4-10 illustrates some sample rotations.
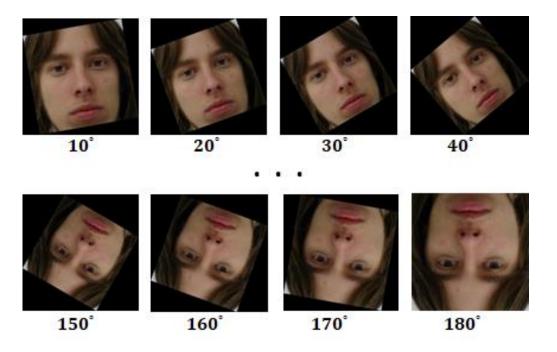
Figure 4-10: Illustration of rotation (10° to 80°) used for the FEI dataset [22]**.**

All rotated images in each subset were passed to the VGGF model to extract features following the same experimental procedure as described previously. Figure 4-11 shows the recognition rates for the rotations with using the two classifiers. The experiments include the addition of rotated facial data into the training set, as well as testing without rotated images.
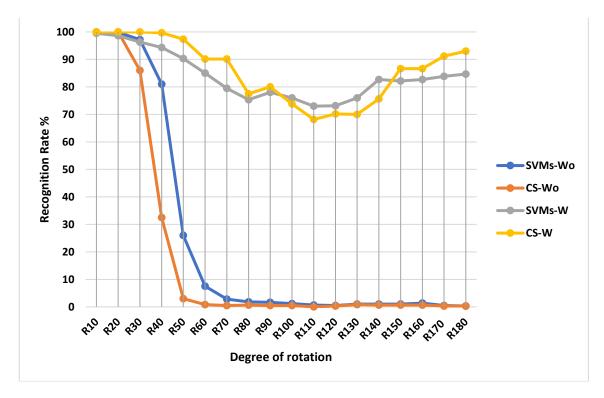
Figure 4-11: Face recognition rates (%) using SVMs and CS classifiers based on face rotations. Without and with rotated faces in the training set for the FEI dataset.

When considering the data without rotated images (SVM-Wo and CS-Wo), it is clear that the rotated faces at $10°$ and $20°$ respectively, show the highest recognition rate at 100% for both classifiers, see Figure 4-11. On the other hand, the recognition rate starts to decrease partially at $30°$ and $40°$ and the rate of recognition achieved is between 98% and 80% for SVM-Wo and CS-Wo. However, the lowest rate of recognition emerges when the degree of rotation is high ($50°$ to $180°$) since the rate of recognition reaches almost 0% in some cases for both the classifiers.

In the second experimental condition, where rotated images are added to the training set, it can observe that the recognition has improved significantly for both classifiers (SVM-W and CS-W), as shown in Figure 4-11. CS-W recorded the highest recognition rate in most of the cases. For example, at $40°$ the previous rate was nearly 33% and enhanced to an impressive 95% with CS-W. Using the SVM-W also gradually increased the recognition rate especially for higher degrees of rotation. For example, at $80°$ of rotation, with rotated data being added to the training dataset, the recognition rate significantly increased from approximately 2% to 76%. As the rotation increased, the rate of recognition became very low for SVM-Wo and CS-Wo (without) while these rates have gone up dramatically from about 0% to between 82% and 84% for SVM-W and from nearly 0% to about 92% in CS-W (with), which again indicate that CS-W outperforms SVM-W.

4.2.1.3 Experiments on Zoomed out Faces using the FEI Dataset

In this experiment, all faces were zoomed out in the test sets from 10% to 90% in order to determine the effect of zooming on the rate of recognition. Here, there are nine groups of test images as shown in Figure 4-12. Similar to the framework adopted for the rotation experiments, the zoomed test faces were also passed to the VGGF model in order to extract features and the features were passed to SVM and CS for classification.

Figure 4-12: An illustrative example of sample images taken from the FEI Database [22] and manipulated at varied degrees of zoom-out modification (10% to 90%).

In the first part of the experiment, the recognition rate was evaluated for images without the added zooming-out effects into the training set (SVM-Wo and CS-Wo). As shown in Figure 4-13, the higher recognition rates were reached (100%) at the degrees 10° to 50° zooming levels by using SVM-Wo. On the other hand, when the images were zoomed out between the range of $70^o - 90^o$ the recognition rates were significantly reduced and ultimately reached approximately 0%. Contrary to this, when using CS-Wo, the recognition at zooming levels of between $10^o - 50^o$ produced recognition rates similar to SVM-Wo, i.e. 100%. Additionally, the recognition for images with a zoom between $70^o - 90^o$ still had the same recognition rate, as compared to the results for SVM-Wo.
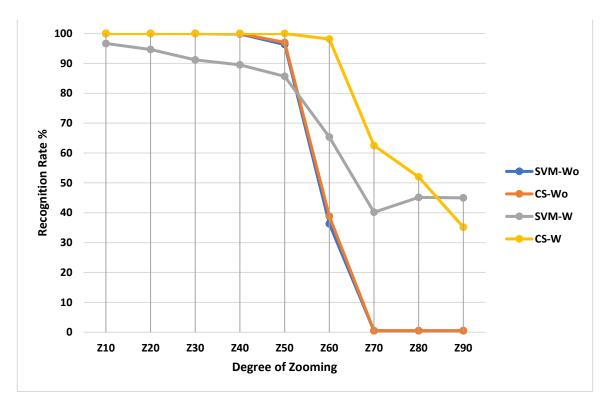
Figure 4-13: The rate of recognition (%) for zoomed out faces using SVM and CS classifiers based for the FEI database.

As the zoomed-out images were added into the training set, it became apparent that there is a slight improvement in the recognition rates, especially in the case of SVM-W at zooming levels of $70^o - 90^o$. However, for zoom-out levels between $10^o \ to \ 50^o$, there is a noted decrease in performance, which, is shown in Figure 4-13. The CS-W method reached a superior rate of recognition at zoomed out levels from $10^o \ to \ 60^o$ since recognition rates of 100% were achieved. Additionally, with 70° and 80° zooming out levels, it is clear that there is a gradual increase in the recognition rate compared with the previous experiment. At the zooming out level of 90°, the recognition rates enhanced from 0% to 45% when using SVM-W in contrast to the CS-W approach which concluded at 36%.

4.2.2 The LFW Dataset

Labelled Faces in the Wild (LFW) is a large dataset of face pictures, which is designed for testing the capability of face recognition in simulated uncontrolled scenarios. All the images have been collected from the Internet and consist of a spectrum of variations in expression, pose, age, illumination and resolution. The LFW database contains images of 5749 subjects with a combined total of 13000 images. As shown in Figure 4-14, the images have variable and significant background clutter. For this reason, during our experiments, some pre-processing were conducted in order to extract the face from the raw image.

Figure 4-14: Some sample face images from the LFW dataset [107]**.**

For the upcoming experiments the same experimental procedure was followed as with the FEI database. Hence, two main experimental conditions were tested; with and without parts for rotation and zooming out. In total we conducted, 14 sub experiments using both the SVM and CS classifiers.

4.2.2.1   Experiments on Parts of the Face using the LFW Dataset

In the case of the partial face experiments, the same procedures were followed as in the FEI dataset and generated 12 datasets for our experiments. Figure 4-15 illustrates some samples of the partial face's dataset.
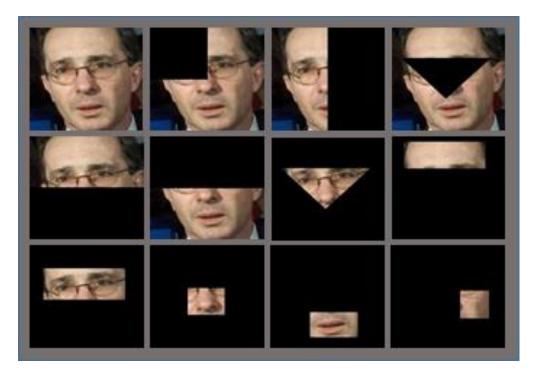
Figure 4-15: A samples of facial parts from the LFW database [107]**.**

All the extracted features from the VGGF model were passed onto both the classifiers (SVM and CS), in both experimental conditions, namely without parts in training (SVM-Wo and CS-Wo) and with parts in training (SVM-W and CS-W). In order to investigate the recognition rates for each facial part, each classifier was applied separately. In the case of without facial parts, it is clear that in general CS-Wo outperforms SVM-Wo for most of the regions of the face. By looking at Figure 4-16, we observe that the recognition rates for the right cheek, mouth, forehead and the nose is low, with about 1% for both the classifiers. In contrast, the rate of recognition increases significantly for facial parts such as the eyes, which reaches 40% using CS-Wo. Noteworthy, that by increasing the

proportion of the face, the recognition rate also improved significantly, with the best recognition rate of 100% for the ¾ face and full face. Again, in all tested cases, it is clear that the CS outperforms the SVMs.
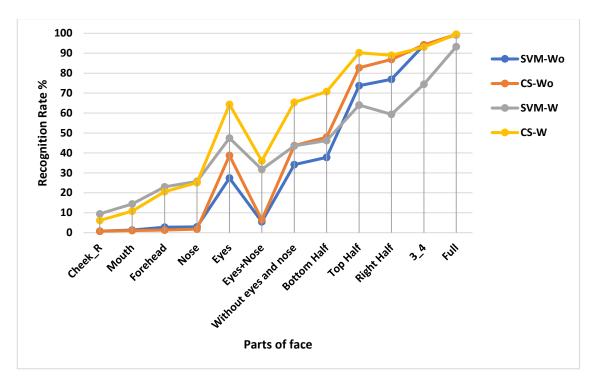
Figure 4-16: Recognition rates (%) using images from the LFW dataset. based on parts of the face using two classifiers: SVM and CS. The tested stimuli include conditions with and without individual facial parts during the training phase.

Similar to the previous experiments, all the experiments were repeated by using facial parts and the same classifiers but this time via adding the facial parts to the training sets (SVM-W and CS-W). The results, as shown in Figure 4-16, indicate a marked improvement in the recognition rates when using SVM-W. For instance, the rate of recognition for the right cheek was about 1% and now reached almost 10%. When considering mouth, forehead, and the nose, it is also observed that there were slight improvements in the recognition rates for both classifiers. By increasing the proportion of the face, the recognition rates also significantly improve with CS-W, which reaches about 68% instead of about 48% for the eyes, however this improvement did also occur when using SVM-W

(recognition rate about 47%). Moreover, for the images with occluded to the eyes and nose, lower, top and right half of the face, the rate of recognition enhanced significantly with CS-W, but slightly decreased for the ¾ of face, from approximately 94% to 93.5%.

4.2.2.2 Experiments on Rotated Face using the LFW Dataset

For experiments using rotated faces, a similar experimental procedure was applied, as with the FEI dataset, however now, all the face images were rotated in eighteen-degree increments. After feature extraction, the rate of recognition subject to each facial rotation was determined in the same manner as the previous experiments.

As shown in Figure 4-17, the results show that in the testing conditions where the data did not contain rotated faces, the rate of recognition was higher for rotated faces of 10° to 30°, since the recognition rates were close to 98% for SVM-Wo. When using CS-Wo the results are even better as in degrees R50 to R180 the recognition rates enhanced from about 0% up to 50%. However, as the degree of rotation increases, the proportion of recognition drops significantly for both the classifiers, with an average of 55% for SVM-Wo and 58% for CS-Wo. Moreover, for rotations between 60° and 180°, the recognition rate drops to almost 0% for both the classifiers.

For the experimental condition where rotated faces were added to the training set, the rate of recognition improved for both classifiers. For example, the recognition rates for images rotated at 150° was initially at 0.5% but improved to approximately 37%.
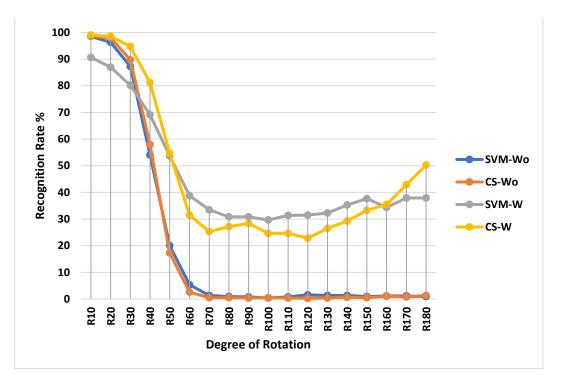
Figure 4-17: Face recognition rates (%) using SVM and CS classifiers based on varied degree of face rotations (10-180) using images from the LFW database. Two experimental conditions were tested and include with and without individual rotated faces within a training set.

4.2.2.3 Experiments on Zoomed out Face using the FEI Dataset.

In this sub-experiment all the faces in the test sets were zoomed out from 10° to 90° in order to determine the effect of zooming out on the rate of recognition. Figure 4-18 demonstrates the achieved results, and it is observed that the rate of recognition is higher for zoomed out images for depths between 10° until 50°, since the recognition rates range from 85% to 100%, for both the classifiers (SVM-Wo and CS-W). However, the rate drops sharply to almost 0% after 50° zoom out, for both the classifiers (SVM-Wo and CS-W). Although once the zoomed faces are added to the training set, it is clear that the performance of the CS-W improves further.
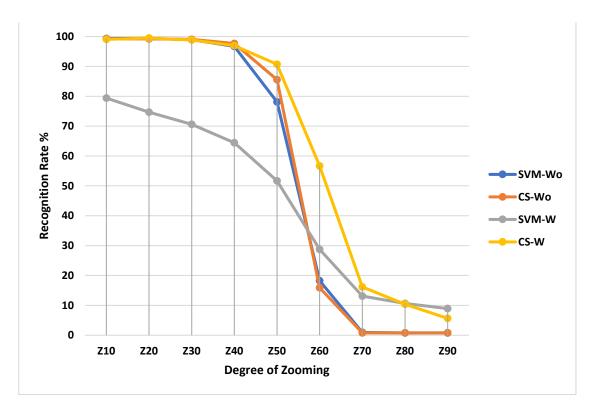
Figure 4-18: The recognition rates (%) for zoomed out images based on SVM and CS classifiers for faces, using the LFW database.

4.3 Discussions

From what have presented for the partial facial experiments using the FEI dataset, the highest recognition rate reported is for the 3/4 face, where a recognition rate of 100% is conclude by using SVM-Wo. In this experimental condition, the training set did not consist of various facial-parts. Also, in the case of CS-Wo, the right half of the face, the top half and the 3/4 face also concluded a recognition rate of 100%. However, the worst recognition observations are for the smaller and perhaps, less significant parts of the face, such as cheek, mouth and nose. When applying the same methodology to the uncontrolled LFW dataset and training with larger proportions of the face a slight decrease in the recognition rates is observed when compared with the FEI dataset, which was between 76% to 99% for SVM-Wo and 83% to 99% for the CS-W classifier. According to the

results obtained for smaller regions of the face, the worst recognition rate is observed for the cheeks, mouth, forehead and nose however the eyes did appear to hold more information.

For the second part of the experiments, when the individual parts of the face were added into the training sets and report a dramatic improvement to the rate of recognition, when using partial faces. For example, the recognition rate for the right cheek improved from 0% to 15% when using the FEI dataset. It is clear that the eyes still have the highest recognition rate when considering other individual parts of the face using the FEI and LFW datasets. Although the combined eyes and nose features report around 90% recognition when using the controlled FEI dataset. However, in the case of the uncontrolled LFW dataset this percentage drops slightly. Furthermore, in general, better recognition results were achieved by using the CS measure.

Thus, an important point to highlight here is that the CS measure, in general, appears to be a better classifier in this case, compared to both Linear and Non-Linear SVMs. SVMs require complete re-training when new data is added which subsequently introduces computational issues. However, in the case of the CS classifier, this is not an issue. Though having said that, in the testing stage, the CS classifier is more computationally intensive but given the greater degree of accuracy and makes logical sense to employ the CS classifier over SVMs.

In order to compare our results with the state-of-the-art techniques, it is clear that He et al., [77], presented work somewhat similar to what have been presented in this research. Whereby He et al., [77], has applied the recognition task to parts of the face, using images from the LFW dataset. The work by He

and colleagues is the closest to have done across in the domain of partial face recognition, with which we can make some form of direct comparison. Table 4-1 presents a comparison of our results with those presented in the literature. It is apparent that our results are significantly better.

