Gender Classification Using Facial Components

Author:
Mayibongwe H BAYANA

Supervisor:
Prof. Serestina VIRIRI
Co-Supervisor:
Dr. Raphael ANGULU

A thesis submitted in fulfillment of the requirements
for the degree of Masters of Science in computer science ,School of Mathematics
Statistics and Computer Science, University of Kwazulu Natal

August 30, 2018

Declaration of Authorship

I, Mayibongwe H BAYANA, declare that this thesis titled, "Gender Classification Using Facial Components" and the work presented in it are my own. I confirm that:

- The research reported in this thesis, except where otherwise indicated or acknowledged, is my original work.

- This thesis has not been submitted in full or in part for any degree or examination to any other university.

- This thesis does not contain other person's data, images or graphs, unless specifically acknowledged as being sourced from other persons.

- This thesis does not contain other person's writing, unless specifically acknowledge as being sourced from other researchers. Where other written sources have been quoted, then;
  (a) their words have been re-written but the general information attributed to them has been referenced;
  (b)where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

- This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Candidate:

Signed: ______________________________ Date: ______________________________
Mayibongwe H Bayana

As the candidate's supervisor i approve the submission of this thesis for examination

Supervisor:

Signed:______________________________ Date:______________________________
Prof Serestina Viriri

Co-Supervisor:

Signed:______________________________ Date:______________________________
Dr Raphael Angulu

Publications

1. **M.H Bayana and S.Viriri**, Gender Classification Based On Fusion of Facial Components, *2018 IEEE Conference on Information Communication Technology and Society*, Durban, South Africa.Link https://ieeexplore.ieee.org/document/8368764/

2. **M.Bayana, S.Viriri and R.Angulu** Gender Classification Based on Fusion of Shape and Texture Based approaches(Accepted)

"If we knew what it was we were doing, we wouldn't call it research.-Albert Einstein"

Abstract

Gender classification is very important in facial analysis as it can be used as input into a number of systems such as face recognition. Humans are able to classify gender with great accuracy however passing this ability to machines is a complex task because of many variables such as lighting to mention a few. For the purpose of this research we have approached gender classification as a binary problem, involving the two classes male and female. Two datasets are used in this research which are the FG-NET dataset and Pilots Parliament datasets. Two appearance based feature extractors are used which are the LBP and LDP with the Active Shape model being included by fusing. The classifiers used here are the Support Vector Machine with Radial Basis Function kernel and an Artificial Neural Network with backpropagation. On the FG-NET an average detection of 90.6% against that of 87.5% to that of the PPB. Gender is then detected from the facial components the nose, eyes among others. The forehead recorded the highest accuracy with 92%, followed by the nose with 90%, cheeks with 89.2% and the eyes with 87% and the mouth recorded the lowest accuracy of 75%. As a result feature fusion is then carried out to improve classification accuracies especially that of the mouth and eyes with lowest accuracies. The eyes with an accuracy of 87% is fused with the forehead with 92% and the resulting accuracy is an increase to 93%. The mouth, with the lowest accuracy of 75% is fused with the nose which has an accuracy of 90% and the resulting accuracy is 87%. These results carried out by fusing through addition showed improved results. Fusion is then carried out between Appearance based and shape based features. On the FG-NET dataset using the LBP and LDP an accuracy of 85.33% and 89.53% with the PPB recording 83.13%, 89.3% for LBP and LDP respectively. As expected and shown by previous researchers the LDP clearly obtains higher classification accuracies as it than LBP as it uses gradient rather than pixel intensity. We then fuse the vectors of the LDP, LBP with that of the ASM and carry out dimensionality reduction, then fusion by addition. On the PPB dataset fusion of LDP and ASM records 81.56%, and 94.53% with the FG-NET recording 89.53% respectively.

Acknowledgements

I would like to thank God for providing me with the opportunity to start this research and providing me with a clean bill of health during the process and keep me going even in times when it seemed like i was beating on the same spot for long periods of times for he has given us a sound mind.

I am grateful for the funding provided by my supervisors Professor Serestina Viriri and Dr Raphael Angulu in aiding my research work. It is through their work that I have been able to sit down and really have guidance as I ventured into image processing having not worked in the area at undergraduate and this guidance helped me learn and be able to move forward. It is through this work that I was able to attend the first ever Deep Learning Indaba held at Witwatersrand University and attend the Neural Information Processing Systems Conference (N.I.P.S) conference in California and the Inaugural Black in Artificial Intelligence Conference California where I presented my work again in 2017.

It has been said, "what gives you strength is not what you have but who you have". I am very glad to have had my sister Gugulethu Bayana, though miles away close by through every day and every challenge encouraging me to finish and move on. I would also thank my parents Leonard and Isabel Bayana for their support and always asking when I was going to finish, which was a very hard question. I would thank my friends Bekithemba Bekinkomo Ntini, Lindani Madzokera Moyo, Augustine Hove for their support and insight whenever called upon not forgetting Nozipho and Nobukhosi Ncube for being there at the lowest point in my life as I started the research after a tragedy had struck my life.

Contents

List of Figures

List of Tables

List of Abbreviations

LBP	**L**ocal **B**inary **P**attern
LDP	**L**ocal **D**irectional **P**attern
SVM	**S**upport **V**ector **M**achine
ANN	**A**rtificial **N**eural **N**etwork
DWT	**D**iscrete **W**avelet **T**ransform
DCT	**D**iscrete **C**osine **T**ransform
PPB	**P**eoples **P**arliament **B**enchmark
LTP	**L**ocal **T**enary **P**attern

This research is dedicated to my loving sister Gugulethu Moreblessings Bayana who has selflessly, continously sacrificed everything to support me as i ventured in this long journey and all through my high school and undergraduate, words will always fall short in describing how often you have backed me up like Cr7 for Madrid, you definitely are my Superwoman and not forgetting to thank God for giving you this strength!!!!!!

Chapter 1

General Introduction

1.1 Introduction

The sphere of influence of Artificial Intelligence is growing rapidly and encroaching into every aspect of society. A number of tasks which were traditionally performed by humans are now being carried out by algorithms such as determining recidivism, job placement, bank loan eligibility[1]. A good example of such an algorithm is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) which was rebranded to "Equivant" in 2017 and has been used in the USA since the year 2000, it has been used to predict a defendant's risk of felony in a space of 2 years using an input of 137 features about an individuals prior criminal record. The use of such algorithms has been brought closer in the field of automated facial analysis, which involves a variety of face perception tasks which involve for example face detection [2], [3], face classification [4], [5], [6], [7] and face recognition [8], which is now built into most phones.

Leading companies such as Google and Facebook have released automated software for facial analysis. Facial analysis has also attracted controversial papers for example [9], whose findings showed that Deep Neural Networks are more accurate at predicting sexual orientation from facial images using facebook and social dating site images. In South Africa, the First National bank recently announced the addition of banking technologies to its services, one of particular interest to this research is the ability to open a bank account using a selfie [10], which would require working together with the Home Affairs in verifying images.

The field of Artificial Intelligence has taken advantage of the fact that the human face is riddled with information which can be extracted for use in various fields. Whenever one seeks to initiate a conversation with someone they have never associated with, they perform gender classification before they greet the person they intend to talk to as either a "Mr", "Miss" or "Mrs" to cater for the genders male and female. Hence from this common everyday scenario we see that humans will accurately classify gender and even when there are occlusions. With that being said it is clear that passing this ability to electronic devices is very important as features and services are immediately personalized according to gender of the user, this ability is very important whenever organisation employees deal with customers as they refer to them as either a "Mr" or "Mrs" which gives customers a good feeling of being appreciated. Regardless of environmental hindrances humans are able to differentiate gender in environments with poor illumination, ageing beards and make-up. Studies carried out a few years back showed that humans can easily differentiate between males and females with a 95% accuracy[11], researchers have also found that this accuracy rate reduces to just above chance when considering child faces [12].

Gender classification is a form of pattern recognition ability which humans have from a young age and improves as they develop. To know what gender classification is, it is important to first understand what gender is. It has been described as "a state of being male or female, especially when considering social and cultural differences instead of biological ones " [13]. Social interactions amongst human beings is determined by many factors one of the most important being gender. This task is carried out with great ease by humans with high accuracy levels however transferring this ability to computers is a very complex task. There are various ways which one can use to carry out gender classification such as hand shape, gait, iris, clothing among many others. For the purpose of this research we will consider only two gender classes which are male and female though there has been a great amount of controversy on the number of genders as some researchers have pointed out that gender is non-binary[14].

Andreu et al. [15] states that the face provides sexual information and as a result gender recognition is made fast and efficient. Gender classification plays an important role in many systems for example it acts as a filtering stage for face recognition mechanisms improving speed and accuracy [16]. Gender can be defined as a soft biometric, as it provides some information about an individual which are however insufficient to precisely differentiate between two individuals. Within the machine learning field gender classification is one of the best applications of pattern recognition [16].

To accomplish the task of gender classification a machine learning algorithm will be used. Machine learning has become one of the mainstays of information technology taking over tasks which most people have no idea are being performed by it. Machine Learning was coined in 1959 by Arthur Samuel an American pioneer in the field of computer gaming and Artificial Intelligence. Some researchers have described machine learning as referring to changes in the systems that perform tasks associated with Artificial Intelligence [17]. Machines are required to learn due to a number of reasons some of which are listed below:

- Some tasks cannot be defined well except by example as we may be able to specify input and output combination but not a concise relationship and hence machines are able to do this.
- The need for machine learning also arises when the information about certain tasks may be too large for explicit encoding by humans and hence machines are able to capture more of it than humans.
- Some environments may change over time and hence machines which apply machine learning are able to adapt to changing environments unlike programs which are rigidly written to performs tasks.
- There may be huge chunks of data which humans cannot sift through to determine existing patterns.

A good example of simple machine learning in everyday use is with Facebook's newsfeed before it had the button for one to determine whose status updates one sees first at the top of their timeline. The algorithm used information from likes and page visits to determine what a user likes most and the individual he often interacting with through likes among others to always show their status updates on the other users newsfeed. There are a number of categories under machine learning which include supervised, unsupervised and reinforcement learning.

In this research gender classification is carried out using feature fusion by addition in an attempt to enhance the classification accuracies of features with lower accuracy. Fusion has led to better classification accuracies [18] and hence we have chose this method.

1.2 Motivation

A human can easily determine gender but this is a great task for a computer to perfrom, however the rate of technological development in the field of Artificial Intelligence has led to huge machine dependence on computer vision techniques such as face and gesture detection to name a few. Effective and improved gender detection stands to benefit sectors such as biometric authentication and security systems. A number of feature extractors have been put forward such as the Local Directional Pattern [19], Local Tenary Pattern [20], Local Binary Pattern [21] have all been used widely in research for either gender classification or face recognition [22].

Most previous researchers have been using facial features individually and the face as a whole to classify gender,thus improving gender classification accuracy by including feature fusion. Fusing different feature extractors is the main desired outcome in this research work. Very few researchers have indulged into feature fusion to improve gender classification rates such as Lian and Lu [23] who classified the facial components and also used the hair. Hence they detected facial features using a texture descriptor, then fused them to investigate the effect of combining features on the classification rate. Hence this research is motivated by the need to find out if feature fusion does improve gender classification.

Fusion of features has not been researched as much, as there are some features from the facial components such as the forehead which have a very high classification rate hence fusing them with other features aims to increase the classification rate. We seek to determine first if fusing components will lead to an increase in classification accuracy as even with human detection certain facial components have a greater impact in determining ones gender. Fusion of appearance features and shape features is also carried out with features extractors like the LBP, LDP and Active Shape Model.

Age has been shown to have an impact on gender classification accuracy [24], as their database of over 8000 images with an age range of 0 to 93 years achieved upto 10% higher on adult faces than on young or senior faces. With this knowledge this research goes a step further by then comparing the performance of the LDP against that of the LBP on the FG-NET dataset. This is important as the LBP makes use of raw pixel intensity and the LDP uses gradient and ignoring the central pixel. Hence we seek to answer the question, in case of gender variation which feature extractor has a better performance when it comes to gender classification and this performance is compared on two datasets.

The use of various facial components such as the eyes, nose, mouth, forehead and cheeks can all be used to determine gender and hence this research will show the facial component with the greatest accuracy when it comes to gender classification and fuse it with that with the lowest accuracy.

1.3 Problem statement

The human face is made up of various facial components which can be individually used for gender classification as identified in the literature review or as a whole. Most gender classification methods are able to classify gender using the whole facial area however in a real world environment obtaining the whole facial area may not always be possible. Hence with factors as those which have been described here, it is not always possible to obtain the whole face in order to perform gender classification, hence we seek to classify gender using the face as a whole, facial components and attempt to improve classification with the fusion of facial components. The fusion approach is chosen to offset the lower accuracies of the other feature it is being fused with.

Gender classification is a classical problem and as a result a lot of research effort has been put into the area, the first notable being [25]. A number of data sets have also been put forward for gender classification such as the FERET and the FG-NET dataset. We compare the findings and performance to the feature extractors on two datasets the FG-NET and Pilot Parliaments benchmark [26].

A number of factors have been shown to clearly affect gender classification such as age [24], illumination, head wrapping beard [27], age [28], [29], [30], ethnicity [31], [32], low resolution [33], pose and facial expression [34].

Gender classification techniques have not achieved an ideal zero Mean Absolute Error(MAE) [35] and hence this makes gender classification a viable research area. Commonly used feature extraction techniques are the Local Binary Pattern [21], Active Shape Model [36] and Local Tenary Patterns [20] to mention a few. These texture descriptors have been used in gender classification and have proven to be effective. The Local Binary Pattern is sensitive to noise and hence in cases of variance in illumination it becomes affected in its capability to determining gender. The LDP on the other hand uses a predetermined number of directional responses to encode image gradient and disregards the central pixel and the $8 - k$ directional responses which may be considered as a loss of information. As each pixel contains valuable information about the image the top k responses could be aligned towards a particular orientation hence encoding. The question this research seeks to answer is, does the fusion of the LDP and ASM improve gender classification accuracy?

Also on the point of fusion we also perform feature fusion at component level to determine if classification accuracy is improved by fusing features of various components using the same extractor.

Facial features have different texture qualities and hence they result in unbalanced classification accuracies and hence this research seeks to answer the question as exposure to Ultra Violet rays may accelerate ageing as skin texture is disturbed on components such as the forehead and nose [37], and hence the use of the PPB dataset with individuals of varying complexion is essential. We also seek to determine which facial components, individually has the best accuracy when it comes to classifying gender.