Area	He et al., [77]	Ours
Upper	39.2%	90.2%
Down	7.8%	70.73%
Right	24.2%	88.9%
Left	27.6%	--

Table 4-1: A table of comparison for the experimental results reported by He et al., [77] and ours, using images from the LFW database. The comparison was done in four parts of the face, upper, down, right and left half of the face.

From the experimental results, it is possible to make further comments about the accuracy between the classifiers (CS and SVM). In general, the CS classifier outperforms the SVM. This can be observed by considering the individual class-level when examining the matching images picked by the classifiers. For example, Figure 4-19 presents the images that were matched correctly for a subject using CS-W and Figure 4-20 shows an example in which the CS classifier got confused.
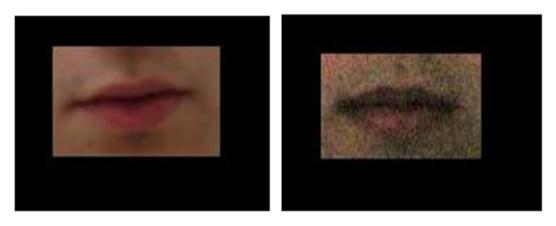
Figure 4-19: The result of correct matching using CS, for the region of the mouth [22].

In some cases, greater matching performance was observed by the CS classifier. For instance, Figure 4-21 shows images of cheeks that have been correctly matched by the CS classifier, though in this case, the classification may have resulted because of the prominent mark on the cheek see Figure 4-21.
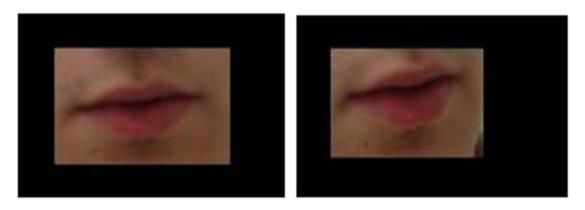
Figure 4-20: The result of an incorrect match using CS, for the region of the mouth.

Figure 4-21: The result of a correct match using CS, for the right cheek.

An interesting question that come to one's mind may relate to the generalisation capability of our proposed approach, especially when eliminating one or two facial parts from within the training sets. To investigate this, further

experiments were conducted whereby randomly left out two facial parts i.e. the right cheek, no eyes + no nose, from the training data of the LFW database. The experiments were run, without these two parts, and utilised the Linear SMV (L-SVM), kernel SVM (k-SVM), and CS for classification. The results of these experiments are summarised in Figure 4-22. As demonstrated when the two facial parts are removed the rate of recognition falls drastically from approximately 25.55% to around 3.88% (for the cheek) using CS. Moreover, for the facial-part without eyes and nose, the recognition rate fell by almost half for all classifiers, and the k-SVM performed worst in this case. As far as recognition for other parts of the face is concerned, in general, the rates for SVMs fell slightly whereas the CS appears to maintain its recognition rates.
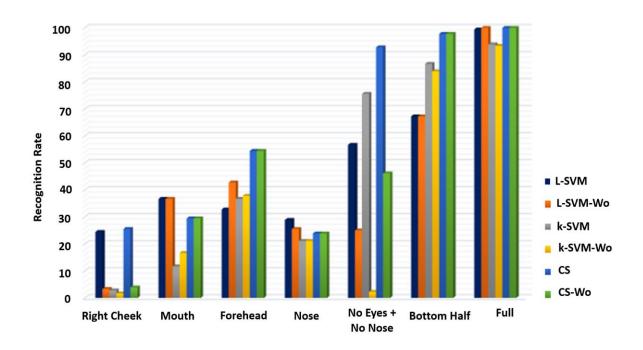
Figure 4-22: The results of the recognition task (%) when some parts of the face (cheeks and facial-part with no eyes and no nose) are removed from the training sets. The tested classifiers include linear SVM, kernel SVM and CS.

4.4 Conclusion

The ability for existing machine-based face recognition algorithms to perform adequately in cases of imperfect facial data such as occluded faces, rotated faces or zoomed out faces, as cues, remains a challenging task for the field of computer vision and visual computing. This work has presented the results of some novel experiments undertaken to highlight such issues of occlusion as well as to outline some potential solutions. This was done by utilising both controlled and uncontrolled public facial-image datasets and demonstrating how deep learning can be utilised for face recognition using imperfect facial cues. Thus, given some partial facial data, we show how feature extraction can be performed using popular CNNs such as the VGG-F model. In addition, showing how classifiers based on popular SVMs and CS can be applied to undertake facial recognition tasks.

This chapter has discussed a rather comprehensive set of experiments relating to face recognition using imperfect facial data. Our results show that as the proportion of the face gets small regardless of the prominent nature of the facial features such as eyes, nose, or mouth the rate of recognition is poor. However, even in the case of machine-based face recognition, the eyes appear to carry more recognition cues compared to other individual facial features. Furthermore, when it comes to rotated faces, it would be better to avoid highly rotated faces (e.g., faces rotated between 110° and 120°) as they appear to perform very poorly in recognition tasks regardless of incorporating rotated faces into the training data. In the case of zoomed-out faces, again it is advisable not to use highly zoomed out faces (e.g., faces zoomed out to 70% to 90%) as

probes. Finally, it clear that the CS measure greatly improves the performance of the classification when compared to both the Linear and kernel SVMs.

5 Age Progression and Regression and the Effect of Ageing in Face Recognition

Age progression and age regression have many applications and concurrently many challenges in the area of face recognition. Automatic aged and de-aged face generation has become an important subject of study in recent years. Over the past decade, researchers have been working on developing face-processing mechanisms to tackle the challenge of generating realistic aged faces. In this chapter, a novel framework is presented to address the problem of template faces based on the formulations of an average face for a given ethnicity which are used to generate a face for a set age.

The key objective of the proposed work is to develop and build a technique which addresses the age progression and regression of facial images based on the corresponding template images computed using different ethnicities as well as gender. The following are the main contributions from this experimental chapter:

- Efficient face ageing is developed, considering ethnic-specific face-templates.

- New, realistic faces are generated by incorporating colour and texture characteristics with the use of the warping technique.

- Face recognition approaches are applied to a new computer-generated face database to ascertain which estimated age could be an obstacle when considering the issue of age progression and regression.

This chapter is organised as follows. First, Section 5.1 discusses the adopted methodology. Then, Section 5.2 outlines our experiments and results. And finally, in Section 5.3 the conclusion of this chapter is presented.

5.1 Methodology

The proposed methodology for face age progression and regression is intended for overcoming some of the key challenges in such systems that currently exist. One key objective we strive to achieve here is the development of a flexible and lightweight method for generating realistic aged faces. The proposed frame-work is based on face templates, which are built by extracting information on the age, gender, colour and texture characteristics from a number of faces corresponding to the principal ethnic groups. Ethnicity based face templates can play a vital role in generating realistic faces, by complementing artefacts that arise from modern and commonly available techniques such as GANs based ageing systems.

The proposed system consists of two key parts. The first part is the mathematical method for building and generating the proposed face templates. It uses an average face for a given ethnicity, age and gender considering a sufficient number of faces for the corresponding category. In the second part, the generated templates are applied to the target face for age generation with two key control parameters, based on the colour and texture of the face. Finally, as part of our methodology, also a framework is proposed for verifying the accuracy of the generated faces through similarity comparison by means of standard face recognition. To compute and verify face similarities, a method based on the state-of-the-art CNNs is used.

5.1.1 Building an Ageing Template

To construct the templates for the calculation of age, based on the technique of calculating the average face is similar to the technique presented in [116]. The templates are generated for five specific ethnicities: Middle Eastern - Arabic, Southeast Asian, African - Black, Caucasian - White, Eastern - Chinese and with eight age-gaps starting from age 10 to 80, with 10 years increments for both the genders. The process of constructing the age template is based on the collection of sufficient data upon which proposed templates can be constructed.

5.1.1.1 Data Collection

The required data for creating the ethnicity specific templates was collected in four phases. Firstly, an Arab educational institution in Bradford, UK, was approached, and participants were recruited for photography. The participants consisted of male and female children and teachers, with an average age of 10-15 years for the children and 31-53 years for the teachers. In the second phase of image collection, colleagues from some Arab countries consented to send images of themselves. Thirdly, students from the University of Bradford were recruited. The fourth and final stage of data collection consisted of downloading readily available images from the Internet, again of various ethnicity and ages.

5.1.1.2 Age Template

All the collected images were categorised into groups based on ethnicity, gender and age. Since all the face images were originally of different dimensions it was necessary to normalise them and bring each image to the same reference

frame. The method of generating the templates of ageing can be broken down into the different steps described below:

1) Facial Features Detection: for face landmark detection, Dlib algorithm [180] is used. Dlib uses a pre-trained model to estimate the position of 68 facial landmark points (x,y) in the face, shown as part of Figure 5-1. However, since the forehead provides information about a person's age and Dlib does not cover that region of the face, five extra points were added based on information given by Dlibs algorithm, as 5-1, 5-2, 5-3, 5-4 and 5-5 below. It is assumed that the forehead is rectangular in shape and maps four points with the fifth point at the middle of the top between points a and b, as in Figure 5-2(a). To find the coordinates of point $a(x,y)$, firstly, computing the height of the forehead and based on our experiments this equals the length of nose D, hence the following equation is applied,

$$a_x = P_x^{19} - D_{nose}, a_y = P_y^{19} - C, \tag{5-1}$$

where P_x^{19} and P_y^{19} are x and y values at point 19, C is a constant value to normalize the distance and, $D_{nose} = \|P_x^{28} - P_x^{31}\|$. Similarly, the points $b(x,y), c(x,y), d(x,y),$ and $e(x,y)$ are computed via the format in Equation 5-2, 5-3, 5-4, and 5-4 respectively. i.e.,

$$b_x = P_x^{18} - round(\tfrac{D_{nose}}{2}) + 1, b_y = P_y^{18} - C. \tag{5-2}$$

$$c_x = P_x^{27} - round(\tfrac{D_{nose}}{2}) + 1, \ c_y = P_y^{27} + C. \tag{5-3}$$

$$d_x = P_x^{26} - D_{nose}, d_y = P_y^{26} + C. \tag{5-4}$$

To compute the middle point $e(x,y)$ we use,

$$e_x = round\left(\tfrac{P_x^{21}+P_x^1}{2}\right) + P_x^{21}, e_y = P_y^{21} - C1, \tag{5-5}$$

where $C1$ is a normalised distance.

After computing the five points to identify the forehead section of the face, the number of facial landmarks increases to 73 points in total, as shown in Figure 5-2(b).
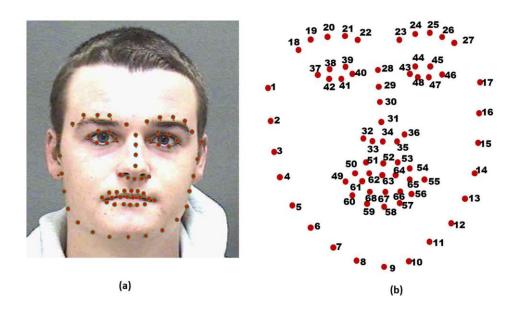
Figure 5-1: Dlib landmarks applied to an image from Morph face images [121]. **a) Facial landmarks. b) The position and order of 68 points on a face.**
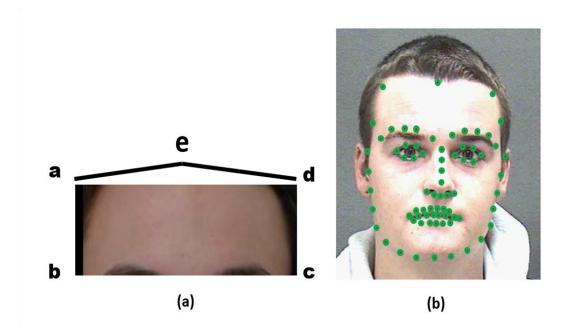
Figure 5-2: New five points. a) The five facial landmarks to cover the region of the forehead. b) The resulted facial landmarks after adding five points for the region of the forehead.

2) Generating the templates: As discussed earlier, the templates for given ages are considered for the five principal ethnicities namely, 1. Middle Eastern - Arabic, 2. Southeast Asian - Indian, 3. African - Black, 4. Caucasian - White and Eastern - 5. Chinese. We consider templates in age increments of 10 years i.e. ages 10, 20, ..., 80 years - for both females and males for each of the five ethnicities.

Consider, for example, the face images I_n of Middle Eastern males at the age of 70, where n is the number of images. Let's assume we want to generate a template for this age category for ethnicity. In the first step, images are pre-processed to resize them to the size and remove the background I_i^p, where $i = 1, 2, ..., n$. Then, the facial features P_i are extracted for all the images I_i^p by the method explained in (1). Before computing an average face and to avoid an unwanted template, all the images are aligned to the main shape by using the Generalised Procrustes Analysis (GPA) [116] [181] shown as Equation 5-6.

$$AI_i = GPA\left(I_i^p, M_s\right), \qquad i = 1, 2, ..., n, \tag{5-6}$$

where, M_s, is a mean shape and computes from Equation 5-7 such that,

$$M_s = \frac{1}{n}\sum_{i=1}^{n} P_i. \tag{5-7}$$

Now, compute the template by warping the aligned faces AI_i to the mean shape M_s and then it is possible to compute the average, as in Equation 5-8 and Figure 5-3(a), such that,

$$Template = \frac{1}{n}\sum_{i=1}^{n} warp(AI_i, M_s), \tag{5-8}$$

where $warp$ is inferred as the spatial transformation [42].
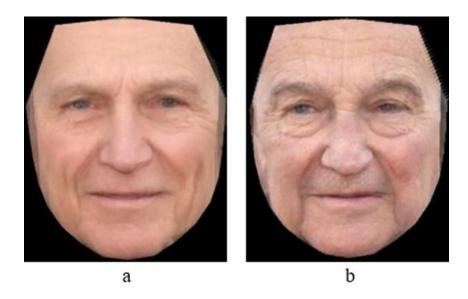
Figure 5-3: a) An example of a generated face template (fake images) at the age of 80. b) The template face after adding wrinkles.

3) Wrinkle map: Wrinkles play an important role in simulating realistic looking faces as they age. In simple terms, a wrinkle map W_m is an image with high quality wrinkles that can be added to ageing templates. This step adds the wrinkle maps W_m to the resulted template as in equation 5-9 and the template images with wrinkles N_t are generated as in Figure 5-3(b),

$$N_t = warp(W_m, template)$$

(5-9)

4) As the final step, all templates are coded to a specific label based on ethnicity, gender and age. For example: 4280 denotes: 4: (Caucasian), 2– (gender- female) and 80– the face-age, which is 80. Figure 5-4 shows some example templates of varied ethnicity, ages and gender. It should be noted that the faces shown in Figure 5-4 are computer-generated and have been produced taking an average from a range of features. Table 5-1 shows some example templates for various ethnicities, ages, and genders.

Ethnicity	Gender	Age	Generated Code
1	1	10	1110
2	1	20	2120
3	2	30	3240
4	2	40	4240
5	2	90	5290

Table 5-1: Illustrative example of sample face- template codes. The first digit represents ethnicity, the second digit is gender, and last two digits represent the face-age.
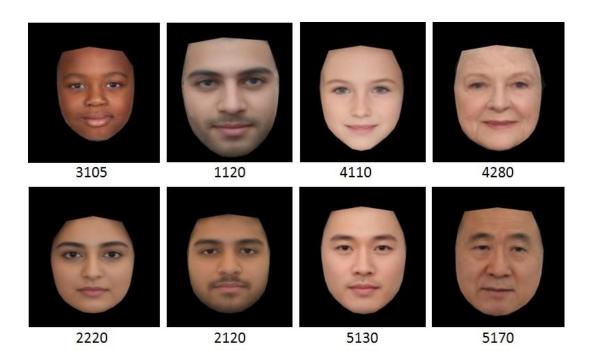
Figure 5-4: Example of face-templates with numeric descriptions of varied genders, ages and ethnicities/race. The image code 3105 denotes a Black male aged 5, 1120 means Middle Eastern male aged 20, 4280 refers to a White female aged at 80 years.

5.1.2 Computing Age Progression or Regression Generation

By this stage, all the face templates have been generated and can utilise them to either progress or regress a facial image to a given age. To do this, image morphing with a cross dissolve technique [182] [183] is utilised as discussed below.

Suppose there is an input face image I_{in} and want to age it to 60 years old for a Middle Eastern male. We invoke the corresponding template, i.e. T_{1160}. First, obtaining the corresponding landmark points for I_{in} and T_{1160} using the modified Dlib algorithm discussed earlier. Those points are referred to as P_{in} and P_t, respectively. Then, generating an intermediate warping field I_{wp} by using interpolation as in Equation 5-10,

$$I_{wp} = \alpha_{shape} \times P_{in} + \left(1 - \alpha_{shape}\right) \times P_t \qquad (5\text{-}10)$$

where α_{shape} is a parameter to control a degree of shape and is between $0.25 \leq \alpha_{shape} \leq 0.75$.

Then, an average between P_{in} and P_t is computed and used to find the corresponding Delaunay Triangulations (*DT*) [184]. However, to avoid ghosting effects in the resulting image, I_{in} and T_{1160} into I_{wp} are warped by applying an Affine Transformation function (AT) [185] [186] as in Equation 5-11 and 5-12, such that,

$$I_{in}^w = AT\left(I_{in}, P_{in}, I_{wp}, DT\right), \qquad (5\text{-}11)$$

$$T_{1160}^w = AT\left(T_{1160}, P_t, I_{wp}, DT\right), \qquad (5\text{-}12)$$

where I_{in}^w and T_{1160}^w are the warped images.

Finally, applying the method of cross dissolving method [182] to the warped images, I_{in}^w and T_{1160}^w, as presented in Equation 5-13 to get an aged face I_{aged},

$$I_{aged} = \alpha_{colour} \times I_{in}^w + (1 - \alpha_{colour}) \times T_{1160}^w, \qquad (5\text{-}13)$$

where α_{colour} is a parameter to control a degree of colour and $0.25 \leq \alpha_{colour} \leq$ 0.75.

Before discussing the experiments and their results, it is worth mentioning the choice of the two parameters α_{shape} and α_{colour}. In order to understand the best choice for these parameters, a number of preliminary experiments were run in which both the parameters were tested for possible values between $0.25 \leq$ and ≤ 0.75. As a result, based on our observations and computing the similarities using Cosine Similarity CS [165] and Structural Similarity Index (SSIM) [187]) between the ground truth and the aged faces, the optimal values α_{shape} and α_{colour} is found as 0.5. Figure 5-5 illustrated the example where we have taken a subject, aged him to 80 years by the various choices of the values for the α_{shape} and α_{colour} and compared the resulting face images with the ground truth. As can be observed in that figure, the highest similarity percentage is recorded at $\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.5$, which is recorded to be 86.15%. In addition, all possible values of two parameters (α_{shape} and α_{colour}) were tested by applying a recognition process between generated aged faces and ground truth faces. As can see in Table 5-2, the best recognition rate was at 0.5 for both parameters.

Ageing parameters α_{shape} and α_{colour}		Ages groups and recognition rate (%)							
Shape	Colour	10	20	30	40	50	60	70	80
0.25	0.25	9.60	28.25	31.63	30.50	11.30	15.82	13.09	17.26
0.25	0.5	76.83	86.44	88.70	86.44	76.83	59.32	74.40	69.64
0.25	0.75	80.56	84.46	87.23	90.25	91.22	70.77	85.21	80.12
0.50	0.25	16.38	39.54	45.76	37.85	19.77	17.85	17.85	25.59
0.50	**0.50**	**90.40**	**97.17**	**96.61**	**93.78**	**94.91**	**72.88**	**88.09**	**82.73**
0.50	0.75	90.30	96.23	95.88	93.12	92.55	69.88	85.25	83.22
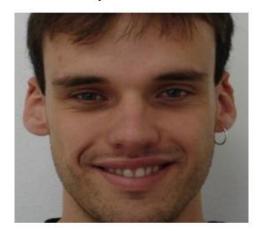
Table 5-2: Comparison between the different values of two parameters by applying a recognition process.

α_colour	α_shape					
	0.25			0.5		
	Observation	Similarity		Observation	Similarity	
		SSIM	CS		SSIM	CS
0.25		64.78	71.39		65.65	81.57
0.5		73.14	73.02		74.68	86.15
	Ground truth image					

Figure 5-5: An example from the tests carried out to determine the optimal values for the two parameters, α_{shape} and α_{colour}. Here a face is taken and aged to 80 years with various choices on the values for α_{shape} and α_{colour}. The similarities are determined using the Cosine Similarity (CS) and the Structural Similarity Index (SSIM). The results for the aged face of 80 years by using different values of α_{shape} and α_{colour}. As a result, the optimal parameter values that should be considered are determined to be α_{shape} and α_{colour}= 0.5.

Additionally, it is observed that wrinkles (present on the forehead, cheeks and eyelids) in the generated faces between the ages of 60 and 80, become clearer and this is achieved by merging the wrinkles map with the template. Figure 5-6 shows the generated face at age 70 alongside the original input face.

Figure 5-6: Illustrative example for the effect of merging the wrinkles template for age generation using FEI images [22]**.**

5.1.3 Recognition Task Using Unrealistic Faces

There are various approaches suggested for face recognition and classification using real faces, see [165] [187]. Here applying recognition approaches to computer-generated faces (i.e. unreal faces) in order to check the reliability of our method. To achieve this, a method based on a pre-trained Convolutional Neural Networks is used for features extraction and popular classifiers for classification.

Due to the low number of images per subject, the most well-known pre-trained model is adopted for facial feature extraction; VGG-Face model [41]. This model is widely used in the case of face recognition task and was developed by Oxford Visual Geometry Group [41]. For the purpose of our experiments, the last fully connected layers were tried and the best results are reported for layer 34.

The extracted facial features from both a ground truth image and an aged image are represented as a vector of dimensions 4096. All these vectors can then be used for training the classifiers such as the Cosine Similarity CS [188],

Decision Trees [169], k-Nearest Neighbours (K-NN) [93] and Linear Support Vector Machines (SVMs) [153].

5.2 Experiments and Results

For performance evaluation, experiments are conducted using two public-domain face databases (FEI [22] and Morph II [14]) and to generate faces of different ages, sex and ethnicities. In addition, optimal shape and colour parameters values are estimated by observing all the experiments and conducting a comparative study with similar published work.

Table 5-3 summarises the facial similarity results between the generated faces and the ground truth faces from the FEI dataset. Note, the results reported in Table 5-3 are for the values _shape = 0.5 and _colour =0.5.

Dataset	CS	SSIM	Average
FEI	85.161%	78.77%	81.97%

Table 5-3: Summary of the facial similarity results between the generated faces and the ground truth for all the faces the FEI dataset.

5.2.1 Using the FEI Database

The FEI University is a Brazilian introduced dataset consisting of 200 faces of students and staff of both the male and female sex [22], more details in chapter 4 section 4.2.1 and the sample of images in Figure 5-7. For each of the experiments, using the FEI dataset, three front face images were selected for each subject totalling to 600 facial images. The faces are then age progressed and regressed using the methodology described earlier. From the experimental results, by setting the two parameters (α_{shape} and α_{colour}) into different values, it is clear that for $\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.5$, our method consistently

produced the best aged face as mentioned in Table 5-2. Figure 5-8 shows a sample of aged faces with different parameter values for an individual whereby the ages considered are between 10 years and 80 years with different parameters.
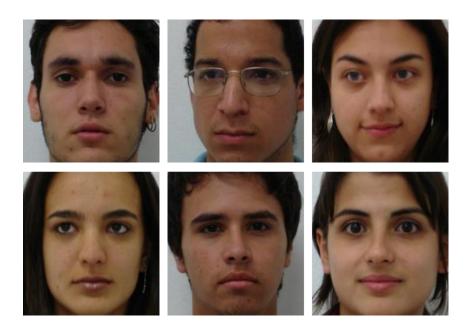
Figure 5-7: Example of the FEI face database images [22]**.**
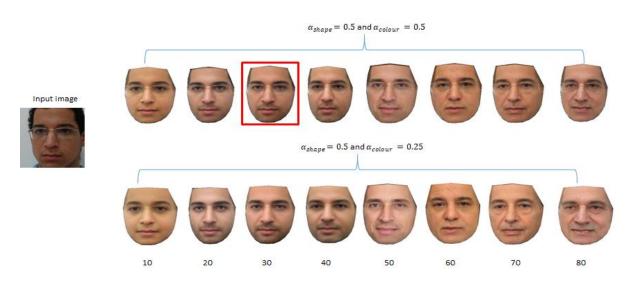
Figure 5-8: Schematic of changing the value of shape and colour. The first row indicates the aged faces when both parameters equal 0.5 which are more realistic and acceptable while the bottom row shows the aged faces with $\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.25$ and look significantly different from the original face. The red box in the first row indicates the best comparable aged face, which is believed to be 30.

As shown in Figure 5-8, it can be seen that the best choice of two parameters (α_{shape} and α_{colour}) are $\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.5$, which generate the most realistic and reliable ages. As a guide for our selection, the proposed method can be used to simulate face images at various ages and generate ages for whole datasets of varied genders and races. For the evaluation of the age regression and progression model, K-Fold Cross Validation was performed by using K = 3, which means that for each subject three original different images were used to produce novel faces. The FEI faces do not include ages for the individuals so an internet application (How-Old.net) [189] was used to estimated ages for each of the faces. Furthermore, similarities between the new faces and the ground truth face at the same ages are computed by using Cosine Similarity (CS) and Structural Similarity Index (SSIM) as seen in the last column of Figure 5-9. Note, crosses indicate that we do not have the templates for the black females between 70 and 80 years. Further, it is demonstrated that the percentage of similarity is very close to the ground truth.

Figure 5-9: Examples of aged face images for individuals from the FEI dataset. The first column represents the input images with estimated age, i.e. 30E means the estimated age is 30 years. The next 8 columns show the aged faces from 10 to 80 years. The faces inside red boxes correspond to those with matching ages between the input and the generated face. Note, crosses indicate that we do not have the templates for the black females between 70 and 80 years. Finally, the last two columns show the percentage similarities between input faces and the aged faces for the same ages.

5.2.2 Using Morph II Database

Similar to using the FEI dataset, experiments were repeated on the faces taken from the Morph II face dataset [121]. This dataset contained roughly 55,000 faces of 13,000 subjects and was collected over four years. It has faces with a range of ethnicities, gender and it consists of face images of individuals between the age of 16 to 77 years. The quality of images in this dataset is generally poor, particularly because the brightness contrast of some faces is very high. As a

result, some of the prominent features of the face in some of the faces are poorly represented. After carefully analysing all the images in the dataset, we selected images corresponding to 200 individuals through which we conducted our experiments. Figure 5-10 shows sample face images from the Morph II face dataset.

Figure 5-10: An example of samples images of Morph II Database [121]**.**

In order to generate new ages, subjects with the most available images were selected. Then applied the methodology described in section 5.1, again by using the same setting for the two parameters ($\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.5$). Figure 5-11 shows some examples of the aged faces. The first row shows the aged faces of a black male resulting from the input of his real face at the age of 53 years whereby it was progressed and regressed to generate aged faces between 10 and 60 years. Similarly, Figure 5-11 shows the aged faces of a white male whereby the input was his real facial image at 57 years. Again, by utilising the input facial image, it was then progressed as well as regressed to generate aged faces between 10 and 60 years. Moreover, in Figure 5-12, we can see how

the generated ages are very close to the ground truth faces when two different classifiers were applied for matching.
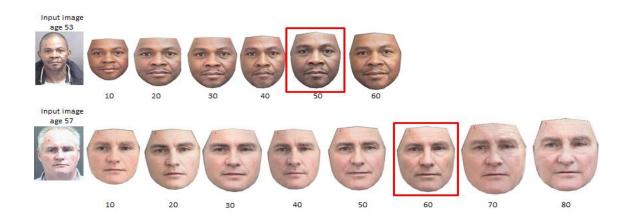
Figure 5-11: An illustrative example of age generated faces using two single input images taken from the Morph II Database [121] **(Input image top: black male, Input image bottom: White male).**
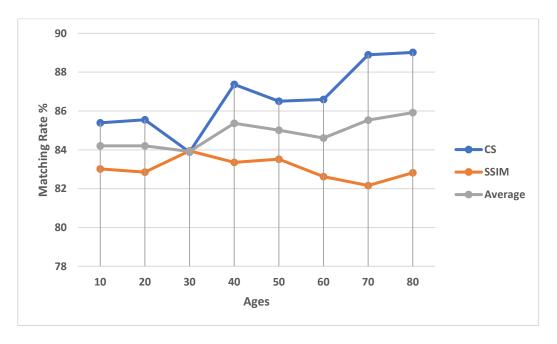
Figure 5-12: The comparison of matching between the generated faces and ground truth images in Morph II faces.

5.2.3 Evaluation and Comparison of the Proposed System

In order to conduct an evaluation of our proposed model, two approaches were utilised. The first method compares with state-of-art technology and the second is a comparison with ground truth images.

5.2.3.1 Comparison with the Most Recent Work

We rigorously compared our method with some of the most recent work in the literature. For comparison, we have selected the use of GANs [131], Recurrent Face Aging (RFA) framework [190] and Coupled Dictionary Learning (CDL) [128]. All these methods are reported to be examples of state-of-the-art on age progression and regression.

In this experiment, firstly, investigating the effectiveness of our method by applying it to the method mentioned in [131]. The advantage of our proposed approach is the ability to progress and regress a face of older ages with wrinkles and make it pragmatic. In Figure 5-13, it can be seen that our method generated faces, which are more realistic especially at the older ages. Furthermore, wrinkles and visual signs of aging are clear for observation. In addition, a comparison with the most recent work (RFA and CDL) was done by following the same style of experiments as presented in [190]. In this case, images were selected from FGENT [14] age dataset. And our results demonstrate that our proposed approach outperforms previously presented methods. Figure 5-14 clearly displays the quality and resolution of the new generated faces.
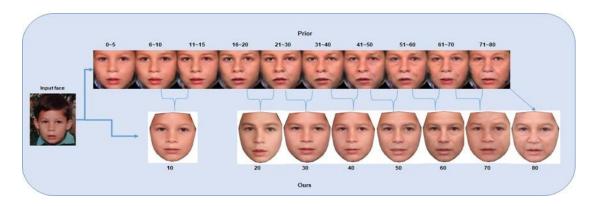
Figure 5-13: The comparison between previously published methodology on face ageing by Conditional Adversarial Autoencoder (CAA) [131] **and our proposed experimental framework using FGNET face data set** [14]**. The left single image is the input at the age of 5 years old. The first row displays the evidence of ageing in ten groups of ages, as published by** [131]**. The second row illustrates the results from our method at ages started at 10, 20, ..., 80. We put the braces between every two images in the first row to indicate the age group and compared the ages with our resulted faces.**
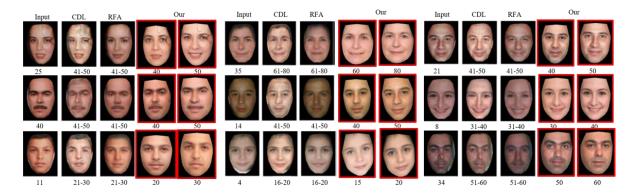
Figure 5-14: A comparison of proposed face ageing methods using the most recent works and FGNET face data set [14]**. The first column in each block represents the input image, the second and third rows show the results of previous methods (CDL** [128] **and RFA** [190]**). Last two columns in each block enclosed with a red square display the results from our proposed method.**

The second part of the comparison process included training the CAAE system with images taken from the UTKFace database [134], which contains 23,000 images. The CAAE system has been used to age images from samples taken from the FEI dataset. Then compared the resulting images with our method. Figure 5-15 demonstrates the resulted ages from images belonging to the FEI database after being passed into the CAAE system, for three females and one male subject. It is clear that the required ages are poorly visualised. However, when the same images were introduced to our system in the same order, a high performance was concluded. Moreover, our system was able to produce smooth aged faces even in a condition with facial occlusion, i.e. subject photographed with glasses, see Figure 5-16.
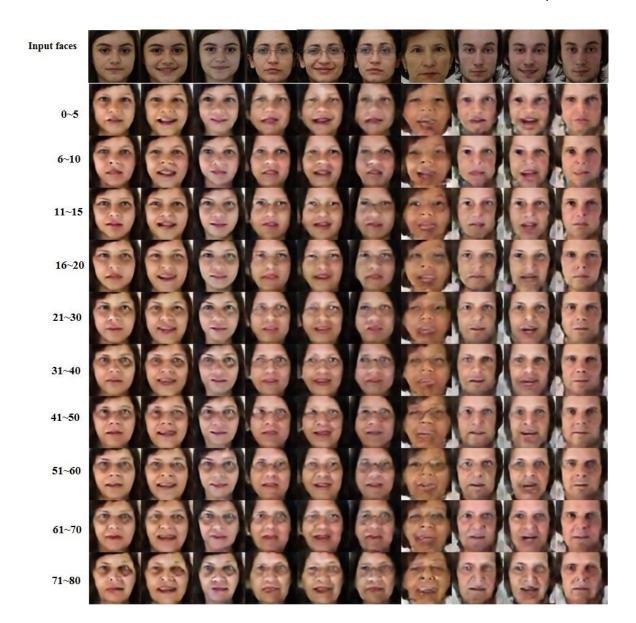
Figure 5-15: The results after applying the CAAE system to some faces taken from the FEI database [22]. The first row illustrations the input faces. The latter rows display the generated ages for the different age groups (left column).
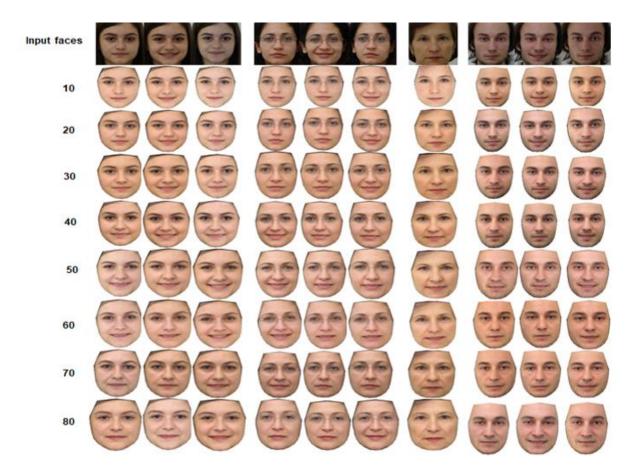
Figure 5-16: Multiple examples of a computer generated aged-faces by applying our method to images taken from the FEI database [22]. The first row displays the input face and the remaining row show the resulted ages (10 to 80 years).