1.4 Dissertation Objectives

- To conduct a critical analysis of the state-of-the-art methods of gender classification from the Literature.
- To detect facial components automatically using Haar Cascades.
- To model a framework for facial component based gender classification.

1.5 Contributions of the Dissertation

This dissertation has made a few contributions with regards to gender classification from facial images. According to the set objective,the following has been achieved:

- Analysis of hindrances to gender classification, image representation and evaluation techniques.
- Research demonstrates that fusing of facial components (nose, eyes, mouth) improve classifications accuracies.
- Fusing of appearance (LBP) and shape features (ASM) was carried out and also showed an improved results from the proposed framework.

1.6 Thesis Overview

This dissertation is organized as follows: **Chapter 2** gives a background in facial gender classification, factors affecting gender classification using the face as a whole or facial components. The chapter also presents a review of related work in gender classification. **Chapter 3** describes the proposed gender classification proposed techniques and materials such as the Local Binary Pattern(LBP), Local Directional Pattern(LDP) and the Active Appearance Model. **Chapter 4** Then presents an analysis and discussion of results and how it relates to the work of previous researchers. **Chapter 5** Includes the conclusion and recommendations for future research work.

Chapter 2

Background Study and Literature Review

2.1 Introduction

Gender is a term which has been commonly used to refer to many aspects accurately or inaccurately about humans. Gender may be defined as attitudes, feelings, and behaviors that a given culture associates with a given sex [38]. Gender classification is a technique of classifying the given images into two classes, males and females. It can be viewed as a binary classification problem, in which one has to predict the given image as belonging to a male or female [7]. Researchers in neuro-psychology have found that the most important means for human communication is the face, which carries identity information such as ethnicity [31], age[39], hence gender is also one of the several characteristics which can be extracted from the face [40]. This chapter will look at developments that have taken place with regards to gender classification improvements and results and the various approaches used in carrying out this classification task.

2.1.1 Applications of Gender Classification

The rise of machine learning and powerful technologies has raised the need for more personalized services, in terms of technological applications hence gender determination is one of the ways in which this can be done. There are a number of applications in which gender classification can be employed as the output of the system or as input into another.

1) Human Computer Interaction
The prevailing, highly digitalized environment has meant that humans interact with computer based systems in many instances. Hence there has been need to find ways in which systems may be able to extract user attributes to respond appropriately, making the system come across as human [41], providing tailored services to users to improve system performance according to gender [32].

2) Input Into Facial Recognition Systems
Gender classification can be used as input into face recognition systems. A good example is searching for an individual in a database after determining the individuals gender reduces the time taken to perform the search assuming both genders have the same number of images in their respective datasets [41]. Other researchers [42]

have found that gender classification can be reduced by employing separate face recognizers trained for the two gender classes.

3) Commercial Development The highly competitive business environment has meant that organisations take their time to effectively advertise to customers using various methods. The use of gender when advertising effectively has been researched for a while and research has shown that there are differences in consumer behaviour between male and female [43]. Hence for marketers for example they may have a camera on the billboard which counts the number of males and females looking at the screen and depending on the majority shows an advert for the majority gender, making it an effective target marketing strategy.

4) Demographic Collection The need for smart ways to collect data from a population has meant that a way had to be found which would employ computer vision for data collection for example to record the number of customers walking into a shop, bar. This data is then used to plan and make decisions by relevant organizations.

5) Mobile Application and Video Games Also related to Human Computer Interaction accurate gender classification improve user experience in mobile applications and games. Various methods have been employed to classify gender such as the use of facial features or gait.

2.2 Face Detection

Face detection is the very first step when it comes to gender classification. It can be regarded as a specific case of object-class detection in which the objective is to find the locations and sizes of all objects in an image that belongs to a certain class. From the work of previous researchers the most popular face detection methods are the appearance based ones, with the Viola Jones [44] coming first. Various techniques have been put forward to review the performance of face detectors with metrics such as learning time, execution time and the number of samples in training [45] being employed. In order for one to extract facial features, one first has to detect the facial area and the various components which is then followed by feature extraction.

The importance of face detection cannot be understated as it determines the input into feature detection and extraction. Face detection is hence one of the most fundamental techniques that enables Human Computer Interaction [46]. Face detection has hence been described as the stepping stone to all facial analysis algorithms, amongst which there is gender and age to mention a few. The goal of face detection is to be able to give the location of a face given an input image, if it is present.

Face detection techniques may be classified into 4 groups [47], which are knowledge-based approaches [48], template matching and appearance based approaches.

1) Knowledge Based Systems
The knowledge Based System uses known information about a face to carry out detection. The human face for example has two eyes, a nose, a mouth roughly at certain common distances from each other, hence we use what we already know about a human face. The rules capture the existing relationships between various facial features and there are therefore the best suited for face localisation [47]. The aim of face localization is to determine the image position of a single face, which considers

a simplified detect problem as the assumption that the input image contains only one face [49].

2) Feature Invariant Approaches
These techniques locate faces using the structural features of the face. The main advantage of this technique over the other techniques is that even with the negative effect of pose, view point or lighting conditions which may affect detection this technique still detects the face easily [47]. After this a statistical classifier is trained and it is used to determine facial and non-facial regions.

3) Template Matching Approaches
In this approach face models are created prior to describe the face as a whole. The correlation between an input image and the stored pattern are computed for face detection [49]. A good example of this approach is the use of deformable templates [50].

4) Appearance Based Method
This approach is different from template methods in the sense that they depend on training face images, which should capture the representative variability of the face to find the face models. The performance of these methods has been proven to be better than other face detection techniques. The Viola-Jones detector has had the best results [51].

2.3 Knowledge-Based methods

The Knowledge-Based methods involves the use of known human knowledge about a face. Usually, they capture the relationships between facial features.

2.3.1 Top Down Methods

In this approach, the face detection methods are developed based on the rules obtained from known human faces for example a face often appears in an image with two eyes which are symmetric to each other. The main disadvantage of this approach is that once a face fails to pass all the rules it is ruled out and yet may still be a face but partially occluded. Yang and Huang [52], used this approach to detect faces, in their system which consisted of three levels of rules where rules at a higher level are general descriptions of what a face looks like and the lower levels rely on facial feature details.

2.3.2 Bottom-Up Feature-Based Methods

This approach seeks to make use of invariant features for face detection as humans can detect faces even when there are different poses. Edge detectors are then used to extract features such as eyebrows, eyes, nose. However Rizvi [53], argued that these feature based algorithms can be easily corruptible for example by illumination.

A number of researchers have used skin colour to detect faces [3], even though skin colour varies most researchers have gone on to show that the major difference lies largely between their intensity than their chrominance [54]. Any of the three color models may be used which are the Red Green Bue(RGB), Hue Saturation Vale(HSV) and YCbCr [3] used a combination of RGB and HSV color spaces to detect faces. Human faces have a texture which is different from other objects which may be in their background. Augusteijn and Skufca [55] developed a method that infers the presence of a face through the identification of texture similar to that of the face [56]. This approach uses standard patterns of a face which are stored with the correlations between an input image and the stored patterns being used for detection. The main weakness of this approach is that it cannot deal with variations in scale and pose [53], which has led to deformable templates for face detection. Sakai et al. [57], used a number of sub-templates for the eyes nose and mouth with each sub-template being defined in terms of line segments. Lines in the input image extracted are based on greatest gradient change after which they are matched against subtemplates, with correlations between sub-images and contour templates being computed first to detect faces.

2.3.3 Appearance-Based

The appearance based method uses models or templates learned from a set of training images which contain variability of facial appearance. This approach makes use of statistical analysis and machine learning to determine relevant characteristics of face and non-face images. Examples of this approach include the Hidden Markov Model, which treats face patterns as a sequence of observation vectors where each vector is a strip of pixels.

2.4 Approaches to Feature Extraction

Feature extraction has been described by researchers as a procedure that is aimed to generate a representative face descriptor by exploiting local operators such as the LBP or processing signals like the Discrete Cosine Transform [58]. The most important goal of feature extraction is cut down on the time of machine training and complexity of space, in order to achieve a dimension reduction. Most researchers [59], [41] are of the notion that gender classification can be categorised into two, the appearance and the geometric approach [29]. However other researchers, Tin et al [60], include two other categories, making them four, which are colour segmentation and template based techniques.

The first reported experiment on gender classification from face images was by Golomb et al [25], who used a multi layer Neural Network approach in which faces where manually aligned using the appearance based approach using 90 images of young adult faces. The backpropagation network termed "SEXNET" had an average error rate of 8.1% compared to the one of humans which is 11.6% these findings set the tone for a better results and techniques.

2.4.1 Appearance (Holistic) Based Approach

The appearance based approach makes use of image pixels to extract features. Hence the approach is based on the calculations which are carried out on the image's pixels [41]. The appearance based approach can be categorised into three classes which are the static, dynamic and apparel features [61]. The static body images may include the face [62] and body [63]. The appearance based approach has a number of advantages firstly being that it does not alter the appearance of face images which can be thought of as naive features and also there is no need for a high level of accuracy as we need a little knowledge of facial features for alignment [64]. The other advantage of the appearance based approach is that it maintains natural geometrical relationships rather than having a vector of facial metric measurements [29]. Appearance based approaches have the disadvantage of being sensitive to many external factors such as position and pose among many others [27]. A good example of this method is the Principal component analysis (PCA) which finds a set the greatest variance in the data.

It has been found that appearance-based methods perform very well in constrained environments but deteriorates in uncontrolled environments [65]. Using the appearance based approach Andreu et al. [15], used an empirical rule put forward by Leonardo da Vinci [66], [67] that states that perfect harmony exists when the face can be broken down vertically into three equal sections whose boundaries match with the eyebrows, hairline and bottom of the nose, division can also be done horizontally into five sections that approximate the width of one eyes. This is illustrated in figure 2.1

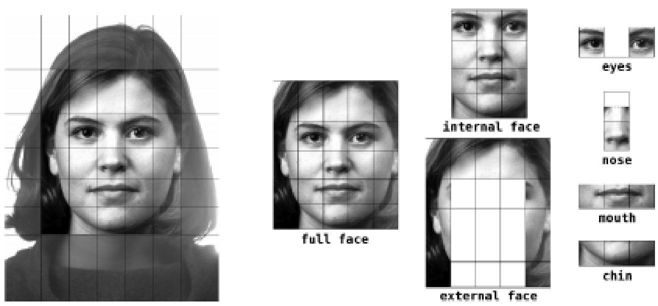


FIGURE 2.1: Facial divisions by Leonardo Da Vinci [66]

Andreu et al. [15] uses this empirical rule to perform sub-image extraction for horizontal division but for the vertical aspect key face features are centered in their corresponding regions. Before feature extraction is carried out, the image is converted to grayscale, other researchers have stated that this reduces computing power required [68]. In the experiments carried out by Andreu et al. [15], an external and internal face is identified. The internal face consists of the eyes, mouth, nose, chin and the external face is made up of hair, ears and contours after which the images are scaled down and pixels are computed by averaging the old ones, using PCA to reduce dimensionality. The work presented by Andreu et al. [15], compared the performance of classifiers, the Support Vector Machine with a polynomial Kernel, the nearest neighbour rules(1,5,10), the Quadratic Bayes Normal classifier and the Parzen classifier. For all four classifiers the full face achieved the highest recognition rates 99.21% for SVM. Also for the SVM the internal face recorded the second highest classification accuracy meaning that it gave a better classification than the eyes, nose, mouth and chin. However for the nearest neighbour with one rule the external face gives an acccuracy that is greater than the internal face as accuracies of 84.22% against 83.28%. In respect to the other facial components individually using the SVM the nose recorded the highest classification accuracy of 86.36%. The findings of Andreu et al. [15], also showed that gender estimation using the global parts of the face was more accurate than when using individual components, also the external parts of the face (hair ear, contour)can also be used to classify gender as there is a correlation between gender and traditional cultural patterns (hair length, earrings).

2.5 Geometric Approach

Unlike the appearance based approach, this approach uses the relationships between the facial features to classify gender. This approach is also known as the non-appearance based approach and it classifies gender through the analysis of physical, biometric or social network-based information [61].

The geometric approach makes use of some known information about a face after which the geometric relationships are extracted. The main advantage of this approach over the other approaches is that it is rotation and translation invariant [61]. An example of the geometric relationships which may be used are the distance of eyebrow from eyes, eyebrows thickness and nose width. It was found that a classification rate of over 95% could be found by [11] from 13 subjects on a set of 180 pictures. This was with only geometric features being used with hair and earrings having been removed from the face. Findings of previous researchers have shown that there are certain features namely the distances of eyebrow from eyes, eyebrows thickness and nose width had the greatest influence in determining gender according to the findings of R.Brunelli and T.Poggio [69] from the sixteen features they had extracted from the face reaching a correct classification of 79%. The geometric features are shown in the figure 2.1. Ashok et al. [70] detected more features than Brunelli [69], 406 geometric features to be precise, performing statistical analysis to show that sexual dimorphism does exist in the human face. The findings of their research showed that a gender classification accuracy of 96% may be obtained through the use of 18-20 geometry features thus supporting the notion that geometry features can be regarded prior knowledge for classification. Figure 2.1(b) shows the ten most significant features Ashok et al. [70], identified using the step-wise discriminant analysis.

2.6 Hybrid Approach

The third approach to gender classification comes up as a result of combining the geometric and appearance based approach to gender classification. In the area of pattern recognition there are two categories when it comes to integrating, one is feature combination and the other is classifier combination [6]. A number of researchers have used the hybrid approach in their experiments [71], FIGURE 2.3 is the architecture of the gender classifier used and is divided into three sections as labelled.

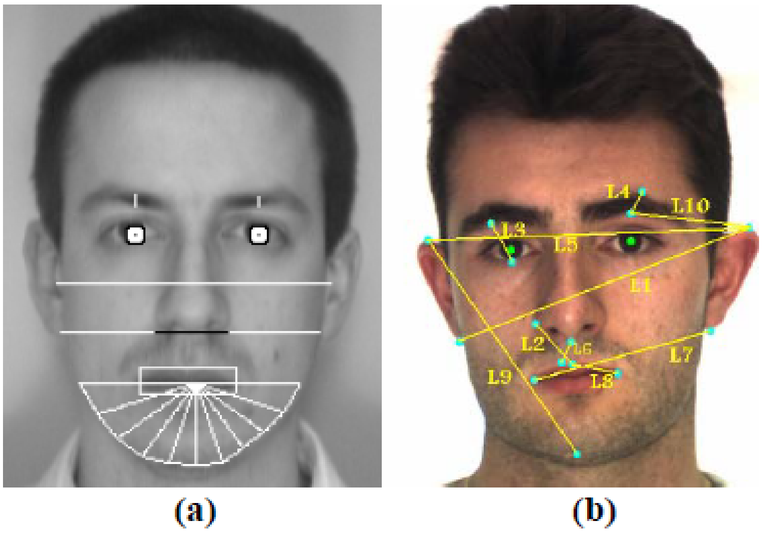


FIGURE 2.2: Examples of geometry features identified by[70]

The first section is one in which the given face image is normalized with gray-level normalization being carried out. In the second stage features are extracted from the normalized image to form a feature vector and this vector is used as input into the final stage of classification.

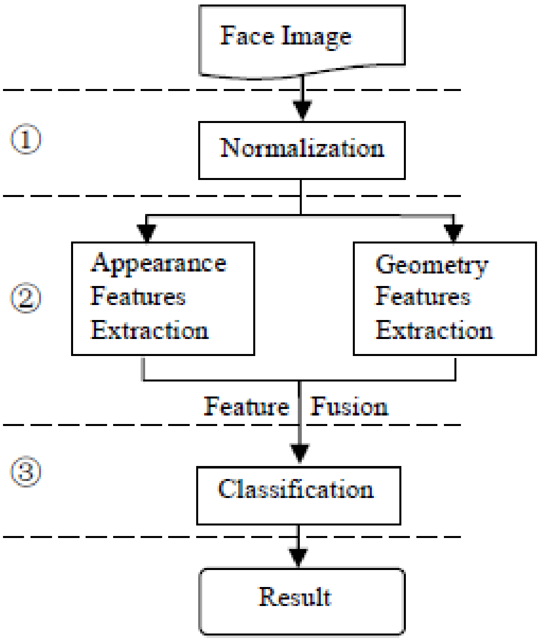


FIGURE 2.3: Architecture of gender classifier by Ziyi et al. [71].

In their approach, Ziyi et al. [71], fused appearance and geometric features using the Adaboost algorithm to choose the best features. The Active Appearance Model was used to extract geometric features, locating 83 landmarks as shown in FIGURE 2.3, from which ten of the most significant features are picked for fusion with appearance based features with normalization being carried out before fusion. The AAM landmark location image is shown below as obtained by Ziyi et al. [71].

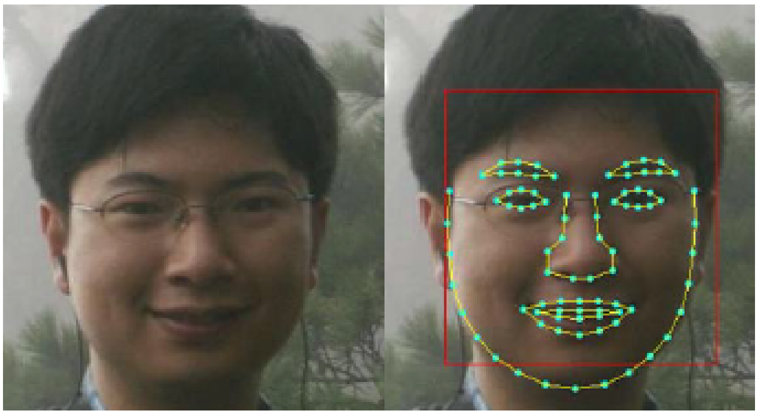


FIGURE 2.4: Landmarks located by AAM

The Adaboost algorithm has been widely used to boost the classification performance of a simple classifier by combining weak classifiers to form a strong one. Haar-like features where used here put forward by Viola and Jones [72] and enhanced by Lienhart [73]. The images used by Ziyi et al. [71], where obtained from the FERET and AR dataset using 1000 pictures, equally divided between male and female using the 5-fold cross-validation. A Support Vector Machine is used with a radial basis function classifier, which outperformed the others like the linear and polynomial kernel. The findings of fusing the appearance and geometric approach is that for both males and females, classification accuracy increased. Initially the appearance based approach records a classification accuracy of 82.97% with 143 weak classifiers and the geometric approach using 10 features records 88.55%, improving to give a 92.38% accuracy.

TABLE 2.1: Gender classification approaches

Recognition rate	**Appearance(143 weak classfiers)**	**Geometry(10 features)**	**Hybrid method**
Total rate	82.97%	88.55%	92.38%
Female	82.91%	89.12%	92.53%
Male	83.74%	87.98%	92.21%

The results by Ziyi et al. [71], showed improved results from the hybrid method (fusing global and local features) as both male and female classification rates where improved by the hybrid method. The hybrid method had a total accuracy of 92.38%, followed by the geometric approach with 88.55%, however there is need for optimization as the AAM locates 83 points but uses a few of them to form the geometry vector. The findings of this paper showed that this approach was effective and robust in terms of expression and illumination. However a shortcoming is identified as the Active Appearance Model locates 83 points on each face with only 10 being used hence time is consumed which can be reduced with further optimization.

Mozaffari et al. [27], put forward results after combining two appearance based features and a geometric feature, the Discrete Cosine Transform (DCT),the Local Binary Pattern (LBP) and geometric distance feature (GDF). In their paper, two different databases are used, the AR consisting of 56 female and 70 male and the ethnic dataset, selecting only frontal facial images with no expression. The ethnic dataset built here had three main differences being that it had females wearing a scarf and more men have facial hair(beard). In their experiments Saeed et al. [27], compared the results of combining two appearance based approaches (LBP and DCT) and combining all three (LBP ,DCT and GDF). Combining the LBP and DCT on the Ethnic and AR dataset yields classification accuracies of 84.96% and 80.3%. Combining the the approaches (LBP, DCT, GDF) led to higher classification accuracies of 97.1% and 96.0% and enhanced classification accuracies on both data sets, the Ethnic by 12.5% and AR by 15.7%. Saeed et al. [27] state that the reason for this is that the appearance and geometric based features are very different and hence combining them leads to very high levels of accuracy.

Chen and Lu [6], further combined facial, hair information and having noted that most existing approaches utilize only the internal facial information and carried out a psychological experiment. Sixteen individuals were to determine gender of 200 images which are divided into two groups. Group A was fed complete faces and Group B was given only inner faces. The accuracy of group A was found to be 100%, while the average accuracy of Group B is 92.5%, hence from the face we conclude that hair can provide clues to help differentiate between male and female. This is true in cases where the face is neutral such as group A, which includes hair and went on to obtain a higher classification accuracy.

In their paper, Chen and Lu [6], use the eyes, nose and mouth regions with the hair being represented by a fragment-based encoding. A fuzzy integration based classifier for combining the 4 classifiers based on Support Vector Machines with probabilistic output is used for fusing the four different classifiers used on the nose, mouth, eyes and hair. The hair is the hardest to extract features from and hence the fragment based method [74], by Laperidza is used, which is insensitive to changes in illumination. The proposed system is shown in Figure 2.5.

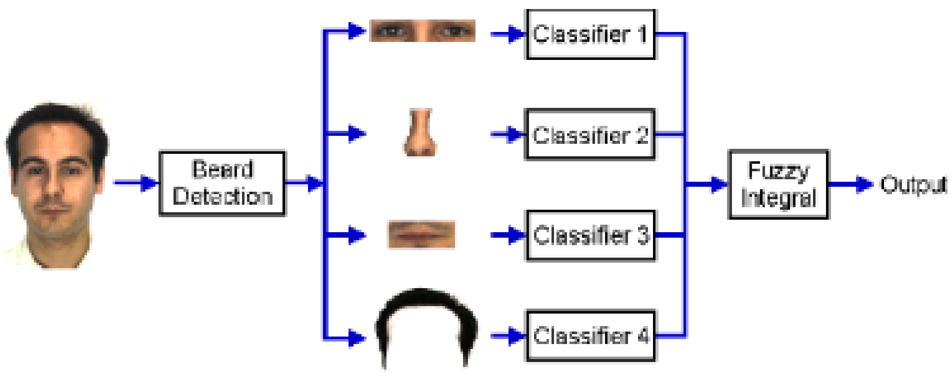


FIGURE 2.5: Proposed gender classification system for combining four different SVM's[6].

From the paper by Chen et al. [6] obtained results which showed that the eyes had the greatest accuracy of 84.25% followed by the mouth with 78.95%, then the hair 78,95% and lastly the nose with 70.94%. Of particular importance is the fact that the hair in this case obtained a higher classification accuracy than the nose, further proving the point that the fragment based method put forward by Lapedriza et al. [74] produces high classification rates when using the hair [74]. They went on to further compare the different combination methods for all four classifiers which are the fuzzy integral with 90.61%, weighted sum 88.73%, product 87.63%, CCA with 77.76% .

2.7 Feature Extraction techniques

The human brain can discriminate many patterns, expressing it using language is challenging as there are a few terms for verbalizing this Xu and Lu [71], who also state that feature extraction may be improved when prior knowledge (geometric approach) about images is used. The two main approaches have been well spelled out, this section goes into depth discussing the various specific examples of the feature extraction techniques.

2.7.1 Discrete Cosine Transform

A Discrete Cosine transform (DCT) is a Fourier related transform similar to Discrete Fourier Transform (DFT) but uses only real numbers [75], as it transforms a signal from a spatial domain to a frequency domain. When extracting features using the the DCT there are two main steps which are applying the DCT on a given image and the second being the selection of coefficients. The second section being subdivided into two groups, the which are static and data dependant. The DCT is computationally efficient and can be implemented using a simple filter convolution [75]. Equation 2.1 is an expression showing how the most significant DCT coefficients can represent and classify an image. DCT transforms a 2 Dimensional(2D) image, f(i,j), with N columns and M rows and defined as :

$$\begin{aligned} F(U,V) = {} & \left(\frac{2}{N}\right)^{1/2} \left(\frac{2}{N}\right)^{1/2} \sum_{N-1}^{i=0} \sum_{M-1}^{J=0} A(i)A(j) \\ & \times cos. \left[\frac{\pi.\mu}{2.N}(2i+1)\right] .cos. \left[\frac{\pi.v}{2.M}(2j+1)\right] \end{aligned} \tag{2.1}$$

In equation 2.1, computes the U and v^{th} entry of DCT of an image and N being the size of te block that DCT is done one for example an 8 *by* 8 image.

2.7.2 Local Binary Pattern

The Local Binary Pattern(LBP) is one the simplest texture based feature extraction techniques. This technique makes use of pixel values performing mathematical operations on them. The Local binary pattern is an image operator which was first used as a measure for local image contrast [21]. The LBP is a non-parametric approach which summarizes local structures by comparing each pixel with its neighboring pixels [63]. The LBP has been used in many other fields as used shown by Mdakane et al. [76], who used both the original LBP, an extension of the LBP.

The initial Local Binary Pattern proposed by Ojala [21], labelled the pixels of an image with decimal numbers. A 3*3 window is passed over the whole image excluding the outer pixels as they do not have any neighbours and hence they are not used. The centre pixel is thresholded against its neighbours and if the neighbour is greater than or equal to the centre pixel it is replaced with a one otherwise a zero. A binary number is obtained by concatenating the binary codes from the top left to the bottom right and that resulting decimal value. The LBP with its radius can be represented as LBP_P^r as shown in the Figure 2.6.

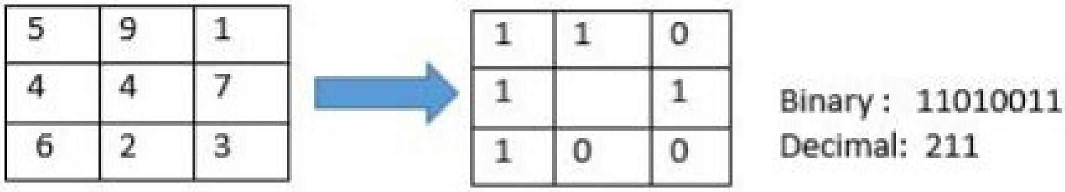


FIGURE 2.6: LBP operator operation

A variant of LBP was put forwardby Lian et al. [23], called the multi-resolution LBP which had the main advantage that it could capture both fine and coarse local micro-patterns. This multi-resolution LBP was based on a simple and powerful concept [77]. This approach considers a hierarchical image description at multiple resolution levels, since the aim is to retain the fine-to-coarse local micro-patterns and spatial information. The findings of the research by Lian et al. [23], were that the LBP and MLBP generalized better than the Gabor with the MLBP recording a higher correct classification rate of 96.5% and a higher correct classification rate of 96.56% and a higher cross validated accuracy of 95.78% being obtained using the Support Vector Machine as the classifier.

Another variant of LBP was put forward by Abbas et al. [78], which incorporated four predefined patterns testing it on the FERET dataset. As a result of combining the two predefined patterns more information would be available to create discriminating features. This approach divided the face into 3 by 3 blocks with four histograms made for each pattern and computed on a different code image with the final feature vector for the input image being a result of concatenating the four block vectors. The SVM was the classifier of choice for this proposed feature extractor along with the KNN and Adaboost. The SVM recorded the highest classification accuracy of of 96.2%, KNN 86.33%, Adaboost 84,59% and decision tree 79.88%.

Due to many limitations associated with the original LBP many extensions have been put forward, a variant was put forward by Yang and Wang [79], by using the low density features and tackling the curse of dimensionality by fusing PCA of the LBP features. The proposed variant is different from the traditional LBP with 8 neighbours. Here the neighbours are split into two operators, the cross neighbours and the diagonal neighbours after which the two low density operators are fused with PCA being used for dimensionality reduction. The SVM classifier with an RBF kernel is used here.

2.7.3 Local Ternary Pattern

The various shortcomings of the Local Binary Pattern such as its sensitivity to noise led to the extension of the LBP by [20], to come up with the Local Ternary Pattern. They extended the basic LBP to a version of three-value codes, which is called the Local Ternary Pattern(LTP). In the LTP, the indicator (x) is further defined as

$$LTP_{(P,R,\tau)} = \sum_{i=0}^{P-1} s(p_i - p_c \times 3^i)$$
$$s(x) = \begin{cases} 1x \geq \tau \\ 0|x| < \tau \\ -1x \leq -\tau \end{cases} \tag{2.2}$$

In equation 2.2, τ is a threshold which is determined by the user and X_c represents the value of the central pixel and x_i for i=0,1,2,3,4,5...7 are the neighbouring pixels of x_c. However selecting the value for τ is a challenge . The main problem that arises is that selecting τ to be an optimum value for every image is no easy task. Tan and Triggs [20], also presented a coding scheme which is represented by splitting each ternary pattern into two parts: which is the positive and negative parts and two histograms are generated for each. These two histograms are then concatenated and used as a feature descriptors.