5.2.3.2   Comparison with Ground Truth Images

A comparison with ground truth images was conducted for a handful of celebrities by computing the similarities between the generated faces and the real faces at known ages for each celebrity. For this purpose, we did some comparative analysis of our generated faces with the ground truth facial images for two celebrities, namely Angelina Jolie and Brad Pitt. The face similarity matrices were computed between the generated faces and real faces with the corresponding ages of the two celebrities. Once new faces are generated, the VGGF model is used as described in 5.1.3 and two other methods to measure the similarities between the facial images.

In the first approach, which is based on a feature map, a total of 4096 features, using the VGGF model, are extracted for all the facial images by using

the convolutional layer 34 in VGGF. These feature parameters are then passed to the Cosine Similarity (CS) classifier to compute the percentage similarity. In the second approach, the Structural Similarity Index (SSIM) was used to compute the similarity between the ground truth and the new faces. For our final approach, the image map method was used in which an online web portal was used to identify the similarity between two facial images, which is referred to as IMG-Online (IMG) [191].

Figure 5-17 shows the face images generated for four different ages for Angelina Jolie. It can be observed in Table 5-3, that the highest similarity measure obtained when comparing the ground truth with the aged faces is for age 20 years which is 90.63% for CS, 96.08% for SSIM and 96.46% for IMG. In contrast, the lowest similarity measure recorded is for the age of 15 years, which is 76.41% for CS, 62.02% for SSIM and 68.93% for IMG.

In the second example, the face images of Brad Pitt were used to evaluate the proposed method. Firstly, four different aged images were generated using our proposed approach. Figure 5-18 shows the gendered faces and the corresponding faces of ground truth. Similar to the previous example, to compute the facial similarities, the features were extracted for all images by using the VGGF. Then compared the aged images with the corresponding ground truths based on the three approaches discussed earlier.

In Table 5-4, it can be observed that the similarity measures for all the generated ages when compared with the ground truth are well above 70%. The highest similarity percentage is for age 20, which is 85.82% by using CS, and the lowest value obtained is 72.37% for age 40 years.
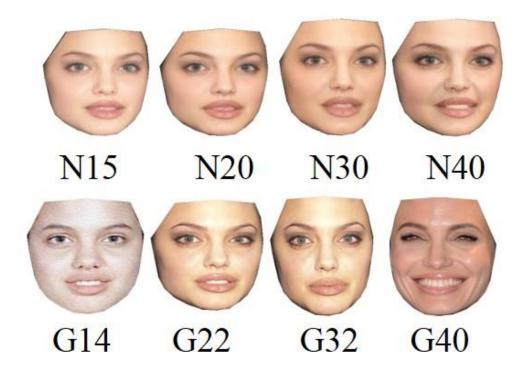
Figure 5-17: Comparison of aged faces with the corresponding ground truth faces for Angelina Jolie. The first row shows the generated faces from age 15 to 40 years with the naming convention NXX, where N indicates the face is computed and XX refers to the age in years. For example, N15 indicates a newly generated face at age 15 years. In the second row, real faces for the ground truth are presented, which were collected from the Internet. The naming convention used here is GXX, where G indicates that image is a ground truth and XX corresponds to the age in years. For instance, G14 indicates a real face of age 14 years.

Method of Comparison	Age range and similarity			
	15	**20**	**30**	**40**
Cosine Similarity (Features)	76.41 %	90.63 %	85.38 %	71.61 %
Structural Similarity	62.02 %	96.08 %	85.33 %	68.36 %
IMG-Online	73.46 %	96.46 %	84.03 %	68.93 %

Table 5-4: The percentage % of similarities between ground truth and generated faces for different ages for Angelina Jolie.
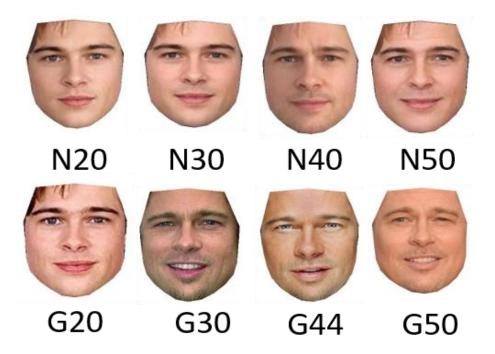
Figure 5-18: A comparison between the new aged faces and the ground truth faces for Brad Pitt. The first row images show the generated aged faces for the age 20 to 50. In the second row, we illustrate the ground truth faces, which were collected from the Internet.

Method of Comparison	Age range and similarity			
	20	**30**	**40**	**50**
Cosine Similarity (Features)	85.82 %	75.83 %	72.37 %	73.61 %
Structural Similarity	80.43 %	75.57 %	83.72 %	83.36 %
IMG-Online	80.14 %	82.28 %	76.56 %	78.63 %

Table 5-5: The percentage (%) of similarities between the ground truth and computer-generated faces for Brad Pitt at different ages.

5.2.4 Unrealistic Face Recognition Trials

We have conducted a range of experiments on generated face databases. These datasets are the new faces, which were aged from the FEI, and Morph II face database and the described method in section 5.1.3 was applied to those faces in order to find the reliability of the generated images.

The first test in the recognition stage was based on the FEI database. As previously mentioned, each individual had 14 images and in order to apply a recognition process, three images per subject were selected for an age generation, while the remaining images were utilized for the purpose of training. In the test set, to demonstrate the best values of two parameters (α_{shape} and α_{colour}), the recognition process was conducted on faces which were generated by ($\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.5$) and ($\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.25$). Furthermore, since we have eight age groups and each group has 600 aged images, it meant that there are 4800 × 2 new faces for a test.

We used the VGG-F model on all images in both the training and test sets to extract features and four different classifiers (see section 5.1.3) were applied. In all experiments, the test set was separated into individual groups representative of a specific age. That allowed us to test each set of age independently and allowed us to identify which age group concluded the best recognition rate.

In case of choosing $\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.5$, we carried out the recognition phase of our experimental framework under four different scenarios of classification for each age and have the summarized the results in Figure 5-19 (a). It is apparent that the recognition rate of CS is significantly higher and outperforms the other classifiers, reaching between 93% and 96% for the ages of 20 up to 50 years. However, we also observe that at the age of 60 the percentage of recognition decrease to approximately 75%. Further, the KNN classifier also has a high recognition rate, and is somewhat similar to the CS results. In comparison Decision Tress classifier concluded the lowest rate of the recognition, which indicates that the classifier is not sufficient for the task of classifying ageing faces.
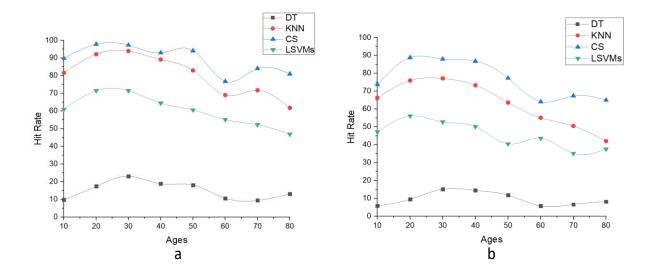
Figure 5-19: The results of the face recognition process (%) based on four tested classifiers (DT, KNN, CS, LSVM) using the FEI database by a) $(\alpha_{shape} = 0.5$ **and** $\alpha_{colour} = 0.5)$ **and b)** $(\alpha_{shape} = 0.5$ **and** $\alpha_{colour} = 0.25)$.

New age faces are generated by putting the values of the two parameters as $\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.25$, and applying the methodology for recognition. From the experimental results, it is clear that the performance of classification does decline slightly. Regardless, this shows that the best values of the parameters are 0.5 for α_{shape} and α_{colour}, as shown in Figure 5-19(b).

In Morph II database, the same approach as in FEI experiments (more information about this database in presented in section 5.2.2) was followed. As part of this experiment, we selected 200 subjects for an age generation task (including progression and regression) and the generated faces were used as a test set. The generated faces are separated into two groups that rely on the colour of skin, namely black and white).

Again, the face recognition process was applied to the aged faces in two cases by changing the values of (α_{shape} and α_{colour}). As mentioned above, there are subsets of the test set; the first is for the black race for testing, while using mixed faces (black and white faces) for training. the classification method was applied to the black group ($\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.5$) and as shown in

Figure 5-20(a), the recognition rates concluded by using CS and KNN classifier are high especially for the ages of 20 and 30 years old. However, the percentage does drop marginally for the other ages. In contrast, when $\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.25$ were used, the decline is below 80%, as seen in Figure 5-20(b).
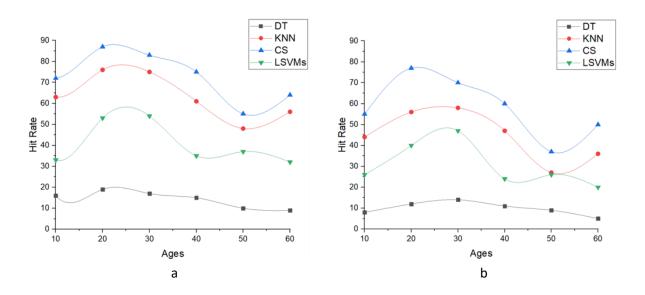
Figure 5-20: A graph to show the application of the face recognition process to images taken from the Morph II dataset (black race) by setting values at: (a) $\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.5$and **(b**$\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.25$**.**

In the second subset where the white ethnicity was used as the test set, the recognition rate improved by utilizing $\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.5$ for the ages of 10 to 50 years and 70 years old; this is represented in Figure 5-21(a). However, in the case where $\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.25$, this rate of recognition declines to <80% for all the classifiers, see Figure 5-21(b).
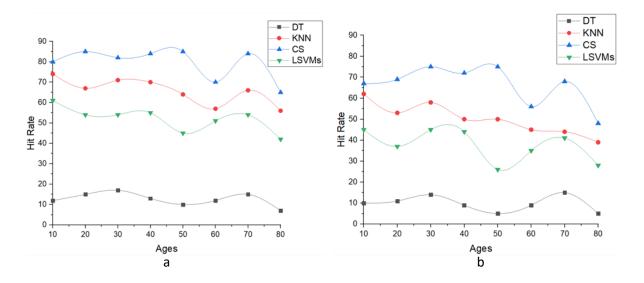
Figure 5-21: A graph to show the face recognition results (%) by using images taken from the Morph II (White race) by setting values at: (a) αshape = 0.5 and αcolour = 0.5 and (b) αshape = 0.5 and αcolour = 0.25.

Finally, the results of the recognition were summarized by averaging the ageing percentages in all cases, to decipher which phase of age may be problematic for recognition. There are changes that occur in the outer appearance of a face because of ageing, such as skin colour and perceived face shape. Based on the experimental outcomes the most challenging ages are typically between early and elderly years. Our experiments report that at age 10 and from the ages of 50 through to 80 prove to be the most challenging, see Figure 5-22. The main reason is that most of the faces in the FEI dataset for example, are between the ages of approximately 20 to 50s years. Additionally, in the Morph II database, the subjects are all between the ages of 30 to 50 years old. Thus, the databases do not contain images of very young or extremely elderly participants.

In addition, the Morph II dataset has a low proportion of recognition and the reason for this is the quality of the original images, which are low. However, the Average of Over All (AOA) illustrates the proportion of recognition by using

$\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.5$ in both sets and the best outcome is around

68% at age 30 and the worse case is roughly 47% for the age of 60 years.
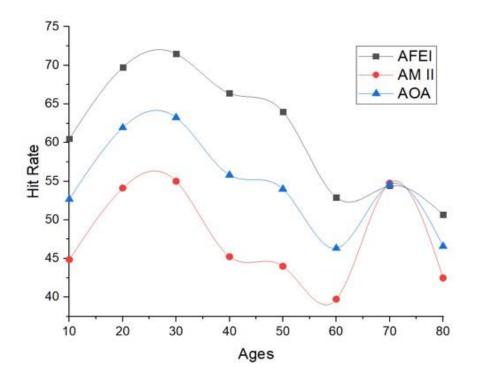
Figure 5-22: Summary of one to many face similarity matching tests (%) on the FEI and Morph II datasets. The results are for similarity rates for the generated faces in the age ranges from 10 to 80 years for $\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.5$. Average for the FEI dataset (AFEI), Average for the Morph II dataset (AMII) and Overall Average overall (AOA).

5.3 Conclusion

This work has presented a novel framework for age progression and regression using face images of varied databases. The proposed approach addresses the problem of face-age progression and regression in a more direct manner, by relying on individualised face age templates. The templates are built based on the formulations of an average face for various ethnicities and gender over a range of ages. Experiments and tested the proposed method were conducted on two public domain face datasets of known and unknown ages, namely the FEI and the Morph II. The experimental results demonstrate that the

age generation method proposed achieves high accuracy. It is observed that assigning a value of 0.5 to the two controlled parameters α_{shape} and α_{colour} can produce optimal results. Additionally, it is noted that aged faces of high quality can be generated even if the input face is of low quality, e.g. as seen in some images from the Morph II database. To find out which estimated age could be an obstacle for regression and progression, all generated faces were passed into the pre-trained model, the VGG-Face, for feature extraction. For the purpose of classification four classifiers were utilized to find a recognition rate per age. It was concluded that the best results are achieved by using CS and the k-Nearest (KNN).

The experimental results demonstrate that the age generation method does achieve high performance in the case of young ages by assigning the value 0.5 to two controlled parameters. In addition, a high-quality generated face is achievable even if the input is of a reduced image quality. With regards to the face recognition, the results show that unreal face recognition can decline gradually especially at the ages of 10 and 60 years old. Further, there are some limitations in our technique, but they relate specifically to the availability of face image data.

Additionally, we also benchmarked our method with the existing state of the art methods such as those based on GANs, RFA and CDL. Based on the extensive experimentation we have carried out, we can confidently claim that the proposed method for age progression and regression is efficient, lightweight yet accurate when compared to the present state-of-the-art in the field.

6   Model based Deep Learning for Partial Face Recognition

One of the major applications of face recognition relate to criminal identification, control- access and security, including biometric security [192]. Given the ample number of identification mechanisms such as iris recognition, handwriting biometrics (i.e. signature) and fingerprint analysis, face recognition has earned the most popularity due to its non-invasive and ubiquitous nature. Over the past decade various contributions have been introduced by avid researchers in face recognition, mostly utilising problems with full frontal-face data or a significant proportion of the face [11]. However, the most exciting phase, which is still in an amateur stage, is the use of occluded faces, specifically when features of the face disappear or just some parts of a face such as the eyes, nose, or mouth are available. This, typically can lead to low-performance and low accuracy in facial recognition [193].