85	32	71
53	50	47
64	38	15

(a) Original image

1	-1	1
0	x	0
1	-1	-1

(b) LTP code $\tau \pm 5$

0	1	0
0	x	0
0	1	1

(c) Negative LTP code $\tau \pm 5$

1	0	1
0	x	0
1	0	0

(d) Positive LTP code $\tau \pm 5$

FIGURE 2.7: LTP code with $\tau = \pm 5$ and corresponding positive and negative codes

Figure 2.7 represents the LTP code for a 3×3 sample region. The LTP can then further be combined with the LDP to make it robust to noise illumination by comparing the raw pixel values. The threshold τ could also be made dynamic by deriving from the weights of neighbour pixels instead of setting it to a static value that is difficult to optimize for all images.

2.7.4 Local Directional Pattern

Building on the inadequacies of the Local Binary Pattern (LBP), the Local Directional Pattern (LDP) was put forward with some advantages over the LBP. The LDP has proven to be robust under face recognition [19]. The LDP, unlike the LBP makes use of edge response values in different directions instead of pixel intensities using a Kirch mask in 8 different positions. The Kirsch is a first order edge detector that gets image gradients by convolving 3×3 image regions with a set of masks.

$$
\begin{aligned}
P(x,y) = max \left\{1, \max_{k=0}^{7}[|5S - K - 3T_k|]\right. \\
where \\
S_K = P_K + P_{(K+1)} + P_{(K+2)} \\
and \\
T_K = P_{(K+3)} + P_{(K+4)} + P_{(K+5)} + P_{(K+6)} + P_{(K+7)}
\end{aligned}
\tag{2.3}
$$

In the equation 2.3 $P(x,y)$ is the Kirsch gradient, a in K_a is evaluated as $a = a$ mod 8 and $P_K[K = 0,1,2,3....,7]$ are the eight neighbouring pixels of $P(x,y)$ as will be shown in FIGURE 2.8.

The LDP features had an accuracy classification of 95.5%, with the SVM and LBP coming slightly below 92.25% using the FERET dataset and from the results LDP had a slightly improved classification accuracy.

The LDP, as proposed by Jabid et al. [19], is an eight bit binary code assigned to each pixel on an input image, which is calculated by comparing the edge responses. In their paper, Jabid et al. [19], uses the Kirsch masks in eight different directionsM_0 to M_7 as represented by Figure 2.8:

$$
\begin{array}{cccc}
\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} &
\begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} &
\begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} &
\begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \\
\text{East } (M_0) & \text{North East } (M_1) & \text{North } (M_2) & \text{North West } (M_3) \\
\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} &
\begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} &
\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} &
\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix} \\
\text{West } (M_4) & \text{South West } (M_5) & \text{South } (M_6) & \text{South East } (M_7)
\end{array}
$$

FIGURE 2.8: Kirsch edge response masks in eight directions

Particular importance here is placed on the K most prominent directions so that we may generate the LDP and as a result after finding the top K values $|M_j|$ and set them to 1 and the rest we set to 0. A histogram has been widely used to analyse and characterize images [80], and has also been used with image descriptors. To find the number of bins when using k=3 after using a Kirsch mask from 8 different directions we use the combination rule since we choose the k most significant edge responses 8C3 to get 56 histograms shown by the equation below where $I_L$represents the encoded image of the Histogram(H) as shown in Figure 2.9.

$$
H_i = \sum_{x,y} P(I_L(x,y) \tag{2.4}
$$

The face is divided the face into small regions and for each small region we extract histograms from each region. It is important to note that this has the benefit of maintaining spatial information which would have been lost if one whole histogram was used. The maintenance of this spatiality is shown in Figure 2.9.

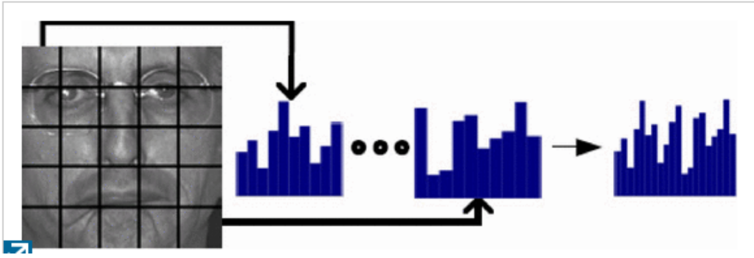


FIGURE 2.9: Spatially enhanced histogram [21]

2.8 Normalizing and Standardization

After features have been extracted geometrically or using the appearance based approach there is need to carry out standardization of the two vectors if fusion is to be carried out. Normalization and standardization have been used interchangeably. Various methods may be used to carry out standardization such as Minimum-Maximum(min-max) or Z-score. Data normalization on one hand performs transformation on all variables (in case of appearance, pixels into one specific range for example) as they range from 0-255 and the gaps may be very wide. A number of normalization methods like min-max(MM), Z-score and adaptive methods have been used obtaining differing results. The MM is perhaps the simplest of the normalization methods and is the preferred model in the work of Ziyi et al. [71].

$$n = \frac{x - min}{max - min} \tag{2.5}$$

In equation 2.5 the quantities max and min specify the end points of the range(maximum and minimum). The normalization method maps the raw scores to the [-1,1] range and n in this case would be referring to normalized value before fusion of the two features.

2.9 Dimensionality Reduction

The fact that the facial image data is always high dimensional has meant that it requires considerable computing time for classification [81]. Hence gender classification like other image recognition practices has the stumbling block of high dimensionality [82], hence dimensionality reduction improves the learning process irrelevant and redundant data are removed. Data dimensionality reduction has been described as the process of deriving a set of degrees of freedom which may be adjusted to reproduce most of the variability as seen in the training set [83]. The problem of high dimensional data is often referred to by most researchers as the "curse of dimensionality" [84]. In face recognition and gender classification the needed for dimensionality reduction arises as images contain a very large number of pixel values and direct operations on them hence leads to high computational and storage demands and hence to avoid this we carry out dimensionality reduction. Dimensionality reduction techniques can be categorised into two classes the linear methods which are PCA, LDA, LPP and non-linear dimensionality reduction techniques include ISOMAP & Eigenmaps. Equation 2.6 shows how dimensionality reduction techniques work.

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \textit{dimensionality reduction} -> \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \tag{2.6}$$

Equation 2.6 represents dimensionality reduction, after undergoing dimensionality reduction the number of variables is reduced from N to M as $M \leq n$ from a high dimensional data space [N] to a lower dimensional dataspace [M].

2.9.1 Principle Components Analysis (PCA)

The PCA is perhaps the most common linear dimensionality reduction technique. It performs an orthogonal transformation to convert a set of correlated variables into a set of linearly uncorrelated variables, PCA has widely been used for dimensionality reduction in the field of face recognition. This dimensionality technique is also known as the Krhunen-Loeve expansion. Sirivic and Kirby [85], where the first to use PCA to efficiently represent pictures of human faces arguing that face images can be reconstructed approximately as a weighted sum of a small collection of images that define a facial basis (eigen image) and the face's mean image. The Eigenface is the most common face application of PCA in the field of face recognition. However the main limitation of PCA is that because it is linear it cannot capture nonlinear data and it is preferred due to its computational and analytical simplicity [86]. Hence the PCA throws away many minor components, retaining fewer principal components on a low dimensional space also known as an eignen-face in face recognition. This is shown in equation 2.7.

$$y_{ik} = V_K^T X_i \tag{2.7}$$

In the equation 2.7, $X_i, i = 1, 2, ..., N$, is the i-th n-dimensional sample (image), where V_k with $k = 1, 2, ...d, d \leq n$ is the k-th largest eigenvalue of the covariance matrix of the data set and y_{ik} is the k-th component of the projected data in the d-dimensional subspace, classification is then carried out on the reduced space after linear projection is carried out.

2.10 Linear Discriminant Analysis

The use of the Principle Component Analysis leads to the loss of discriminative information as classes are not considered and so the collected results are smeared together making classification more difficult [87]. Hence to overcome this problem the LDA is used as PCA only gives the main variance of data and yet with classification but as with most classification problems you need more information from the data to help with classification.

The main aim of LDA is to find optimal projection, vectors simultaneously minimizing the within-class distance and maximizing the between-class distance for the given data [88], as shown in Figure 2.10:

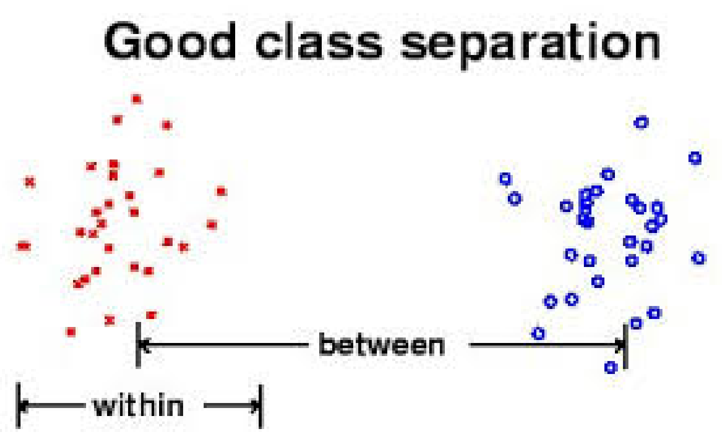


FIGURE 2.10: Within and between class distances

Hence with this in mind the PCA and LDA may be used together in classification problems, with the drawback being that when PCA is used first in cases in which there are few samples it may result in the loss of valuable information [89]. Hence LDA comes in handy when it comes to classification and description, and is hence a supervised learning algorithm which uses class labels. So LDA obtains a bunch of features which all divide in a finest way among classes and increasing proportion among classes towards inside sub-classes.

2.10.1 Independent Component Analysis

Independent Component Analysis (ICA) is a statistical method, which seeks to transform multivariate data into a linear sum of non-orthogonal basis vectors with coefficients which are statistically independent with the main idea also being changing the space from an m-dimensional space to an n-dimensional space. After observing multidimensional data it is expressed as a matrix X with N rows and T columns, where N is the number of variables and T represents the observation for each variable. The goal of ICA is shown in the equation 2.8.

$$Y = BX \tag{2.8}$$

In the equation 2.8 the equation ICA calculates the square matrix B that linearly transforms X into a matrix with variables which are independent. Different variants of ICA have been put forward such as FastICA [90], which is a highly efficient and computable method which perform ICA, one of its main characteristics is that it has high capability of fast convergences leading to faster estimation and refinement.

2.10.2 Locality Preserving Projections

The Locality Preserving Projections(LPP) is a subspace learning algorithm derived from the Laplacian Eigenmap, the main difference between the LPP and the other techniques discussed is that they see only the Euclidean structure of face space, LPP finds an embedding that preserves local information and obtains a face subspace that best detects the essential face manifold structure[91], these Laplacian faces are the optimal linear approximations to the eigen functions of the Laplace Beltrami operator. The main benefit of this is that the unwanted variations in lighting and facial expression to name a few are reduced. Hence according to Shylaja et al. [91], the Eigen face method seeks to preserve the global structure of the image space, whilst the Fisher faces method aims to preserve the discriminating information LPP preserves the local structure of the image space.

The objective function of the LPP is in equation 2.9, as it aims to preserve the intrinsic geometry of the data and local structure.

$$\begin{aligned} Objective\ \ function =& min\sum_{ij}(y_i - y_j)^2 S_{ij} \\ where S_{ij} =& \begin{cases} exp(|x_i - x_j|^2/t), |x_i - x_j|^2 \leq \epsilon \\ 0 \end{cases} \end{aligned} \tag{2.9}$$

In equation 2.9,y_i and y_j, refers to two edge which may be drawn between the two nodes i and j.

2.10.3 Gray-Level Co-Occurence Matrix

This is a simple approach that describes texture by using statistical moments of the intensity histogram of an image [92]. This has the benefit of providing information about the relative position of the neighbouring pixels in an image. GLCM is a tabulation of how often different combinations of pixel values appear in an image and can also be thought of as a matrix in which one of the rows and columns is equal to the number of gray-levels. This approach describes an image texture by comparing each pixel with its neighboring pixel at a specified distance and orientation. This technique extracts second order statistical texture features from gray scale images. The GLCM is hence a matrix whose rows and columns is equal to the number of quantized gray-levels, N_g. The entry $p(i, j)$ is the second order statistical probability for changes between gray-level values i and j at a particular distance and orientation θ.

Suppose there is an $N \times N$ image I(i,j), with N_2 columns and N_Y rows. N_g is quantization of gray-level appearing at each pixel in the image. Let the rows of the image be $N_y = (1, 2.., N_y)$, the columns be $N_x = (1, 2...N_x)$ and the set of N_g quantized grey levels be $G_x = (1, 2, 3...N_(g - 1))$. The image can be represented as a function that assigns some grey level in G to each pixel or pair of coordinates in $L_y \times L_x$;G $\leftarrow L_y \times L_x$. Texture information is specified by the GLCM matrix of relative frequencies $C(i, j)$, hence the value at GLCM(i,j) represents the number of occurrences of the gray-level value i at the reference pixel and gray-level value j at a neighbor pixel, a certain distance d and orientation θ^0. The probability measure can be defined as :

$$P(d, \theta) = p(i, j) \tag{2.10}$$

In the equation 2.10 $p(i, j)$ is defined as:

$$p(i,j) = \frac{GLCM(i,j)}{\sum_{i=0}^{N}\sum_{J=0}^{N} GLCM(i,j)} \tag{2.11}$$

The sum in the denominator in the equation 2.11 represents the total number of gray-level pairs (i,j) within the image and is bounded by $N_g \times N_g$. Dividing every pixel in the GLCM matrix with the denominator results into a normalized GLCM matrix.

2.11 Common Datasets

To obtain results which are comparable with that of other researchers, standardized datasets must be used to accurately evaluate and then compare the various facial algorithms. This section will describe the most popular datasets.

2.11.1 Face and Gesture Recognition Network (FG-NET)

This ageing dataset is made up of 1002 images which belong to 82 different individuals with ages ranging from 0 to 69, meaning that each individual has more than one image. With more than 700 images ranging from 0 to 20 the age distribution is not equally distributed and this is one of the main reasons why the FGNET is a challenging dataset also given that there is a mixture of gray and color images. The FGNET-AD consists of 48 males and 34 females each with 571 and 431 images a set [93]. This dataset has been used by a number of researchers in their papers such as Angulu et al [28].

2.11.2 Face Recognition Technology(FERET)

This dataset contains 1564 sets of images for a total of 14,126 images that includes 1199 individuals and 365 duplicate sets of images, the database was collected as part of a sponsored research. The images in this dataset were collected in 15 sessions from August 1992 to July 1995. Before FERET was used high recognition rates were being claimed which dropped with the use of the FERET. A number of researchers have used this dataset such as [5], used it for gender recognition.

2.11.3 Database of Faces (ORL Database of Faces)

This dataset contains a set of images collected from the period April 1992 to April 1994. It is made up of ten different images of 40 subjects, taken at different times, with varying lighting. All the images here were taken against a dark homogeneous background with subjects in an upright, frontal position.

2.11.4 Labelled Faces In the Wild

Most dataset environments are constrained, not mimicking real life environments and hence it is easy to carry out detection, extraction and classification. By unconstrained it means that the faces show a large range of variations in terms of lighting, expression, age and ethnicity to name a few. The database contains 13, 233 target face images with certain images containing more than one face and most images are in color and a smaller fraction being in gray-scale [94]. A number of researchers have used this dataset such as Han et al [95].

2.11.5 AR Dataset

Most datasets as described above have been restricted to the extent that even head gear, for religious matters is not included however the AR includes females wearing the hijab. This database has been used by a number of researchers such Mozaffarri et al [27], which affected results by reducing classification accuracy as reported in their research.

2.12 Performance Evaluation Protocols

With various algorithms having been used on different standardized datasets assessing the performance of these algorithms is at the core of machine learning [96], other researchers such as Bousquet [97], have stated that a fundamental issue in the design of efficient machine learning systems is the estimation of the accuracy of learning algorithms. In this section we discuss a few of the estimation algorithms. A good evaluation strategy should be independent of training data and representative of the population from which it has been drawn from Budka et al. [98].

2.12.1 Cross Validation

Cross validation has been defined as a statistical method of evaluating algorithm performance by dividing data into segments one for training and the other for evaluation to avoid overfitting. Cross validation seeks to measure the generalizability of an algorithm and compare the performance and find the best algorithm for the available data [99]. Compared to the re-substitution error, CV avoids overfitting by splitting data into a number of parts.

2.12.2 Re-substitution Validation

In this approach, the model is learned from all the available data and then tested on the same data. The error rate is evaluated based on outcome against actual value from the same training dataset, and this error is referred to as the re-substitution error. The main draw back of this approach is that it suffers seriously from overfitting which means that it fails to meet one of the objectives mentioned by Refaeilzadeh [99], as it performs poorly on unseen data.

2.12.3 Hold-Out Validation

To avoid over-fitting, an independent test set is used. This approach enables data to be split into two non-overlapping parts, some for training and the other for testing giving a more generalized performance of the algorithm. However it has been stated that it does not use all the available data and hence the obtained results depend more on the choice for training as stated by Refaeilzadeh et al. [99], instances which may include in the test may be too easy or too difficult to classify leading to skewed results. However these problems may be addressed by repeating hold-out validation a couple of times and it needs to be systematic to ensure that all data is used in training and testing. To address this the k-fold cross validation is used.

2.12.4 K-Fold Cross-Validation

In this approach the data is first divided into k equally sized divisions. K iterations of training and validation are performed such that within each iteration a different fold of the data is held out for validation with the remaining K-1 being used for learning. A good example is with the binary classification problem such as gender, data should be arranged in a way that in every fold each comprises around half the instances. The standard deviation of these errors can be used to approximate the confidence range of the estimate. The main advantage of this validation technique is that eventually all samples will be used for both learning and validating the model. A stratified cross validation approach is used in order to improve accuracy of the estimation.

2.12.5 Leave One-Out Cross-Validation

The Leave One Out(LOO) Cross-Validation [100], may be thought of as a different case of k-fold cross validation as when given a dataset with C classes, $C - 1$ validation experiments are performed. For each experiment, data from $C - 1$ classes is used for training and data from one class that was left out is used for validation. This protocol is commonly referred to as the deleted estimate or U-method. This protocol has a number of properties which are similar to another technique called the Jack-knife [97], as they have many similar characteristics. One record here is used for training and another record is used for testing. The accuracy estimate obtained using the LOOCV is known to be unbiased but has a high variance resulting in unreliable estimates [101].

TABLE 2.2: Validation Methods

Re-substitution Validation	**Advantages**	**Disadvantages**
Re-substitution Validation	Simple	Leads to overfitting
Hold-out Validation	Training &test as data is split	Overlapping training data
LOOCV	Accurate estimation	Large variance is observed

2.13 Performance Measure Criteria

The performance of any proposed model must be measured using various set metric, as discussed in this section.

2.13.1 Confusion Matrix

The confusion matrix is one the metrics most commonly used for finding correctness and accuracy of a model and can be used in cases where there is a classification problem with two or more classes. The confusion matrix is the base from which most or all the performance metrics are based and is illustrated in the diagram below:

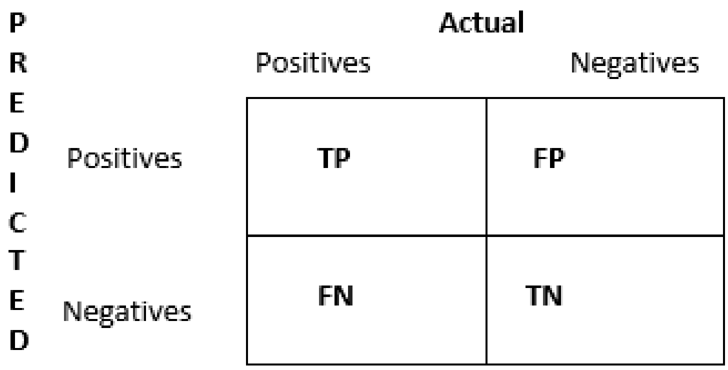


FIGURE 2.11: Confusion matrix

The metrics in TABLE 2.2 are :

- **True Positives (TP):-** Refers to a case or cases in which the actual class of the data point was True and the Prediction is also true for example when an individual is a female and is classified as a female.
- **True Negative (TN):-** Refers to cases when the actual class of the data point was false and the predicted is also false.
- **False Positives (TP):-** Refers to the cases when the actual class of the data point is false and the predicted is true.
- **False Negatives (FN):-** Refers to cases when the actual class on the datapoint is true and it is predicted as false.

2.13.2 Accuracy

Accuracy may then be used to determine the number of correct predictions made by the model over all kind of predictions and is recommended for use when the classes are nearly balanced.

$$Accuracy = \frac{TP + TN}{TP + FN + TN} \tag{2.12}$$

Ngan et al [102], states that accuracy is defined as the number of correctly classified male images, TM, divided by the total number of male images,M as shown below:

$$\text{Male accuracy} = \frac{TM}{M} \tag{2.13}$$

Female accuracy may also be thought of as the number of classified female images, TF, divided by the total number of female images :

$$\text{Female accuracy} = \frac{TF}{F} \tag{2.14}$$

The overall accuracy is then expressed as the sum of the correctly classified male and female images divided by the total number of images as shown below:

$$\text{Overall accuracy} = \frac{TM + TF}{M + F} \tag{2.15}$$

2.13.3 Precision

Precision refers to the proportion of patients classified as being female for example are actually female as shown in the equation below.

$$Precision = \frac{TP}{TP + FP} \tag{2.16}$$

In equation 2.16 we get the precision by dividing the True Positives by the sum of the True Positive and false positive.

2.13.4 Receiver Operating Characteristics Curve

One of the various ways of measuring the performance of a classifier is through the Receiver Operating Characteristics Curve (ROC), which is a characteristics curve of a classifier for a particular problem [103], as it summarizes how well the classifier has performed for that problem at different thresholds hence allowing for the illustration of each classifier between its true positive rate (which refers to the number of correct positives divided by the total number of positive cases) and its false rate (the number of incorrect positive) and its false rate (the number of incorrect positives cases divided by the total number of negative cases) [104]. The selection of the operating threshold is then application specific, depending on the maximum acceptance of false and true positives.

2.14 Previous Work

Age estimation and gender classification of facial images [105] based on Local Directional Pattern has been carried out using different block sizes find the one with the best classification being the 6 by 6 block, hence its uses in the actual experiment. The results obtained the LDP yielded a 95%, PCA 85% and LBP 93% for gender classification.

The Principal component analysis may be used for feature extraction and dimensionality reduction [106]. Since gender classification is a pattern recognition problem, the two main questions which are how to extract and classify subjects. The extraction of features reduces the time taken for classification as there is always a lot of redundant information, Vyas et al [107], compared PCA and LDA techniques for face recognition based feature extraction for various facial expression such as happy, sad, sleepy and surprised. The results of this experiment show that being happy had an accuracy of 93% for the PCA and 86.6% for the LDA. The sad facial expression with PCA gives 60 percent recognition rate and LDA provides 68.75 percent.

Males and females have different facial landmarks and these differences are spread across all ethnic groups and are used for classification [108]. The scale invariant feature transform [65], extracted features which are invariant to any affine transformation, in which angles, length and shape are not preserved. The findings by David et al [65], the SVM with an RBF kernel had accuracy of 86.84% .

The Active Appearance Model, a statistical approach model put forward by Cootes et al [36], is used amongst the other geometrical approaches. A comparison of Active Appearance Model, Gabor and Wavelet Decomposition is carried out by [109] for gender, age and ethnicity. Two databases where used in this experiment which are the FG-NET and Cohn-Canade. The findings of the research showed that the Gabor wavelet had the highest accuracy classification of 91.30% and the MM had the highest of 92.53 on the FG-NET. This showed that the AAM performs well when the input images into a system are of varying ages.

Gabor Wavelets are another alternative when it comes to geometric feature extraction [110], as they exhibit two important characteristics which are spatial locality and orientation selectivity. Other researchers [111], used 4 geometrics which are the midpoint of right eye to left-eye, lips to nose, nose to eyes and lips to eyes attaining a gender accuracy of 95.6%.

A number of biometrics have also been used to classify gender and fingerprints have been used [112]. Previous researchers have shown that males have higher ridge count than females [113].

The Discrete Wavelet Transform takes the approach of analysing at various different frequency bands with different resolutions , this is achieved by decomposing the signals into a different "coarse" approximation and with detailed information, the research employed both the radon transform to project the given input image into radon space then using DWT to extract features for emotion recognition. The system recorded a recognition accuracy of 91.3% on the JAFFE dataset.

Half face images have also been used for gender classification [114]. The research makes use of both the DWT, to get an approximation of the image after which MMDA is then applied to select the most discriminant feature subset. DWT hence helps in reducing face size whilst maintaining all the information. The results of the research showed that after training with a dataset containing half-right face images yielded a 96.8% against a 95.5% after using half the right face for training and left side for testing.Also more importantly this research also used full face images as input with the KNN and SVM used as the classifiers of choice getting an accuracy of 88,8% and 98.4% respectively. This research had great contributions to the field as first of all it considered half-faces which is very common in uncontrolled environments as a result of occlusion and secondly he tested two different classifiers and used two databases making his research valid.

The two dimensional Wavelet Transform has also been used in image classification such as classifying weed as narrow or broad [115], after which, further processing is carried out as coefficient values are still of the the same size as that of the input image as a result and as a result the line measure technique is used to extract the image feature vectors. From the results of the LMT with continuity 7 and angle of 45 obtained the best results classification rate of 86% and 88.4% for the narrow and broad weed response.

Another texture based feature extraction method is the GLCM used to extract statistical second order statistical texture features, of which the third and higher order represent relationships amongst three pixels[116]. The findings of their paper dealing with second order statistical features where that by converting RGB to gray-level image, compression time is reduced using GLCM however this is not the case with DWT techniques.

Jabid et al. [19], showed that the LDP possesses more stability when noise is introduced into an image. This is shown in the diagram below as with the noise addition to the image, the value of the 5th LBP bit changed from 1 to 0, and hence it moved from being uniform (at most two bitwise transitions) to non-uniform code however the LDP's pattern value remains constant as gradients are more stable than gray value. Since the LDP computes a 8-bit binary code for each pixel in the image by comparing edge responses of each pixel in different orientations instead of comparing raw pixel intensities as done by the LBP.

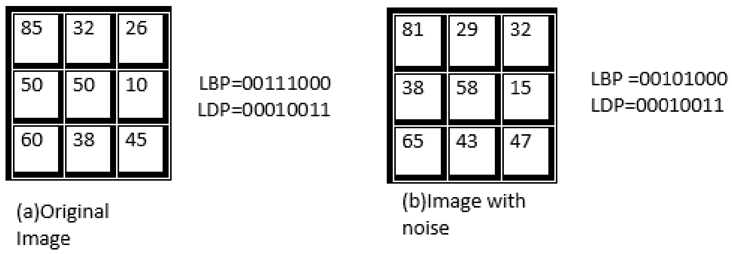


FIGURE 2.12: Illustration of effect of noise on LBP and LDP(a)is the orginal image (b)The image after noise is added

From the results obtained by Jabid et al [19], the feature descriptors LDP and LBP performed at the same level of 0.97 and the PCA at 0.85 under expression variation. However with reference to age the LDP had a higher performance of 0.72, followed by the LBP with 0.66 and PCA with 0.44. The LDP also outperformed the other descriptors with regards to illumination showing that it is more robust to lighting conditions and age variations.

Hybrid approaches have also been used to classify gender by combining two approaches. The fusion of the two approaches to come up with a hybrid approach [71]. Their proposed system had three main modules, one for normalizing the image. The findings of the research showed that the hybrid approach has a higher classification accuracy 92%, than both the appearance approach which made use of 143 weak classifiers 82,97% and 88.6% for the geometry with ten features.

The findings of previous researchers have encouraged the implementation of not only the LBP feature extractor which have been shown to have a number of shortcomings as pointed out by Tabid et al. [19], when the LDP was put forward. Hence we have used the LDP and LBP which make use of gradient and image pixels to extract information.The PCA and LDA are chosen for dimensionality reduction. This is because gender classification is a binary problem and PCA will only show directional variance, however LDA maximises the inter-class distances and hence makes classification easier [88].

Chapter 3

Methodology

3.1 Introduction

As we seek to determine gender using the facial components different techniques and methods are used. In this chapter the methods and techniques used for gender classification in this dissertation are presented.