Many researchers have focused on recognizing faces with occlusion and tried to tackle this issue [74] [73]. Currently, there are several approaches to face recognition tasks [194] [102]. In general, those methods can be categorized into two classes based on a used methodology [195]. In the first class, methods based on traditional algorithms (holistic, feature-based geometry, and hybrid methods), such as Support vector machines (SVMs) [153] and Fisher discriminant analysis, have been used [196]. Such methods have limitations surrounding the use of good features that can be used as needed when recognising a face. The second category is deep learning-based methods [158], the most popular type is Convolutional Neural Networks (CNNs) [197]. A key advantage of using a Deep Learning-based approach for face recognition is that it learns the most and the best features for representation.

Linear Regression Classification (LRC) [198] is an example of face recognition with occlusion. The experimental results of using this method show that LRC can recognize full faces, but when the faces have an element of occlusion, or some parts of the face are missing, the recognition is very poor. Sparse Representation Classification (SRC) is another approach that was tested with occluded faces [198]. Even though a sparse face is robust against noise and highly occluded faces. Nevertheless, depending on SRC, there is a strong presumption that all faces have to be aligned accurately in advance, otherwise, sparsity is hard to achieve.

More recent work has been conducted by Zhe Chen et al., [199], named Sparse Regularized Nuclear norm based Matrix Regression (SR_NMR), to relieve the effect of contiguous occlusion on face recognition issues. This method can impose L1-norm instead of L2-norm representation of NMR structure. The experimental results (as tested on three different databases) demonstrated that the method can achieve better performance of recognition compared with the traditional NMR. In addition, it can be helpful for recovering low-rank error images in the presence of occlusion.

Deep learning approaches have concluded great outcomes for face recognition in several aspects of applications. In 2019 Cen et al., [200] proposed a novel deep dictionary representation-based classification scheme for occluded faces. The researchers utilised the convolutional neural networks model for features extraction and applied a dictionary to code the extracted deep features linearly. The suggested method was computationally effective and was robust to enormous contiguous occlusion. The evaluations of comprehensive experimental methods have demonstrated the exceptional achievement of the proposed method compared with other state-of-the-art techniques.

Wu et al., [198] proposed a method called Occluded Face Recognition Based on Deep Learning. The researchers trained a face recognition model and learning relied on deep convolutional neural networks, which are known to be robust against facial expression, differences of illumination, as well as facial occlusion. The experimental results concluded that a recognition rate of 98.6% was achieved for occluded faces.

Based on the literature above, it is apparent that most methods were applied to occluded face images not to isolated parts of the face, especially in conditions where the facial image contains only a portion of the face, such as the nose or eyes. In this case, methods can achieve a low performance, and the percentage of recognition will be lower. The main reason for this behaviour is due to the structure of the methods used and the way in which the models have been trained i.e., on the data type. Moreover, most face databases do not contain isolated parts of the face from which a testing model can learn new features that concern a specific region of the face.

This chapter presents a method to overcome the challenge of face occlusion based on face parts. Our experimental framework is built on a new model based on Convolutional neural networks (CNNs) for each part of the face and those models are individually trained on individual parts of the face. Through testing and evaluating the criterion-specific models, it became clear that they can achieve high recognition results. In addition, the models use the fewest possible features and concluded a high recognition rate with the benefit of reduced computational complexity.

This chapter is organised as follows: Section 6.1 provides a detailed description of the methodology for the new models. Section 6.2 presents the face dataset and the experimental results for each tested condition when using

different image datasets. Section 6.3 presents a discussion of the reported results. And the chapter concludes with a discussion relating to the advantages of the novel algorithm.

6.1   Partial Face Recognition Based CNNs Models

This section offers a comprehensive explanation of the proposed methodology which has been used to generate face recognition models for sectioned regions of the face. The novel models have been created based on CNNs inherited from VGG16 [40] and the models are named as follows: (1) Eyes-CNNs (ECNN), (2) Nose-CNNs (NCNN), (3) Mouth-CNNs (MCNN), (4) Forehead-CNNs (FCNN), and (5) Eyes & Nose-CNNs (ENCNN) based models. The architecture of VGG16 was introduced by the Visual Geometry Group [41] for Large-Scale Image Recognition. This model consists of 13 convolutional layers (CONVs) with the same size filter (3x3). These layers are divided as follows, the first two layers have depths of 64, two layers have 128, three layers have 256, and six layers have 512. After these layers, there are three Fully Connected (FC) layers, two layers have 4096 units and last layer has 1000 units, which are based on several classes.

VGG16 is considered a good example of the Convolutional Neural Networks architecture. In this model, instead of using an enormous number of training hyper-parameters, it uses the same arrangement of layers and hyperparameters in the whole architecture to reduce this number. This arrangement is structured as follows, CONV with filter size 3x3, stride 1, and 2x2 mask for padding and max-pool layers with stride 2.

The suggested experimental framework encompasses two stages as mentioned in Figure 6-1. In the first stage, the first model is constructed and trained on a dataset of eye images only. In the second stage, the generated model is used for building the rest of the models by utilising the same structure but this time with different parts of a face; this process is called Fine tuning [201].
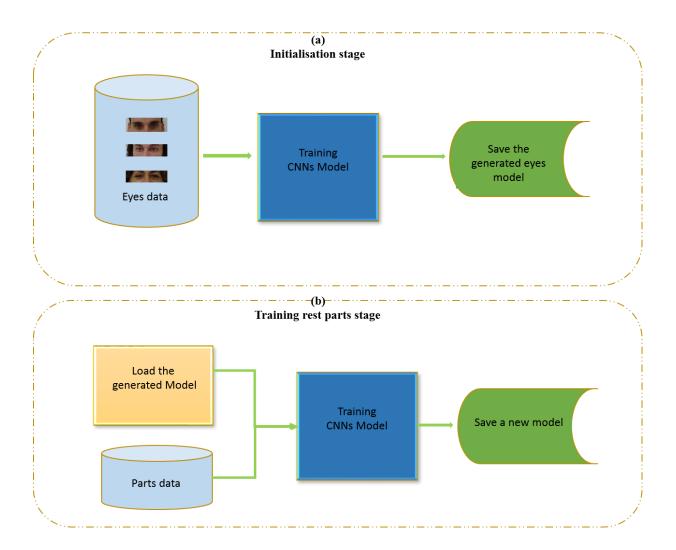
Figure 6-1: An overview of the experimental architecture for our proposed method of partial face models. a) Initialisation stage, which is used to prepare the first model by using a dataset of eyes images only and b) the generation of new models based on the rest of the human face.

6.1.1 Initialisation Stage

This section illustrates the configuration of the proposed CNNs model which have been used for all the experiments, and then, how the models are trained.

6.1.1.1 Architecture of the Network Model

Our proposed model is inspired by the VGG16 model and requires an input image with a pre-fixed size 224x224x3, five convolutional blocks, each containing a number of convolutional layers followed by non-linearities activation functions. As shown in Figure 6-2, in the first block there are two convolutional layers (conv1) with a filter size of 3x3 and 64 features maps. They produce the same size of the output which is 224x224x64 and a different number of trainable parameters which are 1792 and 36928. The resulted features maps are subjected to max pooling layer with kernel size 2*2 and stride 1, and the output size is (112, 112, 64). The structure of second block (conv2) has the same structure as conv1, but the number of filters is 128, and the number of trainable parameters is 73856 and 147584, respectively. After applying the same max pooling, the resulted shape is (56, 56, 128). The third block (conv3) has two convolutional layers with the same number of filters (256). The resulted shape of (con3) after applying max pooling (2*2) is (28, 28, 256) and the number of trainable parameters is 295168, and 590080. In the fourth and final blocks (conv4 & conv5), there are two convolutional layers with 512 filters followed by max pooling with size (2*2). Block four produces a total number of 1180160, and 2359808 trainable parameters respectively, and the output shape is (28, 28,512). Block five generated 2359808 trainable parameters for each sub-convolutional layer and the output shape is (14, 14, 512). After those blocks, there are three Fully Connected (FC) layers; FC1, FC2, and FC3 and each one has a different number of trainable parameters;

20,070,600, 40200, and 40200, respectively. Thus, the final total trainable parameters are 29,555,992 which is less compared to the original number 134,268,738. At the end of the model, there is a final layer known as the soft-max which functions to classify and predict the probability of 200 classes.

6.1.1.2 Training Phase

Training a CNN model is a procedure to minimize the variation between the ground truth label and the predicted output from within a training dataset. This is achieved by locating learnable parameters (kernels and weights) within the convolutional and fully connected layers. In our model, the weights of filters are initialised by using a Gaussian and standard deviation, biases are initialised to zero. To evaluate the performance of the model through forwarding propagation, the most common loss function is used for a multiclass classification problem, called cross-entropy function [202]. The value of the loss function identifies how well or poorly a model conducts after every iteration of the optimization. In addition, based on the error gradient in the current state of the model, all learnable parameters are updated during the optimization method, a process called gradient descent and backpropagation. Learning rate is another hyper-parameter that has the role of controlling how fast the CNNs model learns the presented data. The value of this parameter is positive and in the range of between 0 and 1, in our case, the best value of the learning rate is 10^{-4}.
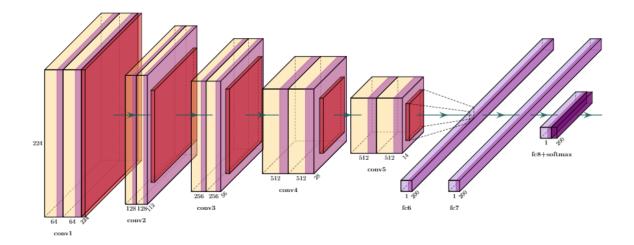
Figure 6-2: A visual schematic of the network architecture for our proposed model. Each colour represents a different type of layer. The yellow regions indicate the convolutional layers, the pink regions denote activation function, the red regions represent the max pooling layers, the light purple denotes the fully connected layer, and lastly, the dark purple represents the softmax function.

The last two hyper parameters are the number of epochs, which indicates the number of times the learning method works during the whole training dataset, and a batch size, which determines several batches or sets, needed to finish one epoch (in all experiments, batch size= 64). In our methodology, there are two stages to conduct the training process. In the first section, the model was trained using 20 epochs with a batch size of 64. After that, the resulted weights were saved to utilise them for initialising the weights in the second section of the training. During phase two, 50 other epochs were conducted to train the model from the last weights. Those epochs were divided into ten sections and each section (SN section no) has five epochs (N-epoch, which increasing by 1 each time up to 5) as shown in Figure 6-3. In SN1, we load the weights from the previous training as the initialisation phase, and train the model for five epochs, the resulted new weights are saved to use for next SN. This procedure continues until SN10 is reached, then the model is saved. For the remaining experiments, we follow the same methodology used for training, especially since the validation

and training loss reached <0.02 approximately, while the validation accuracy reached 85.42%. This indicates that the performance and prediction of the model is one of the best. More detail relating to the training phase is provided in Figure 6-4.
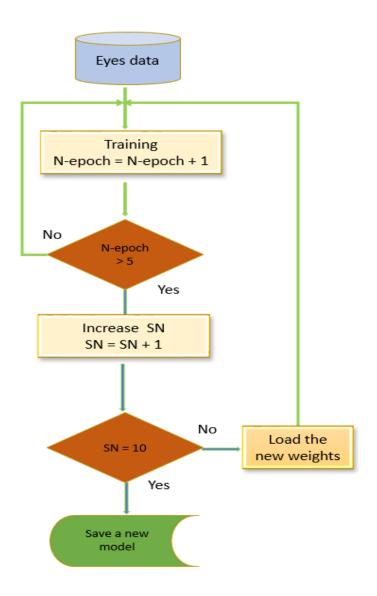
Figure 6-3: A flowchart to show the process of training the model by conduced iterations producer.

The training data that was used to create the first model, is selected from the face image database named VGG-Face dataset [203]. Further details are provided in section 6.2.1. The first model, called ECNN (Eyes CNN), has been trained using eye images originating from 200 people and consist of more than 70,000 images, see Figure 6-5. All images in the training set have different

dimensions, thus, before the training process, they are resized to 224*224 to fit the size of the input layer. The dataset was objectively split into two groups: 70% training group and 30% validation group. The training set was used to train the model, where the loss is computed by forwarding propagation and updating the learnable parameters by backpropagation. As far as the validation group is concerned, they were used for monitoring the performance of the model through the training process.

6.1.2 Fine-Tune Stage

The process of fine-tuning for machine learning algorithms is a procedure whereby the CNNs model already trained for a given job and/or data type, is utilised for performing a similar task [201]. This is done by replacing an output layer, which originally was trained for recognizing previous classes, with another layer, which can now recognize a new number of classes for a new task. An advantage of using fine-tuning is the reduction of computational time during the training phase. Subsequently, the first layers already have the efficiency to deal with a new task, and training will utilise the last layers, without the need of training from scratch. Another advantage is improved performance because the pretrained models are typically trained on large datasets.

To generate new models for NCNN, MCNN, FCNN, and ENCNN we fine-tuned the ECNN model. To train the NCNN model by means of the nose dataset, we removed the last layer of the eyes model (ECNN) and replaced it with the nose classes. After that, we trained the model using 50 epochs (see section 6.1.1.2) with the same hyperparameters used previously. The new nose model is stored as NCNN and this process is then applied to all remaining parts of the face datasets.
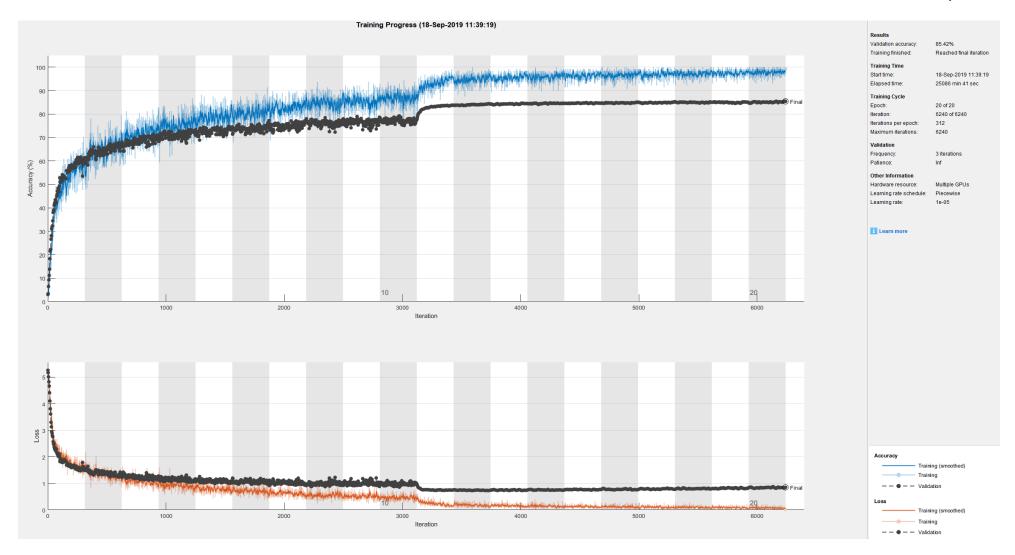
Figure 6-4: A visual displaying the training progress by using the eyes image dataset [203]**. At the top, there are two line graphs, the blue line denotes the training accuracy and the black line denotes the validation accuracy. The lower graph also shows two line graphs (indicated by the black and orange colour which denote training and validation loss). The two columns on the top right have the details about Results, Training Time, Training Time, Training Cycle, and Validation.**

Figure 6-5: A sample of the eye images generated from VGGface data [203]**.**

6.2 Experiments

This section presents our experimental results and evaluate the performance of our proposed system. Our implementation was done by using Maltab R2018a and carried on a machine with the following descriptions: OS Name Microsoft Windows 10 Home Premium; Processor Intel(R) Core (TM) i12-3770 CPU @ 3.40GHz, 3401 Mhz, 4 Core(s), 8 Logical Processor(s); Installed Physical Memory (RAM) 16.0 GB. The image data set which was used to train all the models originated from the VGGFace dataset, and method evaluation was conducted via face dataset, namely, the FEI [22].

6.2.1 Data Set

Our training images originate from a large-scale face dataset namely VGGFace image database [203]. This database comprises of more than 2 million high-resolution labelled images, which belong to 2622 subjects. The images were downloaded from the internet and they consist of multiple variances, relating to age, illumination, pose, and ethnicity. Figure 6-6 shows some examples of the images. For the whole-face trials, randomly approximately 70000 images were picked and in order to achieve promising results, all images were cropped using Dlib algorithm [180] to remove unwanted objects (such as backgrounds). The resultant images after the cropping process are shown in Figure 6-7. From the

cropped images, five different collections were generated for each part of the face (Eyes - Nose - Mouth – Forehead and combined Eyes+Nose) via Dlib face features detection as described in section 5.1.1 of Chapter 5 (See Figure 6-8 for the illustrative example).

Figure 6-6: A sample of whole-face images taken from the VGGFace image dataset(pre-cropping) [203]**. All of the selected images still have their background, and a majority of the images have objects such as hands, glasses, and a mic present.**
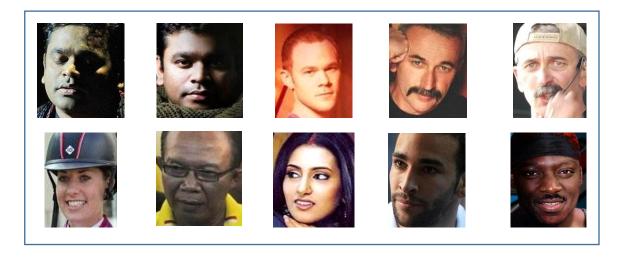
Figure 6-7: A sample of face images taken from the VGGFace dataset [203]**, (post-cropping). All of the face images were cropped by removing their backgrounds via the Dlib algorithm.**
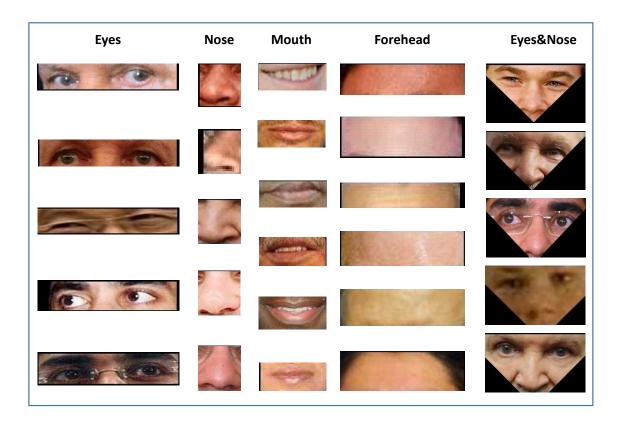
Figure 6-8: An example of some facial regions originally taken from VGGface data [203] **and cropped using the Dlib algorithm** [180]**. The first column shows images of eyes for five different people and in two cases, with and without glasses. The second column illustrates cropped noses, column three presents cropped mouths, Forehead crops are shown in column four, and lastly, images of just the eyes and nose is present in column five.**

6.2.2 Experiment Methodology

For each testing condition, the data from each part of the face was divided into two-sets; training and a validation set with the following split: 70% and 30% respectively. Moreover, to validate our results, the prediction performance of the five generated models was compared by two different methods, and Tables 6-1 and 6-2 in sections 6.2.3 and 6.2.4 present more details about the comparison and evaluation of the system.

6.2.3 Comparison with Other Models

The prediction performance of the five generated models were compared by two methods when utilising the same dataset. In the first method, the transfer learning approach was used to train an existing AlexNet [204]. Secondly, a pretrained model based method was selected for features extraction and known

classifiers Linear Support Vector Machines(LSVMs) [153] and Cosine Similarity (CS) [165] were used for classification.

A transfer learning strategy is one of the methods that is used to train models when a dataset does not have enough data for training. The mechanism of transfer learning is a transfer of previously acquired knowledge from a model that has been trained to perform a specific task to improve the learning process and perform another task, instead of training it from scratch. In this research, the entire model was trained for transfer learning because this approach enables the re-updating of all network weights and increases the performance of the model.

The transfer learning was conducted on the architecture of pre-trained models with weights. The Alexnet [204] model is trained on a large database containing roughly 1000 categories and generates feature maps of 2048. Before the process of fine-tuning, we replaced the classification layer in the Alexnet model with the categories of our data, then the same training method referred to in Section 6.1.1.2 was followed to train the model on five different facial databases (eyes - nose - mouth – forehead – combined eyes and nose). The training results show that the recognition rates from our models are better when compared to Alexnet in five chosen parts of the face. In addition, our five models generate features of 200 dimensions compared to 2048 from Alexnet. A comparison of our results is presented in Table 6-1.

Method	Number of Features	Percentage Rate (%) for the Recognition of Each Face Part				
		Mouth	Nose	Forehead	Eyes	Eyes & Nose
VGGFace [41] **& CS**	4095	25.71	28.54	16.87	38.25	76.51
VGGFace [41]**& LSVMs**	4095	32.12	39.08	13.69	51.25	80.72
FaceNet [87] **& CS**	128	11.91	15.16	12.11	24.97	74.46
FaceNet [87] **& LSVMs**	128	11.28	16.13	10.92	32.21	77.38
Alexnet [204]	2048	41.25	42.24	37.95	64.75	74.13
Our Model	**200**	**69.00**	**84.02**	**75.16**	**85.42**	**92.47**

Table 6-1: A table demonstrates the results of our model's performance compared to five different models when using images from the VGGFace data set [203]**. This comparison illustrates the recognition percentage (%) for each facial part (mouth, nose, forehead, eyes and eyes and nose).**

For the second experimental approach we used pre-trained models for features extraction, whereby an input image was pre-processed by a model to give an output i.e. vector of numbers. These features were then used as inputs to train a new model. In our experiments, two models were used for feature extraction (VGGface [41] and FaceNet [87]) and for classification CS and LSVMs were utilised. As shown in Table 6-1, our method outperformed all the other models by achieving the highest recognition within each experimental condition.

6.2.4 System Evaluation

This section evaluates the results of five models on a wider face image dataset, namely FEI (see section 4.2.1 in chapter 4).

In these experiments, previously generated models were utilised for the purpose of feature extraction by using the Fully Connected layer (FC1). Images from five databases were used to represent parts of the face (eyes, mouth, nose, forehead, combined eyes and nose). Each feature group was divided into two sets, the first set was used for training and consisted of 70% of the images, while the remaining 30% of images were used for testing. For the purpose of classification, two popular classifiers were applied; Cosine Similarity (CS) and Linear Support Vector Machines (LSVMs). From the experiments conducted using partial face features, the results demonstrated that 200 features are sufficient and credible for distinguishing between faces during multi-classes problems. The recognition rates for our five models (ECNN, NCNN, MCNN, FCNN, and ENCNN) exceeded the state-of-the-art work as published by Zheng et al., [106], who proposed Reinforced Centrosymmetric Local Binary Pattern (RCSLBP) as an efficient method for feature extraction. Furthermore, our models outperformed methods published by Elmahmudi and Ugail [44] who applied the VGG-face model for features extraction with 4095 features per face part, see Table 6-2.

Method		Number of features	Percentage Rate (%) for the Recognition of Each Face Part				
			Mouth	Nose	Forehead	Eyes	Eyes & Nose
Elmahmudi et al., [45]	CS	4095	13.83	14.00	34.33	64.66	90.00
	LSVMs	4095	16.00	11.33	16.83	38.33	56.33
FaceNet[87]	CS	128	10.66	**31.00**	33.5	79.83	**97**
Zheng et al., [106]		XX	16.48	16.91	38.27	69.30	93.01
Our		200	**51.00**	30.5	**44.5**	**90.33**	84.5

Table 6-2: An evaluation of our models using the FEI face image dataset [22]. **This table presents the results of testing our generated models for a face recognition task against three other methods based on the number of features and the recognition rate (%).**

6.3   Discussion

From what is presented as part of the experiments, using a small number of features and reducing the number of convolutional layers concludes good results, when compared to using 4095 features. After building the first model (ECNN), it was possible to use the transfer learning process to build other models (NCNN, MCNN, FCNN, and ENCNN), subsequently leading to a reduction in training time. We noticed that by presenting the results obtained using VGGFace and Alexnet models, for the five parts of the face (mouth, nose, forehead, eyes, and combined eyes and nose), the recognition rate (%) increased from 41.25%, 42.24%, 37.95%, 64.75% and 74.13%) for Alexnet, and (32.12%, 39.08%, 13.69%, 51.25% and 80.72%) for VGGface model respectively, to 69.00%, 84.02%, 75.16%, 85.42% and 92.47% when applying our models. In addition, a significant improvement was observed in the rate of recognition for smaller regions of the face such as the mouth and nose, which increased from 41.25% and 42.24% (using the previous models) to 69.00% and 84.02% using MCNN and NCNN. This indicates that when the model is trained on a specific part of the

face using the proposed architecture, it helps to improve the training process in terms of time and accuracy. Moreover, when the five generated models (ECNN, MCNN, FCNN, NCNN, and ENCNN) were selected for feature extraction using FEI face database images, and trained using SVMs and CS, the results obtained are greatly improved. For example, the recognition rates for the region of the eyes enhanced from 79.83% using FaceNet [87] to 90.33% via ECNN.

6.4 Conclusion

The work presented in this chapter was aimed to identify effective models for partial-face recognition issues. Five novel models were introduced for common regions of the face (eyes, nose, mouth, forehead, and combined eyes and nose), based on deep convolutional neural networks. The proposed models were based loosely on the VGG16 structure. By reducing the number of convolutional layers and the size of the features vector, this led to a reduction in the number of training hyper-parameters and a reduction in the training time and required memory. Firstly, the Eyes model structure (ECNN) was built and trained it on an ample dataset of eye images, then the generated model was used to fine-tune other models. The experiments conclude that the five generated models surpass the other models, when trained on an uncontrolled face database (namely VGGFace) and evaluated using a controlled face database (FEI). The results show that the ability to distinguish between small parts of the face is improved by using only 200 features instead of 4095 features.

Moreover, our results also indicate that small parts of the face can be used as biometric markers for many applications such as forensic science.

7 Conclusions, Limitations and Recommendations for Future Work

This chapter summarizes the research work and highlights some of the possible limitations of the work conducted on this subject. Further discussion is also presented about the future direction in which the research can be taken forward.

7.1 Conclusions

Computer based face recognition is one of the most interesting subjects in present-day visual computing. Face recognition has come a long way but still, there are many challenges and hurdles to overcome. Those challenges include, for example, recognising faces in poor illumination, pose variations, partial face, inverted faces, and aged faces. This work detailed some of these challenges along with some of the techniques we have developed to improve face recognition, specifically towards enhancing computer-based face recognition. Through our research, we intensively studied different aspects of the above challenges and we have introduced efficient and robust algorithms for detecting and analysing face recognition.

Firstly, we investigated the use of averaging over faces as an efficient computational entity for face processing. More specifically, utilising the concept of EVD to compute average weights over facial images in the Eigenspace. Our approach has a competitive advantage for face processing and analysis, e.g., compared to the standard average face computed in the pixel space through image pre-processing and transformations, the proposed method uses a handful of training images for computing the average weights in the Eigenspace for faces, resulting in a decreased level of complexity. Our approach computed the average

weights from a selection of known images avoiding the need to manually separate the face from the background. To demonstrate the capacity of this approach, face recognition is used as an example, i.e. recognition is performed by comparing the average weights of a probe face image to that of pre-computed average weights of the query images available within a database. Additionally, a varied image selection technique has been used to compute the average weights, through which we were able to measure the effect of image numbers on recognition rates. It is observed during experimentation that beyond a certain low number, increasing the number of images to compute the average weights has no significant impact on the rate of recognition.

For dimensionality reduction, a novel modification to the method of PCA was introduced, utilising the inherent averaging ability of the discrete Biharmonic operator as a pre-processing step called BiPCA. New images of reduced size can be generated by applying the Biharmonic operator to images keeping the features in them intact. Moreover, the resulting images of lower dimensionality can significantly reduce computational complexities.

We explored the question that surrounds the idea of face recognition using facial features. We explored them by developing novel experiments to test the performance of machine learning using parts of the face. In particular, we studied the proportion of recognition for parts, such as eyes, mouth, nose and cheek. In addition, we also study the effect of face recognition subject to facial rotation as well as the effect of recognition subject to zoom out of the facial images. Our experiments are based on two approaches, the first methodology implemented the state of the art Convolutional Neural Network based architecture along with the pre-trained VGG-Face model through which we extracted features. We then utilised two classifiers namely the Cosine Similarity and the Linear Support

Vector machines to test the recognition rates. In the second method, we introduce five new models for partial face recognition and each model was trained on a different region of the face (Mouth, Nose, Forehead, Eyes, and combined Eyes and Nose). Firstly, we built the Eyes model structure and trained it on a large dataset of eyes with a smaller number of trainable parameters, and then we used it to fine-tune other models. The experimental results indicate that the new models achieved higher performance compared to other models, by training them on an uncontrolled face database namely VGGFace, and evaluating them using a controlled face database (FEI). The results showed that the ability to distinguish small parts of the face has improved by using only 200 features instead of using 4095 features. In addition, based on our results, the small parts are promising and can be used as part of biometric systems for many applications within security.

Finally, we investigated the concept of age progression and regression by building a novel approach based on an average template to generate a new age for an input face. We conducted experiments and tested the proposed method on a public domain face dataset of known and unknown ages (FEI & Morph II). Our results demonstrate that the proposed method achieves high quality aged faces even if the input face is of low quality. To find out which estimated age could be an obstacle for regression and progression, all generated faces were tested for face recognition via the pre-trained VGG-Face model, for feature extraction, and classification utilised four classifiers to find a recognition rate per age. The results shown there is a high similarity between generated faces and the ground truth images.

7.2   General Limitations

This section lists limitations of the work which can be considered for future work as described below:

7.2.1   Face Pose Correction

Through the analysis of some face images taken from the face datasets and model evaluation for partial face recognition, there was some shortage of performance of recognition based on certain parts of the face. This shortage occurs when our algorithm tries to extract face features in cases such as face profile when it is only possible to see the cheek or one eye. Such features cannot probably be seen in certain situations and can affect the process of classification.

7.2.2   Environmental Condition

The age progression and regression framework report a high performance when addressing the ageing problem uses carefully selected images. However, running the method on faces taken from an uncontrolled environment such as illumination (faces with very high or low lighting) and cosmetic modifications are still a challenge.

7.2.3   Face Datasets

Our face parts datasets were generated by using face landmarks detection algorithm after which all extracted facial features were checked manually to remove unwanted items, especially the landmarks algorithm in some cases extracts irrelevant objects. In addition, our proposed system of age progression

and regression considers five different ethnicities. However, in some ethnicities we do not have enough representative images for different ages, for example, there are no images for Black Africa subjects at ages 40 and 50 years. Thus, dataset can be made more heterogeneous, diverse and complex, which as a result will increase the robustness and efficiency of the proposed model with a richer training.

7.3   Future Work

Firstly, following on from chapter 4 and 6, more investigation is required to investigate thoroughly face recognition based on parts of the face, by studying a combination of faced regions. There are other possibilities to combine other parts of the face such as the forehead and eyes or nose and mouth. In addition, generating new datasets of face parts can assist and save time for researchers. Further, we have demonstrated the robustness of CNNs, in particular, the application for feature extraction and the analysis of imperfect facial data. In addition, following on from chapter 6, this is scope for researcher to build other models for the rest of the face, which can be done in future by utilising Fine-Tuning strategies for our pre-trained. Furthermore, combing all the face part models in one implementation system can promote a significant performance of the recognition.

From an application perspective, we believe the experimental framework is still in its early stages, since only publicly available datasets have been utilised, and the images are far removed from practical scenarios. Therefore, it will be useful to extend this work to assess its practical applicability in terms of extending

our experiments where for example, images of faces from CCTV footage may be used as recognition cues.

Moreover, there are other frameworks and techniques that can be employed to develop and test facial recognition cues, using various parts of the face. General Adversarial Networks (GANs) is an example, which is increasingly becoming popular as a tool for machine learning. It can be used as a process to complete missing parts of a face, based on some criteria of learning methodologies.

Finally, in chapter 5, we have demonstrated the possibility of using the formulations of an average face of a given ethnicity, gender, and age to generate a new face in different ages. As we mentioned in the previous section, we used just five various ethnicities for given ages. However, it is plausible to extend that by increasing the number of ethnicities and age groups. Further, it is possible to develop our method to generate new ages even if an input face is just a profile face. There is also a suggestion to correct a pose of a face and after that apply the ageing process.

Reference

[1]   A. K. Jain, "Biometric recognition," *Nature*, vol. 449, no. 7158, pp. 38–40, 2007.

[2]   V. Bruce and A. W. Young, *Face perception*. Psychology Press, 2012.