3.2 Proposed Gender Classification Model

The model used in this thesis initially begins with obtaining a dataset and in this case the Face and Gesture Recognition Research network(FG-NET) and the Pilots Parliament Benchmark (PPB) are used. The PPB contains individuals with varying ages and hence we are able to test our model on gender with different age ranges and the latter has individuals with varied skin colour. The proposed system is shown in Figure 3.1.

The main modules that make up the system are feature detection, feature extraction and then classification being commonly used by many other researchers [71].

3.2.1 Image Acquisition and Pre-processing

The FG-NET provides a mixture of grayscale and colour images, whilst the PPB dataset consists of images which are all colour, hence since grayscale requires less computational power for processing we convert images from, colour to grayscale.

3.2.2 Face and Feature Detection

After the two datasets are obtained, PPB and the FG-NET, the face is first detected using the Viola-Jones algorithms [72], after which the facial components, the eyes, nose, mouth are detected using the pre-trained cascades. The rest of the components, the forehead and cheeks are detected using the locations of the three components detected first.

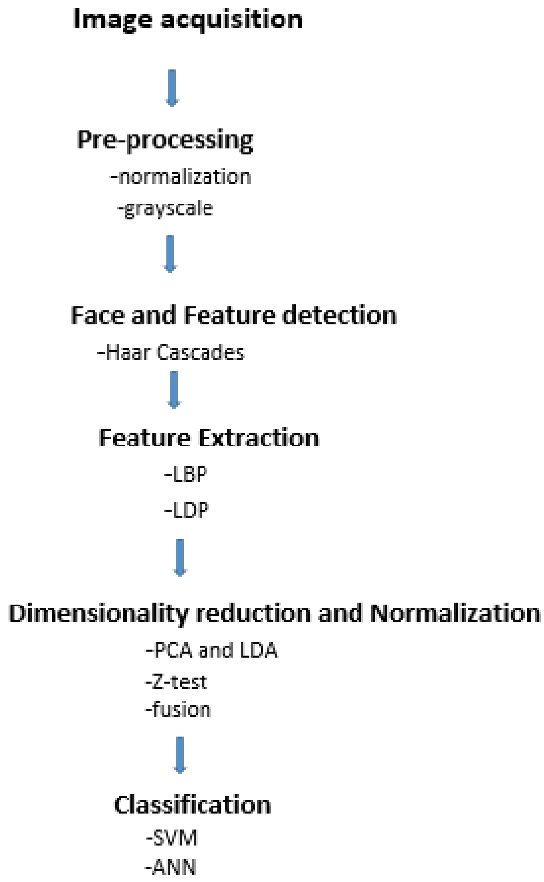


FIGURE 3.1: Proposed System Model

3.2.3 Feature Extraction

The appearance based approach to feature extraction and the Shape features are both used in this research. The Local Binary Pattern (LBP) and the Local Directional Pattern (LDP) are used here which use pixel and gradient to analyse images. The Active Shape Model is used to detect and extract shape features.

3.2.4 Dimensionality Reduction and Normalization

The face has many features some of which are not useful when it comes to classifying gender resulting in what is called, "the curse of dimensionality" [84], hence the PCA, an orthogonal transformation is used to show the main variance of the extracted features and the LDA is used to reduce in-class distances. Normalization by fusion is carried out to ensure that matrix values lie within the same range.

3.2.5 Gender Classification

Classification is the carried out using an Artificial Neural Network and the Support Vector Machines.

3.3 Image Acquisition and pre-processing

After obtaining the datasets operations are then carried out on the given images to enable the resulting image to be more suitable than the original image to carry out operations on it [92]. Images from the two datasets are shown in 3.2 and 3.3

Original Images FG-Net

FIGURE 3.2: Images from FG-NET dataset

From the images in FIGURE 3.2, we see that the PPB dataset has only colour images and the FG-NET has a mixture of grayscale and colour, however in our techniques all the images are converted to grayscale.

Previous researchers have noted that the use of grayscale images reduce the computational power required to perform operations on images [94]. Preprocessing may include image alignment, rotation, illumination equalization and smoothing. Hence the first step carried out under this section is converting the images from colour to grey scale as for most applications colour information does not help us with valuable information on edges or any other information which may be valuable. Figure 3.4 illustrates the image after grayscale conversion.

FIGURE 3.3: Images from PPB dataset

FIGURE 3.4: Conversion of image to grayscale

The colour image is converted to grayscale using the openCV function colour to grayscale method [117]. In grey scale images, the pixels represent different shades of gray without apparent colour. Hence this conversion from Colour to grayscale results in a reduction from a three-dimensional colour data into a singular dimension and resulting images are composed exclusively of shades of grey ranging from black with zero intensity to white at 255. According to Reddy et al [118], intermediate shades of gray are represented by equal brightness levels the three primary colours (red, green, blue). These RGB values have a constant applied to them. This is shown in the equation below :

$$\text{RGB to Grayscale} : Y \leftarrow 0.299.R + 0.587.G + 0.114.B \tag{3.1}$$

In equation 3.1, Y refers to the pixel value obtained by operating on the Red, Green and Blue pixel values.

After the image has been converted to grayscale we then perform normalization, which results in a change of the pixel intensity values. This process may also be referred to as histogram stretching done to have the image into a range that is more familiar or normal to the senses. Normalization takes an n-dimensional grayscale image.

$$I : X \subseteq R^n \rightarrow Min, .., Max \tag{3.2}$$

The original grayscale image is converted to the new normalized image below

$$I_N : X \subseteq R^n \rightarrow newMin, ..; newMax \tag{3.3}$$

Initially in the equation 3.3, the original grayscale image is used as input with pixel intensity values ranging from $Min \rightarrow Max$ which is then transformed by the normalization process to the new image with a range of pixel intensity values ranging from new extreme values $newMin \rightarrow newMax$.

3.3.1 Face and Feature Detection

To carry out feature extraction of either the individual components or the face as a whole,the face should first be detected. According to Zhang et al [51], appearance based methods compared to face detection have had the best performance when it comes to face detection and we have used them to carry out detections. The Viola Jones algorithm has three main modules which make it efficient which are the integral image which efficiently computes the sum of values in a rectangle subset, the adaboost Learning which combines many weak hypothesis and the attentional cascade. Of great importance when looping through the datasets is the use of the globe function as shown below which is an extract of looping over the facial images, converting the image to grayscale among others taking in a vector of file-names and a path to the dataset.

Once the face has been detected, we start by detecting the eyes, nose, mouth using Viola and Jones [44]. The eyes, nose, mouth and detected using pre-trained cascades as shown in Figure 3.5. The forehead, cheeks are detected using face geometrics. The pseudo code for detecting facial features as shown in Algorithm 1.

After detecting the face, nose, eyes and mouth we locate and extract the forehead and cheeks using a geometric approach.

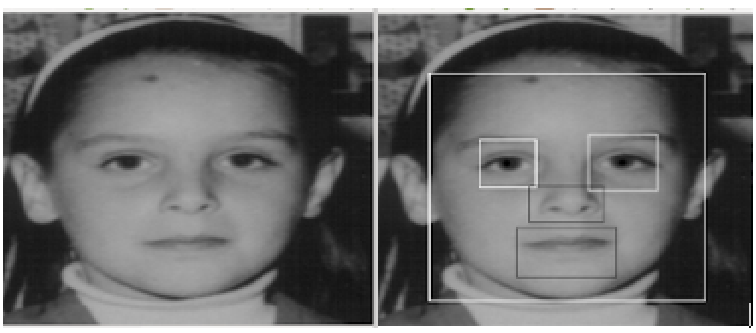


FIGURE 3.5: Facial components detected using Haar cascades

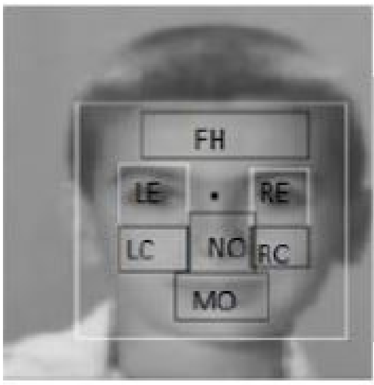
Algorithm 1: Detecting Facial Components

```
function Detect Facial features;
Input  : faces, cascades (eyes, nose ,mouth), labels
Output: detected(eyes, nose, mouth, forehead, cheeks)
for each image i in dataset do
    set paths to dataset and cascades;
    if file path or face cascade path are empty then
        print error message;
    convert image to gray scale;
    detectface ;
    draw bounding box on face(ROI);
    save ROI;
    for each face detected do
        if eye cascade is not empty then
            if eyes are being detected in ROI then
                mark center of eyes and draw bounding rectangle;
                crop and save eyes image;
                use eye width and height to approximate cheek location ;
                crop and save cheeks;
        if nose cascade is not empty then
            draw rectangle around nose;
            crop and save nose;
        if mouth cascade is not empty then
            if mouth is in ROI then
                draw bounding box around mouth;
                crop and save the mouth;
        draw forehead bounds using corner center of eyes and half of eye width
          distance;
        crop and save forehead;
    end
end
```

High speed and detection rate of Viola and Jones classifier guided its choice for this study. The rectangular region around, and including the eyes is detected and cropped. Haar-Cascade classifier is used to detect nose and mouth separately. The bounding rectangles around the nose and mouth are found and Region of Interest (ROI) related to them are detected. Due to high chances of false detections for nose and mouth, threshold is used to ensure that the mouth is always below the nose.

To carry out this detection the forehead refers to an area that is above the eyes as shown by the rectangle from the top left to the bottom right corner which we have drawn directly above the centre of the eye. Coordinate values x1 and x2 correspond to x coordinates of left eye and right eye respectively. The top-left corner of the cheek bone area is found as point $(x; y + h)$ where x and y are the x and y coordinate of the top-left corner of respective eye, h is the height of the eye. Bottom-right corner of cheek bone area is found as point $(x + w; y + 1 : 5 \times h)$ where w is the width of the respective bounding rectangle around the eye. Coordinate values $y1, y2$ are found by going over the dataset to determine a range of values which may be applicable and selecting one.

The top-left corner of RE is considered as $RE(Righteye)andLE(LeftEye)$ and its coordinates are referred to as Right eye top left x-coordinate and Right eye top left y-coordinate $RE.TLx, RE.TLy$ respectively. Geometric information about the eyes is then used to detect the corresponding cheeks, as cheeks are always below the eyes. The centre of LE denoted as LE:C and centre of RE denoted as RE:C are derived. The y-coordinates of the forehead's top left corner and bottom right determined by the top left and top right points of the left and the right eye.



Key	
FH	Forehead
LE	Left Eye
RE	Right Eye
LC	Left Cheek
NO	Nose
RC	Right Cheek
MO	Mouth

FIGURE 3.6: Bounding rectangle and component detection

3.4 Feature extraction

This section focuses on the feature extraction techniques which are used in this research. Two texture descriptors are used here which are the Local Binary Pattern and the Local Directional Pattern the Active Shape model.

3.4.1 Local Binary Pattern

In this texture based approach we divide the image into 9 blocks. Images in computers are represented in a complex way ,however since we intend to perform some mathematical operations may be applied upon them easily. The original Local Binary Pattern is used in this case taking a 3×3 block, where each pixel is thresholded against the center value as shown below with the final binary value converted to a decimal and these are the Local Binary Patterns. The labeling of each pixel of an image by thresholding its P-neighbor values with the center value and converts the result into binary using the equation 3.4.

$$
\begin{aligned}
LBP(x,y) &= \sum_{n=0}^{k-1} s(i_n - i_m)2^n \\
s(x) &= \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}
\end{aligned}
\tag{3.4}
$$

In the equation above i_m represents the middle pixel at coordinate(x,y) and i_n represents the neighbourhood pixels.

The Local Binary Pattern code is calculated with R=1, P=8 and histogram with $2^8 = 256$ for each region is generated after which the histograms are concatenated into one feature vector of size 9×2^8 which is used for texture as shown in Figure 3.7.

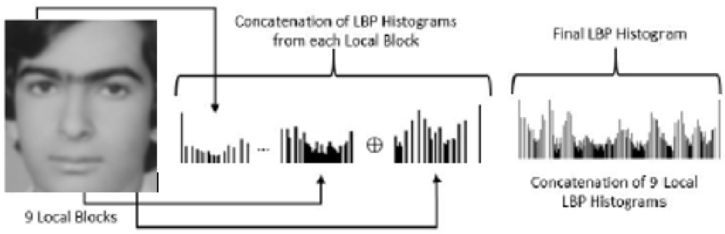


FIGURE 3.7: LBP feature extraction [21]

3.4.2 Local Directional Pattern

The Local Directional Pattern (LDP) feature extraction method assigns an 8 bit binary code to each pixel. Hence to find the response edge of a pixel we have used the Kirsch mask though there are a number of other filters which may be used such as the Prewitt edge detector. The LDP compares neighbouring pixel gradient magnitudes along a specific direction [19], researchers have found that the presence of edges and corners lead to high response values and hence we are only interested in the K-most important directions and setting them to 1 and the remaining gradients are converted to zero.

$$LDP_K = \sum_{i=0}^{7} b_i(m_i - m_k)2^i$$

$$b_1 = \begin{cases} 1 & \text{if } a \geq 0 \\ 0 & \text{if } a < 0 \end{cases} \tag{3.5}$$

In the equation 3.5, m_k refers to the k-th most significant directional responses. As with the LBP, after encoding the image we now have an encoded LDP image. A histogram has been widely used to analyse and characterize images [119]. Hence to find the number of bins when using k=3 after using a Kirsch mask from 8 different directions we use the combination rule since we choose k-most significant edge responses $8C3$ to get 56 histograms as shown by the equation below where I_L represents the encoded image of the Histogram(H) in equation 3.6.

$$H_i = \sum_{x,y} P(I_L(x,y) \tag{3.6}$$

At this stage every feature or face is represented as a histogram the image has to be divided into small regions initially which will each have a histogram of its own after which these histograms are concatenated into one which represents the whole feature and pushed into a vector.

3.5 Dimensionality Reduction

According to Jadhao [81], facial image data is always high dimensional and this has meant that it requires considerable computing time for classification and concatenation of histograms leads to the huge number of features with a considerable amount of the features being redundant and this has been referred to as "the curse of dimensionality", which has been described as the problem of finding structure in data embedded in a highly dimensional space as the more features we have, the more data point which will be needed to fill space [83]. To carry this Process out we have used the Principal Component Analysis and the Linear Discriminant Analysis.