[3]   A. F. Chapman, H. Hawkins-Elder, and T. Susilo, "How robust is familiar face recognition? A repeat detection study of more than 1000 faces," *R. Soc. open Sci.*, vol. 5, no. 5, p. 170634, 2018.

[4]   N. M. Blauch, M. Behrmann, and D. C. Plaut, "Computational insights into human perceptual expertise for familiar and unfamiliar face recognition," *Cognition*, p. 104341, 2020.

[5]   H. Wechsler, Ed., "Face Representation BT - Reliable Face Recognition Methods: System Design, Impementation and Evaluation," Boston, MA: Springer US, 2007, pp. 65–96.

[6]   S. M. Andersen, C. A. Carlson, M. A. Carlson, and S. D. Gronlund, "Individual differences predict eyewitness identification performance," *Pers. Individ. Dif.*, vol. 60, pp. 36–40, 2014.

[7]   A. W. Young, D. Hellawell, and D. C. Hay, "Configurational information in face perception," *Perception*, vol. 42, no. 11, pp. 1166–1178, 2013.

[8]   J. Murphy, K. L. H. Gray, and R. Cook, "The composite face illusion," *Psychon. Bull. Rev.*, vol. 24, no. 2, pp. 245–261, 2017.

[9]   J. W. Tanaka and D. Simonyi, "The 'parts and wholes' of face recognition: A review of the literature," *Q. J. Exp. Psychol.*, vol. 69, no. 10, pp. 1876–1889, Oct. 2016.

[10] J. W. Tanaka, M. D. Kaiser, S. Hagen, and L. J. Pierce, "Losing face : impaired discrimination of featural and configural information in the mouth region of an inverted face," *Attention, Perception, Psychophys.*, vol. 76, no. 4, pp. 1000–1014, 2014.

[11] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Comput. Vis. Image Underst.*, vol. 189, p. 102805, 2019.

[12] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep Face Recognition: A Survey," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2018, pp. 471–478.

[13] H. Zhou and K. M. Lam, "Age-invariant face recognition based on identity inference from appearance age," *Pattern Recognit.*, vol. 76, pp. 191–202, 2018.

[14] T. Cootes and A. Lanitis, "FG-NET aging database," 2008. [Online]. Available: http://www.fgnet.rsunit.com/. [Accessed: 07-Jul-2018].

[15] D. Gong, D. Tao, X. Li, and J. Liu, "A Maximum Entropy Feature Descriptor for Age Invariant Face Recognition Shenzhen Key Lab of Computer Vision and Pattern Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5289–5297.

[16] U. Park, Y. Tong, and A. K. Jain, "Age-Invariant Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 947–954, 2010.

[17] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs, "Face Verification Across Age Progression Using Discriminative Methods," *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 1, pp. 82–91, 2010.

[18] Z. Li, U. Park, and A. K. Jain, "A Discriminative Model for Age Invariant Face Recognition," *IEEE Trans. Inf. forensics Secur.*, vol. 6, no. 3, pp. 1028–1037, 2011.

[19] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

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

[21] X. W. X. Wang and X. T. X. Tang, "Random sampling LDA for face recognition," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, 2004, vol. 2, pp. 259–265.

[22] C. E. Thomaz, "FEI Face Database," *The Artificial Intelligence Laboratory of FEI in So Bernardo do Campo, So Paulo, Brazil*. [Online]. Available: https://fei.edu.br/~cet/facedatabase.html. [Accessed: 23-Oct-2018].

[23] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning Continuous Face Age Progression: A Pyramid of GANs," *IEEE Trans. Pattern Anal. Mach. Intell.*, p. 1, 2019.

[24] E. Pantraki and C. Kotropoulos, "FACE AGING AS IMAGE-TO-IMAGE TRANSLATION USING SHARED-LATENT SPACE GENERATIVE ADVERSARIAL NETWORKS," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2018, pp. 306–310.

[25] D. Burt and D. Perrett, "Perception of age in adult Caucasian male faces :

computer graphic manipulation of shape and colour information," *Proc. R. Soc.*, vol. 259, no. 1355, pp. 137–143, 1995.

[26] C. Choi and C. Engineering, "Age change for predicting future faces," *Proc. IEEE Int. Fuzzy Syst. Conf.*, vol. 3, pp. 1603–1608, 1999.

[27] Y. H. Kwon and V. Lobo, "Age Classification from Facial Images," *Comput. Vis. Image Underst.*, vol. 74, no. 1, pp. 1–21, 1999.

[28] T. Zheng, W. Deng, and J. Hu, "Age Estimation Guided Convolutional Neural Network for Age-Invariant Face Recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2017, vol. 10, pp. 503–511.

[29] W. Hwang, W. Kim, S. Suh, J.-J. Han, and J. Kim, "Face recognition using average example image," *Electron. Lett.*, vol. 50, no. 25, pp. 1921–1923, 2014.

[30] A. M. Aguilera, M. Escabias, and M. J. Valderrama, "Using principal components for estimating logistic regression with high-dimensional multicollinear data," *Comput. Stat. Data Anal.*, vol. 50, no. 8, pp. 1905–1924, 2006.

[31] R. do E. Santo, "Principal Component Analysis applied to digital image compression," *Einstein (Sao Paulo).*, vol. 10, no. 2, pp. 135–139, 2012.

[32] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proceedings. 1991 IEEE computer society conference on computer vision and pattern recognition*, 1991, pp. 586–587.

[33] A. M. Burton, R. Jenkins, P. J. B. Hancock, and D. White, "Robust

representations for face recognition: The power of averages," *Cogn. Psychol.*, vol. 51, no. 3, pp. 256–284, 2005.

[34] R. Jenkins and A. M. Burton, "100% Accuracy in automatic face recognition," *Science (80-.).*, vol. 319, no. 5862, p. 435, 2008.

[35] J. Shah, M. Sharif, M. Raza, and A. Azeem, "A survey: Linear and nonlinear PCA based face recognition techniques," *Int. Arab J. Inf. Technol.*, vol. 10, no. 6, 2013.

[36] Y. Wang and Y. Wu, "Complete neighborhood preserving embedding for face recognition," *Pattern Recognit.*, vol. 43, no. 3, pp. 1008–1015, 2010.

[37] M. Marletta, "Eigenvalue problems on exterior domains and Dirichlet to Neumann maps," *J. Comput. Appl. Math.*, vol. 171, no. 1–2, pp. 367–391, 2004.

[38] K. Meena and A. Suruliandi, "Local binary patterns and its variants for face recognition," *Int. Conf. Recent Trends Inf. Technol. ICRTIT 2011*, pp. 782–786, 2011.

[39] A. Smola and S. V. N. Vishwanathan, "Introduction to machine learning," *Cambridge Univ. UK*, vol. 32, p. 34, 2008.

[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

[41] P. Omkar, V. Andrea, and Z. Andrew, "Deep Face Recognition," in *Procedings of the British Machine Vision Conference 2015*, 2015.

[42] A. Elmahmudi, H. Ugail, and A. Bukar, "Average Weights of Eigenvalue

Decomposition for Efficient Face Image Processing," in *INTECH 2018*, 2018.

[43]   A. Elmahmudi and H. Ugail, "The Biharmonic Eigenface," *Signal, Image Video Process.*, vol. 13, no. 8, pp. 1639–1647, 2019.

[44]   A. Elmahmudi and H. Ugail, "Deep face recognition using imperfect facial data," *Futur. Gener. Comput. Syst.*, vol. 99, pp. 213–225, 2019.

[45]   A. Elmahmudi and H. Ugail, "Experiments on Deep Face Recognition Using Partial Faces," in *2018 International Conference on Cyberworlds (CW)*, 2018, pp. 357–362.

[46]   A. Elmahmudi and H. Ugail, "A framework for facial age progression and regression using exemplar face templates," *Vis. Comput.*, 2020.

[47]   R. Koenen, "Mpeg-4 Project Overview, International Organisation for Standardization," ISO/IEC/JTC1/SC29/WG11, La Baule, 2000.

[48]   D. J. Robertson, R. S. S. Kramer, and A. M. Burton, "Face averages enhance user recognition for smartphone security," *PLoS One*, vol. 10, no. 3, Mar. 2015.

[49]   B. Subbiah and S. C. Christopher, "Image classification through integrated K-Means algorithm," *Int. J. Comput. Sci. Issues*, vol. 9, no. 2, pp. 518–524, 2012.

[50]   T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.

[51]   N. Dalal and B. Triggs, "Histograms of oriented gradients for human

detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 1, pp. 886–893 vol. 1.

[52] D.-H. Liu, K.-M. Lam, and L.-S. Shen, "Optimal sampling of Gabor features for face recognition," *Pattern Recognit. Lett.*, vol. 25, no. 2, pp. 267–276, 2004.

[53] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, 2001.

[54] I. Dagher, "Incremental PCA-LDA algorithm," in *2010 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, 2010, pp. 97–101.

[55] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Guide to the CMU Multi-PIE database," *Tech. report-Carnegie Mellon Univ.*, 2007.

[56] W. Lee, M. Cheon, C.-H. Hyun, and M. Park, "Best Basis Selection Method Using Learning Weights for Face Recognition," *Sensors* , vol. 13, no. 10. 2013.

[57] M. Çarıkçı and F. Özen, "A face recognition system based on eigenfaces method," *Procedia Technol.*, vol. 1, pp. 118–123, 2012.

[58] "Face Recognition Data." [Online]. Available: https://cswww.essex.ac.uk/mv/allfaces/index.html. [Accessed: 29-Mar-2020].

[59] P.-C. Hsieh and P.-C. Tung, "A novel hybrid approach based on sub-pattern technique and whitened PCA for face recognition," *Pattern*

Recognit., vol. 42, no. 5, pp. 978–984, 2009.

[60] A. J. Bell and T. J. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vision Res.*, vol. 37, no. 23, pp. 3327–3338, 1997.

[61] A. Hyvärinen, "Independent component analysis: recent advances," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 371, no. 1984, p. 20110534, 2013.

[62] "Extended Yale Face Database B (B+)." [Online]. Available: http://vision.ucsd.edu/content/extended-yale-face-database-b-b. [Accessed: 23-Mar-2020].

[63] "Weizmann face database," *Computer Vision DataBases.* [Online]. Available: http://www.wisdom.weizmann.ac.il/~vision/databases.html. [Accessed: 23-Mar-2020].

[64] G. Ghinea, R. Kannan, and S. Kannaiyan, "Gradient-Orientation-Based PCA Subspace for Novel Face Recognition," *IEEE Access*, vol. 2, pp. 914–920, 2014.

[65] C. Paige and C. Van Loan, "A Schur decomposition for Hamiltonian matrices," *Linear Algebra Appl.*, vol. 41, pp. 11–32, 1981.

[66] M.-. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," in *Proceedings of 12th International Conference on Pattern Recognition*, 1994, vol. 1, pp. 566–568 vol.1.

[67] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," *Am. Stat.*, vol. 46, no. 3, pp. 175–185, Aug. 1992.

[68] "The Database of Faces, Cambridge University Computer Laboratory." [Online]. Available: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html. [Accessed: 06-May-2017].

[69] B. Poon, M. A. Amin, and H. Yan, "Improved methods on PCA based human face recognition for distorted images," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2016, vol. 1.

[70] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using LDA-based algorithms," *IEEE Trans. Neural Networks*, vol. 14, no. 1, pp. 195–200, 2003.

[71] S.-K. Oh, S.-H. Yoo, and W. Pedrycz, "Design of face recognition algorithm using PCA -LDA combined for hybrid data pre-processing and polynomial-based RBF neural networks : Design and its application," *Expert Syst. Appl.*, vol. 40, no. 5, pp. 1451–1466, 2013.

[72] S. Suganya and D. Menaka, "Performance Evaluation of Face Recognition Algorithms," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 2, no. 1, pp. 135–140, 2014.

[73] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognit.*, vol. 39, no. 9, pp. 1725–1745, 2006.

[74] M. Kan, S. Shan, Y. Su, X. Chen, and W. Gao, "Adaptive discriminant analysis for face recognition from single sample per person," in *Face and Gesture 2011*, 2011, pp. 193–199.

[75] M. Savvides, R. Abiantun, J. Heo, S. Park, C. Xie, and B. V. K. Vijayakumar, "Partial & Holistic Face Recognition on FRGC-II data using Support Vector Machine Kernel Correlation Feature Analysis," in *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on. IEEE*, 2006, pp. 48–48.

[76] M. Savvides, "Face Verification using Correlation Filters," *3rd IEEE Autom. Identif. Adv. Technol.*, no. 1, pp. 56–61, 2002.

[77] L. He, H. Li, Q. Zhang, and Z. Sun, "Dynamic Feature Learning for Partial Face Recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7054–7063.

[78] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.

[79] S. Li, D. Yi, Z. Lei, and S. Liao, "The casia nir-vis 2.0 face database," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2013, pp. 348–353.

[80] Y. Long, F. Zhu, L. Shao, and J. Han, "Face recognition with a small occluded training set using spatial and statistical pooling," *Inf. Sci. (Ny).*, vol. 430–431, pp. 634–644, 2018.

[81] a M. Martinez and R. Benavente, "The AR face database," 1998.

[82] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," 2008.

[83]   B. Lahasan, S. Lebai Lutfi, I. Venkat, M. A. Al-Betar, and R. San-Segundo, "Optimized symmetric partial facegraphs for face recognition in adverse conditions," *Inf. Sci. (Ny).*, vol. 429, pp. 194–214, 2018.

[84]   Y. Duan, J. Lu, J. Feng, and J. Zhou, "Topology Preserving Structural Matching for Automatic Partial Face Recognition," *EEE Trans. Inf. Forensics Secur.*, vol. 13, no. 7, pp. 1823–1837, 2018.

[85]   K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, 2005.

[86]   N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 365–372.

[87]   F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[88]   R. Verschae, J. Ruiz-del-Solar, and M. Correa, "Face recognition in unconstrained environments: A comparative study," 2008.

[89]   H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang, "Probabilistic elastic matching for pose variant face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3499–3506.

[90]   J. Cai, J. Chen, and X. Liang, "Single-Sample Face Recognition Based on Intra-Class Differences in a Variation Model," *Sensors*, vol. 15, no. 1, pp. 1071–1087, 2015.

[91]   J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.

[92]   H. Li and C. Y. Suen, "Robust face recognition based on dynamic rank representation," *Pattern Recognit.*, vol. 60, pp. 13–24, 2016.

[93]   Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Ann. Transl. Med. Vol 4, No 11 (June 2016) Ann. Transl. Med.*, vol. 4, no. 11, 2016.

[94]   M. G. Calvo, A. Fernández-Martín, and L. Nummenmaa, "Facial expression recognition in peripheral versus central vision: role of the eyes and the mouth," *Psychol. Res.*, vol. 78, no. 2, pp. 180–195, 2014.

[95]   B. Olk and A. M. Garay-Vado, "Attention to faces : Effects of face inversion," *Vision Res.*, vol. 51, no. 14, pp. 1659–1666, 2011.

[96]   J. Taubert, D. Apthorp, D. Aagten-Murphy, and D. Alais, "The role of holistic processing in face perception: Evidence from the face inversion effect," *Vision Res.*, vol. 51, no. 11, pp. 1273–1278, 2011.

[97]   J. J. Richler, M. L. Mack, T. J. Palmeri, and I. Gauthier, "Inverted faces are (eventually) processed holistically," *Vision Res.*, vol. 51, no. 3, pp. 333–342, 2011.

[98]   A. de Heering, B. Rossion, and D. Maurer, "Developmental changes in face recognition during childhood: Evidence from upright and inverted faces," *Cogn. Dev.*, vol. 27, no. 1, pp. 17–27, 2012.

[99]   J. Murphy and R. Cook, "Revealing the mechanisms of human face

perception using dynamic apertures," *Cognition*, vol. 169, pp. 25–35, 2017.

[100] A. M. Megreya and A. M. Burton, "Unfamiliar faces are not faces: Evidence from a matching task," *Mem. Cognit.*, vol. 34, no. 4, pp. 865–876, 2006.

[101] X. Yan and B. Rossion, "A robust neural familiar face recognition response in a dynamic (periodic) stream of unfamiliar faces," *Cortex*, vol. 132, pp. 281–295, 2020.

[102] L. Du and H. Hu, "Nuclear norm based adapted occlusion dictionary learning for face recognition with occlusion and illumination changes," *Neurocomputing*, vol. 340, pp. 133–144, 2019.

[103] M. Liao and X. Gu, "Face recognition based on dictionary learning and subspace learning," *Digit. Signal Process.*, vol. 90, pp. 110–124, 2019.

[104] T. Xiao, P. Liu, W. Zhao, H. Liu, and X. Tang, "Structure preservation and distribution alignment in discriminative transfer subspace learning," *Neurocomputing*, vol. 337, pp. 218–234, 2019.

[105] A. Appice, P. Guccione, and D. Malerba, "Transductive hyperspectral image classification: toward integrating spectral and relational features via an iterative ensemble system," *Mach. Learn.*, vol. 103, no. 3, pp. 343–375, 2016.

[106] W. Zheng, C. Gou, and F.-Y. Wang, "A novel approach inspired by optic nerve characteristics for few-shot occluded face recognition," *Neurocomputing*, vol. 376, pp. 25–41, 2020.

[107] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-miller, "Labeled Faces in the Wild : A Database for Studying Face Recognition in Unconstrained

Environments," in *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008, pp. 1–11.

[108] S. A. Peixoto *et al.*, "A high-efficiency energy and storage approach for IoT applications of facial recognition," *Image Vis. Comput.*, vol. 96, p. 103899, 2020.

[109] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," in *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 2016, pp. 78–83.

[110] L. Zhou, S. Pan, J. Wang, and A. V Vasilakos, "Machine learning on big data: Opportunities and challenges," *Neurocomputing*, vol. 237, pp. 350–361, 2017.

[111] V. Bruce, P. J. B. Hancock, and A. M. Burton, "Human Face Perception and Identification," in *Face Recognition: From Theory to Applications*, H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulié, and T. S. Huang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 51–72.

[112] I. Goodfellow *et al.*, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2672–2680.

[113] Y. Lu, S. Wang, W. Zhao, and Y. Zhao, "WGAN-Based Robust Occluded Facial Expression Recognition," *IEEE Access*, vol. 7, pp. 93594–93610, 2019.

[114] S. Li, W. Deng, and J. P. Du, "Reliable crowdsourcing and deep locality-

preserving learning for expression recognition in the wild," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, pp. 2852–2861.

[115] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 18–31, 2019.

[116] A. M. Bukar, H. Ugail, and D. Connah, "Individualised model of facial age synthesis based on constrained regression," in *2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2015, pp. 285–290.

[117] T. F. Vieira, A. Bottino, A. Laurentini, and M. De Simone, "Detecting siblings in image pairs," *Vis. Comput.*, vol. 30, no. 12, pp. 1333–1345, 2014.

[118] K. A. Dalrymple, J. Gomez, and B. Duchaine, "The Dartmouth Database of Children ' s Faces : Acquisition and Validation of a New Face Stimulus Set," *PLoS One*, vol. 8, no. 11, pp. 1–7, 2013.

[119] A. M. Bukar and H. Ugail, "Convnet features for age estimation," in *11th International Conference on Computer Graphics, Visualization, Computer Vision and Image Processing*, 2017.

[120] H. WOLD, "11 - Path Models with Latent Variables: The NIPALS Approach**NIPALS = Nonlinear Iterative PArtial Least Squares.," in *International Perspectives on Mathematical and Statistical Modeling*, H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon, and V. B. T.-Q. S. Capecchi, Eds. Academic Press, 1975, pp. 307–357.

[121] K. Ricanek and T. Tesafaye, "MORPH: a longitudinal image database of

normal adult age-progression," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, 2006, pp. 341–345.

[122] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 112–119.

[123] G. Guo and G. Mu, "Human age estimation: What is the influence across race and gender?," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2010, pp. 71–78.

[124] S. Riaz, U. Park, J. Choi, and P. Natarajan, "Age progression by gender-specific 3D aging model," *Mach. Vis. Appl.*, vol. 30, no. 1, pp. 91–109, 2019.

[125] W. Li, Z. Zhang, Y. Wang, and R. Morales, "Super-resolution in Tensor Space and Active Appearance Model Based Simulation of Adult Aging Effects on Face Images," in *2010 Chinese Conference on Pattern Recognition (CCPR)*, 2010, pp. 1–5.

[126] I. Matthews and S. Baker, "Active Appearance Models Revisited," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 135–164, 2004.

[127] Y. Wang, Z. Zhang, W. Li, and F. Jiang, "Combining Tensor Space Analysis and Active Appearance Models for Aging Effect Simulation on Face Images," *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 42, no. 4, pp. 1107–1118, 2012.

[128] X. Shu, J. Tang, H. Lai, L. Liu, and S. Yan, "Personalized Age Progression with Aging Dictionary," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3970–3978.

[129] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval," in *European conference on computer vision*, 2014, pp. 768–783.

[130] I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S. M. Seitz, "Illumination-aware age progression," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 3334–3341.

[131] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 4352–4360.

[132] Y. Wen, Z. Li, and Y. Qiao, "Latent Factor Guided Convolutional Neural Networks for Age-Invariant Face Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4893–4901.

[133] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," in *In International Conference on Learning Representations*, 2016, pp. 1–16.

[134] I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S. M. Seitz, "Illumination-aware age progression," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3334–3341.

[135] T. Lindeberg, "Scale invariant feature transform," vol. 7, no. 5. Computational Biology, CB, Computational Biology, CB, School of Computer Science and Communication (CSC), KTH, p. 10491, 2012.

[136] C. Xu, Q. Liu, and M. Ye, "Age invariant face recognition and retrieval by coupled auto-encoder networks," *Neurocomputing*, vol. 222, pp. 62–71, 2017.

[137] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, vol. B 39, no. 1. 1977.

[138] H. Zhou, K. Wong, and K. Lam, "Feature-aging for age-invariant face recognition," in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2015, pp. 1161–1165.

[139] D. Gong, Z. Li, D. Lin, J. Liu, and X. Tang, "Hidden factor analysis for age invariant face recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2872–2879.

[140] C. Herrmann, D. Willersinn, and J. Beyerer, "Low-resolution Convolutional Neural Networks for video face recognition," in *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2016, pp. 221–227.

[141] M. Deswal, N. Kumar, and N. Rathi, "Face recognition based on singular value decomposition linear discriminant analysis method," *Int J Eng Innov Technol*, vol. 3, p. 12, 2014.

[142] S. Cha, "Comprehensive Survey on Distance / Similarity Measures between Probability Density Functions," *City*, vol. 1, no. 4, 2007.

[143] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, 2001.

[144] P. Korshunov and T. Ebrahimi, "Using face morphing to protect privacy," in *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2013, pp. 208–213.

[145] E. Naz, U. Farooq, and T. Naz, "Analysis of Principal Component Analysis-Based and Fisher Discriminant Analysis-Based Face Recognition Algorithms," in *2006 International Conference on Emerging Technologies*, 2006, pp. 121–127.

[146] C. Clausen and H. Wechsler, "Color image compression using PCA and backpropagation learning," *Pattern Recognit.*, vol. 33, no. 9, pp. 1555–1560, 2000.

[147] K. Tan and S. Chen, "Adaptively weighted sub-pattern PCA for face recognition," *Neurocomputing*, vol. 64, pp. 505–511, 2005.

[148] R. Gottumukkal and V. K. Asari, "An improved face recognition technique based on modular PCA approach," *Pattern Recognit. Lett.*, vol. 25, no. 4, pp. 429–436, 2004.

[149] G. Chen, Z. Li, and P. Lin, "A fast finite difference method for biharmonic equations on irregular domains and its application to an incompressible Stokes flow," *Adv. Comput. Math.*, vol. 29, no. 2, pp. 113–133, 2008.

[150] G. D. Smith, G. D. Smith, and G. D. S. Smith, *Numerical solution of partial differential equations: finite difference methods*. Oxford university press, 1985.

[151] M. Shiffer, P. R. GARABEDIAN, and M. SCHIFFER, "Variational Problems in the Theory of Elliptic Partial Differential Equations," *J. Ration. Mech. Anal.*, vol. 2, pp. 137–171, Mar. 1953.

[152] T. N. Sainath, A. Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for LVCSR," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 8614–8618.

[153] S. Amarappa and S. V Sathyanarayana, "Data classification using Support vector Machine (SVM), a simplified approach," *Int. J. Electron. Comput. Sci. Eng*, vol. 3, pp. 435–445, 2014.

[154] M. A. Alsmirat, F. Al-Alem, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Impact of digital fingerprint image quality on the fingerprint recognition accuracy," *Multimed. Tools Appl.*, vol. 78, no. 3, pp. 3649–3688, 2019.

[155] S. Atawneh, A. Almomani, H. Al Bazar, P. Sumari, and B. Gupta, "Secure and imperceptible digital image steganographic algorithm based on diamond encoding in DWT domain," *Multimed. Tools Appl.*, vol. 76, no. 18, pp. 18451–18472, 2017.

[156] A. M. Bukar and H. Ugail, "Automatic age estimation from facial profile view," *IET Comput. Vis.*, vol. 11, no. 8, pp. 650–655, 2017.

[157] H. Ugail and A. Al-dahoud, "Is gender encoded in the smile? A computational framework for the analysis of the smile driven dynamic face for gender recognition," *Vis. Comput.*, vol. 34, no. 9, pp. 1243–1254, 2018.

[158] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[159] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

[160] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[161] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[162] Y. Zhong, R. Arandjelović, and A. Zisserman, "Faces In Places : Compound query retrieval," in *In BMVC-27th British Machine Vision Conference*, 2014.

[163] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1717–1724.

[164] A. Vedaldi and K. Lenc, "MatConvNet Convolutional Neural Networks for MATLAB," in *Proceedings of the 23rd ACM international conference on Multimedia. ACM*, 2015, pp. 689–692.

[165] G. Sidorov, A. Gelbukh, H. Gómez-Adorno, and D. Pinto, "Soft similarity and soft cosine measure: Similarity of features in vector space model," *Comput. y Sist.*, vol. 18, no. 3, pp. 491–504, 2014.

[166] S. Jilani, H. Ugail, and A. Logan, "The Computer Nose Best," in *2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, 2019, pp. 1–6.

[167] A. M. Bukar and H. Ugail, "CONVNET FEATURES FOR AGE ESTIMATION," in *International Conferences Computer Graphics, Visualization, Computer Vision and Image Processing*, 2017, no. 2010, pp.

94–102.

[168] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

[169] S. K. Murthy, "Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey," *Data Min. Knowl. Discov.*, vol. 2, no. 4, pp. 345–389, 1998.

[170] X. Han, H. Ugail, and I. Palmer, "Gender Classification Based on 3D Face Geometry Features Using SVM," in *2009 International Conference on CyberWorlds*, 2009, pp. 114–118.

[171] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. W. B. T.-P. of T. 33rd I. C. on M. Learning, "Autoencoding beyond pixels using a learned similarity metric," in *International conference on machine learning. PMLR*, 2016, vol. 48, pp. 1558–1566.

[172] H. V Nguyen and L. Bai, "Cosine similarity metric learning for face verification," in *Asian conference on computer vision*, 2010, pp. 709–720.

[173] A. C. Lorena, A. C. P. L. F. De Carvalho, and J. M. P. Gama, "A review on the combination of binary classifiers in multiclass problems," *Artif. Intell. Rev.*, vol. 30, no. 1, pp. 19–37, 2008.

[174] D. Wu, C. Li, J. Chen, D. You, and X. Xia, "A novel multi-class support vector machines using probability voting strategy and its application on fault diagnosis of gearbox," *Int. J. Performability Eng.*, vol. 10, no. 2, pp. 173–186, 2014.

[175] M. a. Bagheri, G. A. Montazer, and S. Escalera, "Error correcting output codes for multiclass classification: Application to two image vision problems," in *The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012)*, 2012, pp. 508–513.

[176] A. Passerini, M. Pontil, and P. Frasconi, "New results on error correcting output codes of kernel machines," *IEEE Trans. Neural Networks*, vol. 15, no. 1, pp. 45–54, 2004.

[177] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.

[178] T. Fletcher, "Support vector machines explained," *Tutor. Pap. Mar*, p. 28, 2009.

[179] A. Al-dahoud and H. Ugail, "A method for location based search for enhancing facial feature detection," in *Advances in Computational Intelligence Systems*, Springer, 2017, pp. 421–432.

[180] D. E. King, "Dlib-ml: A Machine Learning Toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, 2009.

[181] R. Larsen, "L 1 generalized procrustes 2D shape alignment," *J. Math. Imaging Vis.*, vol. 31, no. 2–3, pp. 189–194, 2008.

[182] M. Grundland, R. Vohra, G. P. Williams, and N. A. Dodgson, "Cross dissolve without cross fade: Preserving contrast, color and salience in image compositing," *Comput. Graph. Forum*, vol. 25, no. 3, pp. 577–586, 2006.

[183] A. Patel, "A Basic Knowledge of Image Morphing," *Int. Hournal Eng. Sci.*,

vol. 1, no. 2, pp. 2–5, 2015.

[184] S. W. Cheng, T. K. Dey, and J. R. Shewchuk, *Delaunay mesh generation*, 1st Editio. New York: CRC Press, 2012.

[185] P. Dong and N. P. Galatsanos, "Affine transformation resistant watermarking based on image normalization," *Image Process. 2002. Proceedings. 2002 Int. Conf.*, vol. 3, pp. 489–492, 2002.

[186] G. Yu and J. Morel, "A fully affine invariant image comparison method," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 1597–1600.

[187] D. Brunet, E. R. Vrscay, and Z. Wang, "On the Mathematical Properties of the Structural Similarity Index," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1488–1499, 2012.

[188] Y. Gao, J. Ma, and A. L. Yuille, "Semi-Supervised Sparse Representation Based Classification for Face Recognition With Insufficient Labeled Samples," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2545–2560, 2017.

[189] "How-Old.net." [Online]. Available: https://www.how-old.net/?q=#. [Accessed: 21-Jan-2019].

[190] W. Wang, Y. Yan, Z. Cui, J. Feng, S. Yan, and N. Sebe, "Recurrent Face Aging with Hierarchical AutoRegressive Memory," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 654–668, 2019.

[191] "IMGonline.com.ua," 2018. [Online]. Available: https://www.imgonline.com.ua/eng/similarity-percent.php. [Accessed: 12-

Mar-2019].

[192] V. Talreja, M. C. Valenti, and N. M. Nasrabadi, "Multibiometric secure system based on deep learning," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2017, pp. 298–302.

[193] Y. Zhang, C. Xu, S. Yu, H. Li, and X. Zhang, "SCLPV: Secure Certificateless Public Verification for Cloud-Based Cyber-Physical-Social Systems Against Malicious Auditors," *IEEE Trans. Comput. Soc. Syst.*, vol. 2, no. 4, pp. 159–170, 2015.

[194] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face recognition systems: A Survey," *Sensors*, vol. 20, no. 2, p. 342, 2020.

[195] C. Peng, N. Wang, J. Li, and X. Gao, "Re-Ranking High-Dimensional Deep Local Representation for NIR-VIS Face Recognition," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4553–4565, 2019.

[196] M. Wan, H. Zheng, G. Yang, T. Zhan, and W. Huang, "Tri-Direction 2D-Fisher Discriminant Analysis (T2D-FDA) for Feature Extraction," *IEEE Access*, vol. 7, pp. 114714–114720, 2019.

[197] D.-X. Zhou, "Universality of deep convolutional neural networks," *Appl. Comput. Harmon. Anal.*, vol. 48, no. 2, pp. 787–794, 2020.

[198] G. Wu, J. Tao, and X. Xu, "Occluded Face Recognition Based on the Deep Learning," in *2019 Chinese Control And Decision Conference (CCDC)*, 2019, pp. 793–797.

[199] Z. Chen, X.-J. Wu, and J. Kittler, "A sparse regularized nuclear norm based matrix regression for face recognition with contiguous occlusion," *Pattern*

Recognit. Lett., vol. 125, pp. 494–499, 2019.

[200] F. Cen and G. Wang, "Dictionary Representation of Deep Features for Occlusion-Robust Face Recognition," *IEEE Access*, vol. 7, pp. 26595–26605, 2019.

[201] J. Kim, H. Jo, M. Ra, and W. Kim, "Fine-tuning Approach to NIR Face Recognition," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 2337–2341.

[202] L. Busoniu, D. Ernst, B. De Schutter, and R. Babuska, "Cross-Entropy Optimization of Control Policies With Adaptive Basis Functions," *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 41, no. 1, pp. 196–209, 2011.

[203] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," *Proc. - 13th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2018*, pp. 67–74, 2018.

[204] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.