3.5.1 Principle Component Analysis

After feature extraction we have to select the most important features and throw away the rest, as the data would be of high dimensionality and hence has to be reduced using Principal Component Analysis (PCA). PCA is a very well-known method of identifying patterns in data and expressing data in such a way that to highlight their similarities and differences. High dimensional data usually contains data which may have many irrelevant variables [84]. Hence we used PCA to reduce the feature vectors columns(dimensionality reduction) before which we would have converted the vector into row-matrix form and then dimensionality reduction using PCA is carried out.

To carry out PCA it takes in an m-by-n matrix X, consisting of n number of m-dimensional vectors $\vec{x}_i \in R^m$. The first step is to compute the mean and the co-variance of the data matrix, which we will call$S \in R^{m \times m}$. The covariance of the matrix is defined by the equation 3.7.

$$S = \frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i - \vec{x})(\vec{x}_i - \vec{x})^T \tag{3.7}$$

In the equation 3.7 $\bar{x} \in R^m$ refers to the mean of each row in X and is defined in equation 3.8.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}\vec{x}_i \tag{3.8}$$

After calculating the mean we extract the Principal Components using the Singular Value Decomposition(SVD) to obtain the Singular Value Decomposition S, as PCA is performed via SVD.

$$S = UEV^T \tag{3.9}$$

In the equation above V is a matrix of eigen vectors, with each column being an eigen vector and E being a diagonal, matrix.

3.5.2 Linear Discriminat Analysis

Since we seek to solve a binary problem we then use Linear Discriminant Analysis, as PCA only gives us the main variance of data. The main aim of LDA is to find optimal projection, vectors simultaneously minimizing the within-class distance and maximizing the between-class distance for the given data [88]. The drawback of using PCA before LDA is that in cases where there are few samples is that it may result in valuable texture information being discarded [89]. The LDA is also performed in this experiment as to maximize between the classes scatter and minimizing within class scatter without losing data and hence for all sample classes the LDA defines two measures, the first being the within class scatter:

$$S_W = \sum_{j=1}^{c}\sum_{N_{i=1}}^{N_j}(x_i^j - \mu_j)(x_i^j - \mu_j)^T \tag{3.10}$$

In the equation 3.10, c is the number of classes and$x(i)^j$is the i-th sample of the class j, μ_j and N_j the number of samples in class j. Equation 3.11 is the equation for the between class scatter.

$$S_b = \sum_{j=1}^{c}(\mu_j - \mu)(\mu_j - \mu)^T \tag{3.11}$$

In equation 3.11 μ represents mean of all classes. After dimensionality reduction has been carried out the vector is then fed into a classifier.

3.5.3 Feature Fusion and Standardization

The fusion of features requires us to carry out standardization in the cases in which feature fusion is performed. The primary goal of feature fusion is to have one feature offset the classification weakness of another feature. As one component may also have a high level of noise associated with it, hence we fuse the mouth, which had the highest classification rate, from the findings of a number of researcher such as Belhumeur [120], with the eyes.

3.5.4 Standardization

Before feature fusion is performed we need to standardize our feature matrices, there are a number of ways to perform this and we have performed Z-score standardization on the two feature matrices before fusing them by addition. This can be done by performing the three steps that follow:

$$\mu = \sum_{1}^{n} \frac{x}{n} \tag{3.12}$$

Above μ is the mean ,n is the number of samples and x is the particular value to be standardized. One has to find the standard deviation as shown below

$$\sigma = \sqrt{\frac{\sum(x-\mu)}{n}} \tag{3.13}$$

Then standardizing each matrice value Z-score = $\frac{x-\mu}{\sigma}$

After standardizing each matrice value in cases where feature fusion will be carried out we then add the two matrices using matrix addition function in openCV as shown below:

$$\begin{bmatrix} a & b & c \\ d & e & f \\ h & i & f \end{bmatrix} + \begin{bmatrix} k & l & m \\ o & p & q \\ s & t & u \end{bmatrix} = \begin{bmatrix} a+k & b+l & c+lm\, d+o & e+p & f+q \\ h+s & i+t & f+u \end{bmatrix} \tag{3.14}$$

In the equation above the two normalized matrices are added to get the final fused matrix and then dimensionality reduction is performed.

3.6 Classification

This is the final stage of the gender classification system in which we determine from the given image if one is a male or a female, binary classification. A Multi-Layer

Perceptron Neural Network with back-propagation is used along with a Support Vector Machine. These two classifiers have been selected due to their significant success in pattern recognition in recent years [121].

3.6.1 Artifical Neural Network

The number of layers in the Multi Layer Perceptron are determined empirically with different values for the parameters being tried in the evaluation dataset and then being tested on unseen test data. In each case the input layer is assigned neurons equivalent to dimensionality of the input features. The MLP is structured to have the size of the input layer being equal to the number of columns in the matrice. We then use seven nodes at the two hidden layers.

A sigmoid function is used in the hidden layers as we are dealing with a binary classification problem and maps resulting values between 1 and 0, as we have also set our result to be either male or female as we seek to scale data in some specific range with a threshold.

3.6.2 Support Vector Machine

The SVM has been found to be the most popular and successful binary classification method [18]. In the research we used the Support Vector Machine(SVM) with a (Radial Basis Function (RBF) kernel as defined in equation 3.15.

$$K(x_i, x_j) = e^{-\gamma |x_i - x_j|^2} \tag{3.15}$$

Different pairs of (C, ϵ) are tried and parameters that give lower cross validation accuracy are chosen The parameter ϵ determines the number of support vectors which will then be used to construct a regression function controlling the width of the $\epsilon - insensitive$ zone. The search for the best parameters is known as grid-search optimization.

3.7 Validation and Evaluation Protocols

3.7.1 Validation Protocol

We split the data into train, validate and test made up of 300 images each. From the validation we chose the best hyper-parameters and then used these directly on the Test data which was unseen to our algorithm from which we measure the performance of our model. In the validation data set we determine the best value to be used for the SVM and ANN parameters such as the number of nodes, layers, SVM types to mention a few which are then used on unseen data, the test data set.

3.7.2 Evaluation Protocol

From the evaluation protocols described, we chose accuracy as our datasets are nearly well balanced between male and female. The equation is shown below:

$$Accuracy = \frac{TP + TN}{TP + FN + TN} \tag{3.16}$$

In the equation above TP refers to True Positive,TN refers to False Negative and TP refers to True Positive.

3.7.3 Image Data set

The FG-NET and PPB ageing database is used to evaluate the proposed age estimation approach. The first a publicly available ageing dataset that has 1002 images of 82 subjects aged between 0 and 69 years. Images have wide variation in illumination, colour and expression. Some images have poor quality since they were scanned distribution of images in FG-NET ageing dataset in 10-year age groups. The limitation of this dataset is that it has few images in adult age groups (above 30). The gender distribution for FG-NET is 34 females and 48 males made up of 431 images and 571 images respectively. The PPB has images which have been labelled phenotypically with skin type being chosen, it contains 1270 images of different individuals from South Africa, Senegal, Rwanda, Sweden, Iceland and Finland.

Chapter 4

Results and Discussion

4.1 Programming Environment

This chapter puts forward the various algorithms which are used in this research. The research was carried out and run with Intel Core(TM)i7-3520M CPU @2.90Hz with), 8GB installed RAM. Microsoft Visual Studio 2013 is used as the development environment of choice.

4.2 Datasets

This investigation made use of two different datasets the FG-NET [122] and the PPB [26]. The FG-NET is made out of 1002 images made up of on average 12 age-separated images per individuals. The PPB dataset is made up of 1270 different individuals, it was put together as part of a research on the "Accuracy disparities in Commercial Gender Classification. The breakdown of the datasets is given below:

TABLE 4.1: Datasets used

Property	PPB	FG-NET
Release year	2017	2004
Subject number	1270	82
IM Width	160-590	300-400
IM Height	213-886	300-350

4.3 Face and Feature Detection

The face and facial components are automatically detected as illustrated in Figure 4.1 from the grayscaled facial image and the cropped out and saved into an labelled folder for the facial components. Pre-trained haar-cascades are used to detect the face, after which geometric operations are carried out to detect the forehead and cheeks as illustrated in Figures 4.2 -4.4.

The performance of the pre-trained Haar-Cascades detect 915 face areas out of 1003, giving an accuracy of 91.2%. After which we then move on to detecting the various facial components the eyes, nose, forehead, cheeks and the mouth. As shown shown in FIGURES 4.1-5:

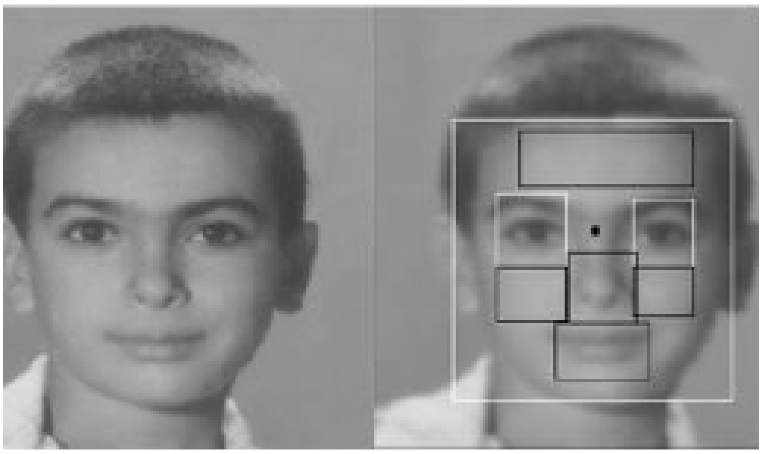


FIGURE 4.1: Detected face components

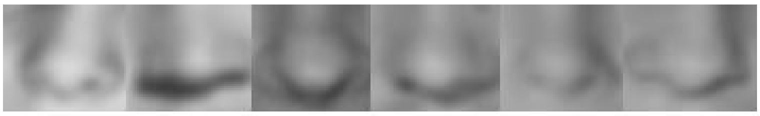


FIGURE 4.2: Detected Nose Components

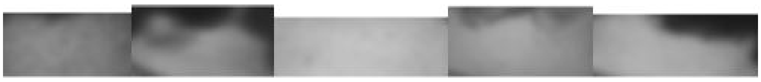


FIGURE 4.3: Detected forehead components

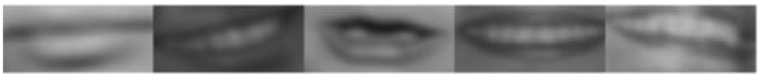


FIGURE 4.4: Detected mouth components

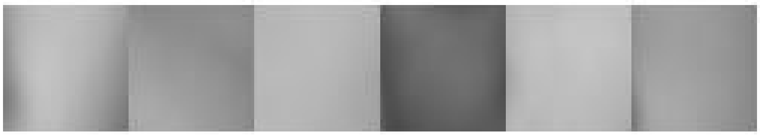


FIGURE 4.5: Detected cheek components

The images shown in FIGURE 4.1-4.5 the ones which are detected accurately by our proposed model. However though we detect some facial components successfully some are classified inaccurately and the following example in Figure 4.6 shows this.

As can be clearly seen from the above results the face, recording a 92.2% and 89.7% accuracy for the PPB dataset. The forehead for the FG-NET records the lowest accuracy rate which may be a result of the fact that to detect it, we first have to determine the centre of the eye and a height above the eye to determine the top left corner for example and as a result some forehead images are detected although having the hair covering a certain fraction of the extracted image.

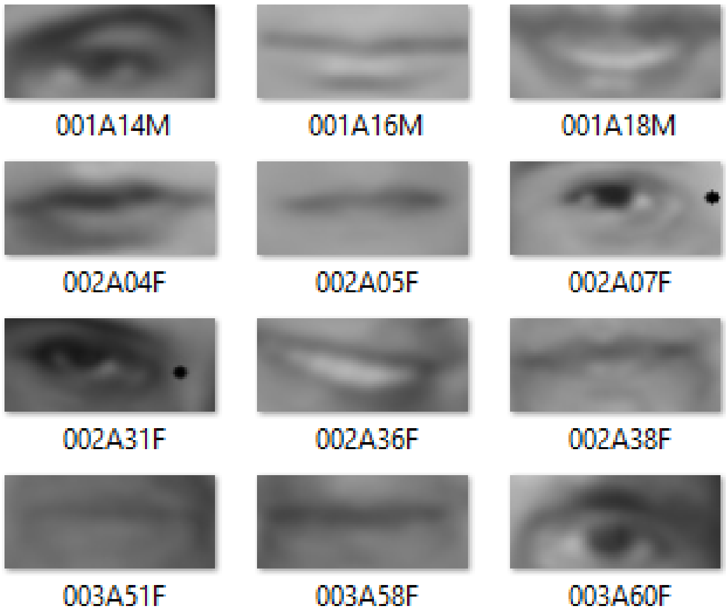


FIGURE 4.6: Eyes are also mis-classified as mouths

TABLE 4.2: Component detection accuracy

Component	FG-NET	PPB
Face	92.2%	89.7%
Nose	90.6%	88.2%
Mouth	90.3%	87.3%
Cheeks	89.6%	86.8%
Mouth	91.4%	87.4%
Forehead	89.32%	87.1%

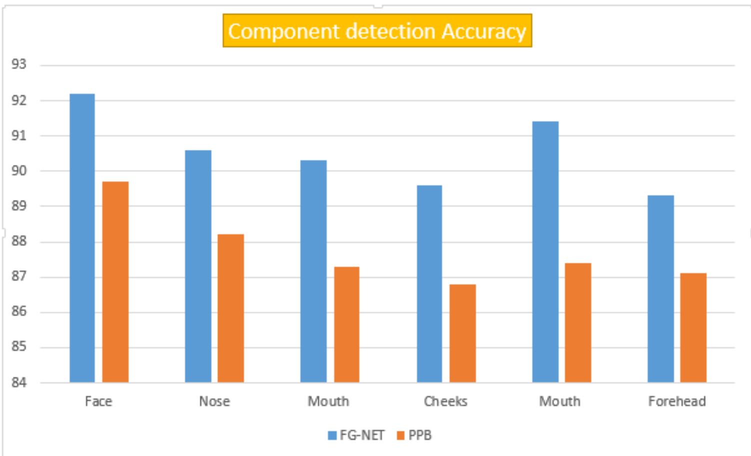


FIGURE 4.7: Component detection Accuracy

4.4 Gender Classification Using an Artificial Neural Neural Network With Backprogation

A neural network with two hidden layers is used to classify gender first and has been chosen as it is simple and also applicable to a wide variety of areas [123]. Neural network had been used in Deep Learning with great success either in image processing or in gaming as shown by Silver et al. [124].

TABLE 4.3: Gender Classification using single components

Component	Accuracy
Nose	90%
Eyes	87%
Mouth	75%
Forehead	92%
Cheeks	89.2%

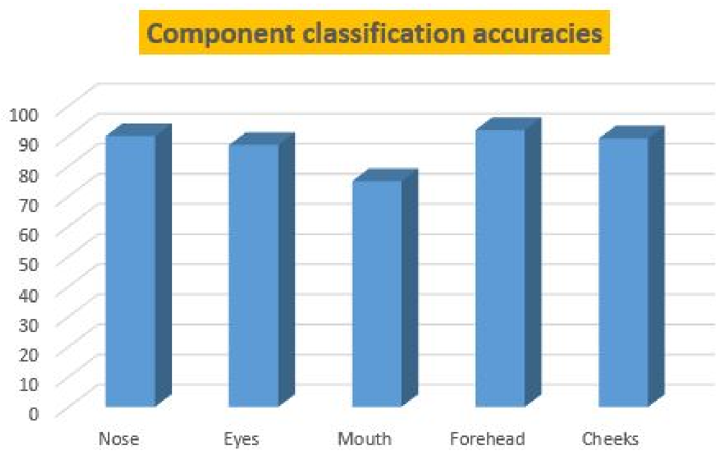


FIGURE 4.8: Bar chart for component detection

From the results obtained in this research using the Artificial Neural Net (ANN) with backpropagation the forehead showed the greatest gender prediction accuracy with 90% of the image genders being predicted accurately and an average classification accuracy of 86.64% being recorded. This experiment on individual components was inspired by the findings of Brown et al. [125], by carrying out experiments which showed that individual components such as the eyes, nose, mouth among others in isolation carried information about gender. We hence compare our results to that of Chen and Lu [6], who however adopted an Active Shape Model to get location of three facial components on the FERET data set using the Support Vector Machine as a classifier.

TABLE 4.4: Comparison of classification accuracies on FG-NET

Component	Proposed Model	Results of Chen et al[6]
Nose	90%	85.98%
Eyes	87%	66.82%
Mouth	75%	82.71%

From the above results the LBP with an Artificial Neural Network outperformed the SVM and ASM used by Chen et al [6], for the nose and mouth but is however outperformed with regards to the mouth, hence we perform feature fusion to improve it aiming to improve the feature classification accuracies. This is clearly shown in Figure 4.9.

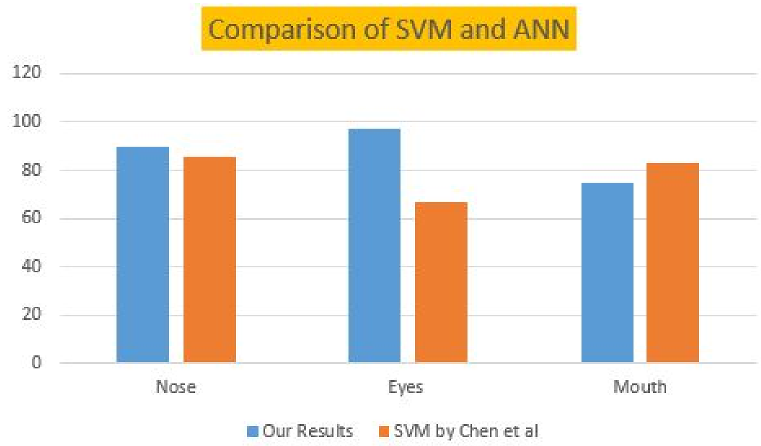


FIGURE 4.9: Bar chart for component detection of our results and those of [6]

The presence of individuals with a beard in the FG-NET dataset may be the cause of the mouth having the lowest classification accuracy, hence we then seek to improve this accuracy by fusing it with a component such as the nose, which has a higher accuracy. The nose, mouth and nose as shown above recorded an average accuracy of 84% compared to that of 78.5% by Chen et al [6].

We perform component fusion on our proposed system above and the results are shown in the table below, we have fused the forehead and the nose with all the other components as Table 4.3 showed the two having the highest classification accuracies of 92% and 90%.

TABLE 4.5: Gender Classification by fusing features on FG-NET

Features	Results
Forehead + eyes	93%
Forehead +Nose	94.6%
Forehead+Cheeks	93.7%
Nose+Eyes	92%
Nose+Cheeks	91.3%
Nose+Mouth	87%

Table 4.5 shows clearly that gender classification is improved by fusion of features based on FG-NET dataset images, especially when we fuse the highest individual classifiers with the features with the lowest classification accuracy. The mouth's accuracy when fused with the Nose increases from 75% to 87% with an enhancement of 12%. The forehead having an accuracy of 90% is fused with the eyes with an accuracy of 87% recording a slight improvement from both individual to a fused accuracy of 93%. The accuracy enhancements obtained by fusion is shown in the following bar chart to illustrate the enhancements obtained. The table showing enhancements is shown in Figure 4.1.

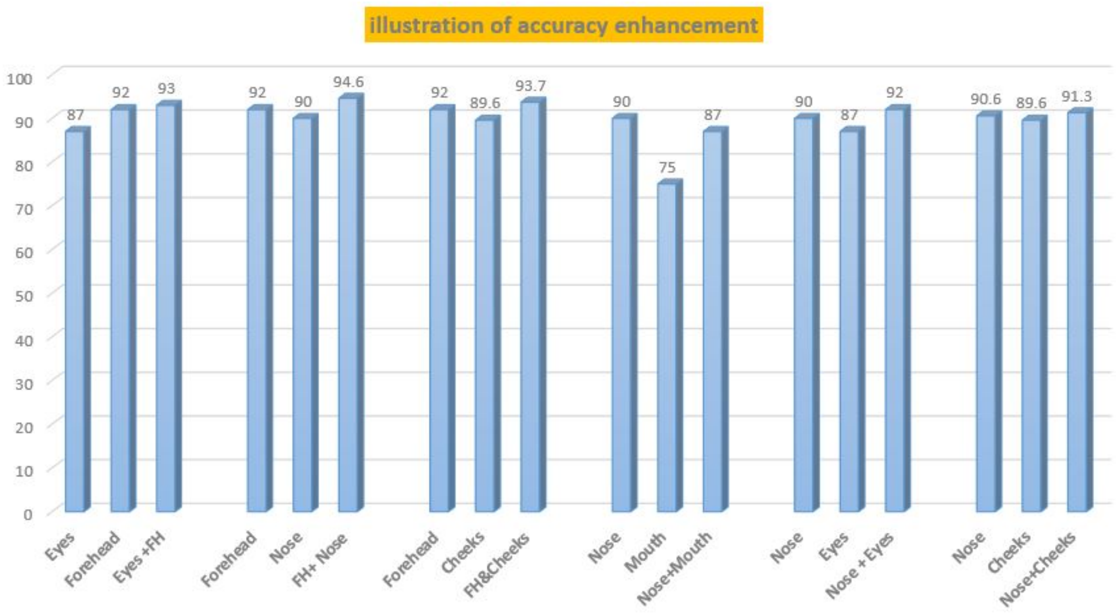


FIGURE 4.10: illustration of fusion results

FIGURE 4.10 is a digrammatic illustration using bargraphs of Table 4.5 after fusing the various facial components.

4.5 Comparison of Local Binary Pattern,Local Directional Pattern and Active Shape Model Feature Extraction with Fusion

This research moved on to compare the performance of two appearance based approaches to feature extraction which are the LDP and the LBP, along with a shape feature descriptor Active Shape Model using the detected facial points. Sixty-eight points are detected in the face as shown in Figure 4.11.

The findings of this research showed that the LDP outperformed the LBP as with the findings of Jabid et al [19], as pixel values have been shown to be affected by noise. The LDP outperforms the LBP on FG-NET and the PPB dataset.

TABLE 4.6: Gender Classification using SVM with various feature extractors

Extractors	FG-NET	PPB
LBP	85.33%	83.13%
LDP	92.85%	89.26%
LDP + ASM	94.53%	81.56%
LBP +ASM	89.53%	85.43%

Feature fusion is then carried out hereby fusing shape features of both the LBP and LBP with ASM as show below and enhancement is clearly in Figure 4.12. Fusing the LDP and the ASM results in an enhancement of 1.68% and that of the LBP and ASM achieved an enhancement of 4.17% both on the FG-NET dataset. On the PPB dataset fusing the LDP and ASM resulted in an increase of 7.7% compared to that of

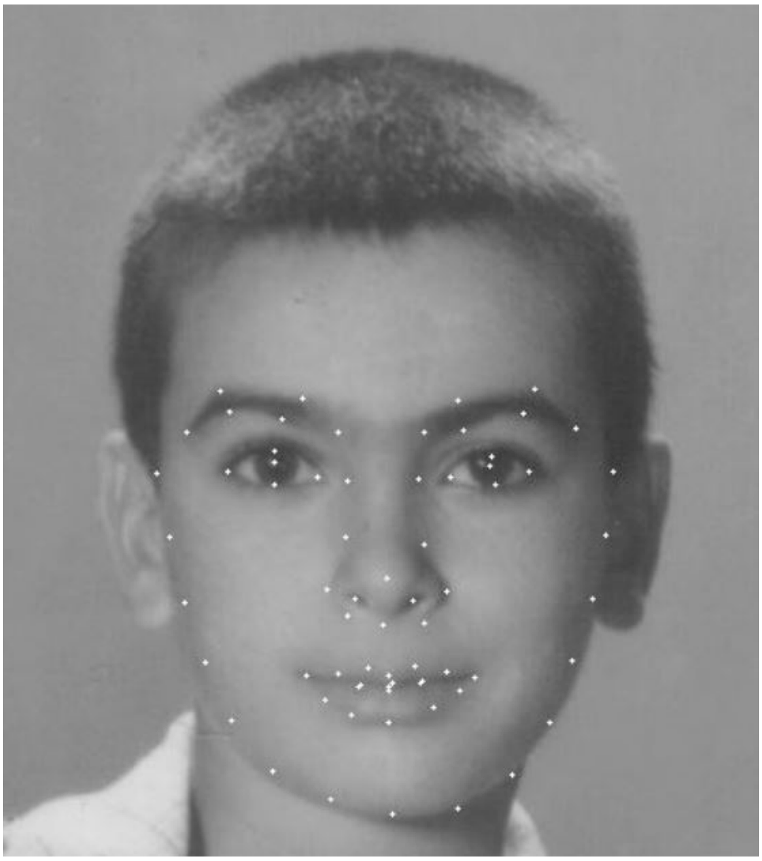


FIGURE 4.11: Detected points in a facial image

2.3% of LBP and a lesser enhancement of 2.3%, hence fusion here achieves accuracy enhancements.

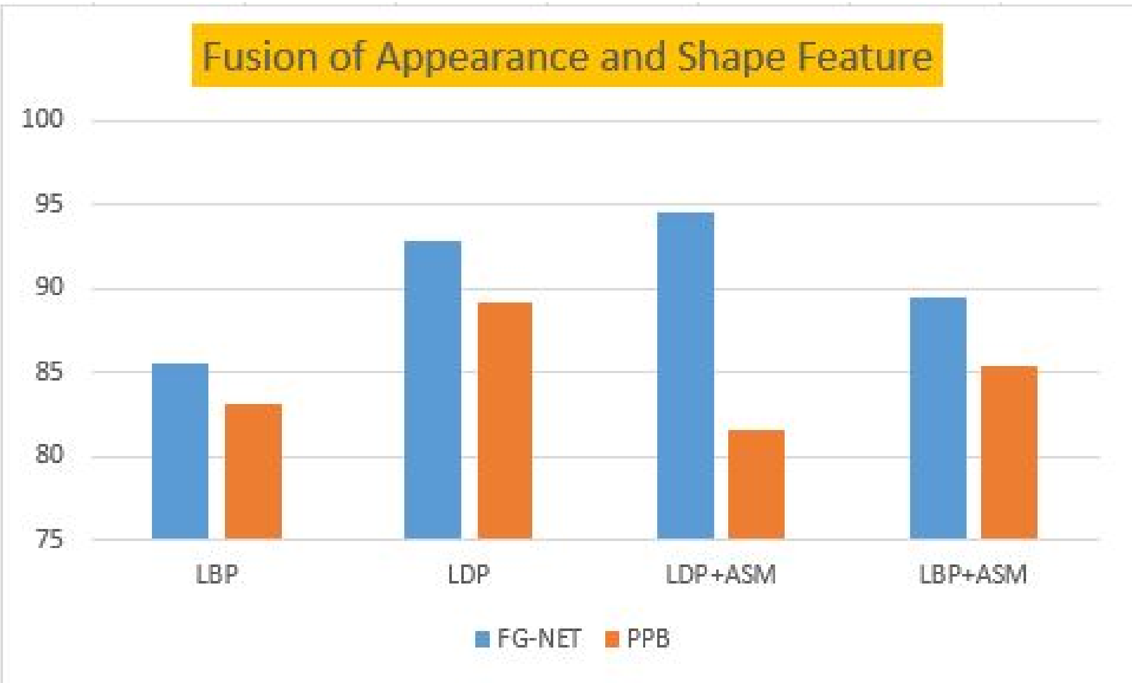


FIGURE 4.12: Enhancement from fusing shape and appearance features

Chapter 5

Conclusion and Future Work

5.1 Summary of work

This research provided a comprehensive analysis of gender classification and feature extraction techniques the LBP, LDP and the ASM. The appearance based approach was catered for by the LDP and the LBP with the shape features being represented by the ASM. Gender classification was carried with the whole facial area as input and on individual facial components. On reviewing performance of the gender classifiers on the individual components from previous researchers and our own we then used feature fusion in an aim to improve the accuracy rates.

To extract feature the Local Binary Pattern is one of the extractors used and we chose the 3×3 (8 neighbours) neighbourhood, the original LBP and thresholding each centre pixel by its neighbours and converting the resulting binary codes are converted to decimal and histograms are used to represent the resulting decimal representations which are then concatenated to form a feature vector. The LDP however looks at the edge response values in all different eight directions rather than surrounding pixels as done by the LBP and hence each pixel is considered from the 8 directions using the Kirsch mask. Shape feature points are extracted using the ASM to then form a feature vector which is also preprocessed before being fed into a classifier.

Dimensionality reduction is carried out, as the "curse of dimensionality" results in many redundant features being present as witnessed by the great of LBP columns and hence PCA and LDA are applied. In the case of fusion, normalization is carried out to ensure features are of the same range before fusing the two vectors. Two classifiers are implemented, the SVM and Multilayer Perceptron(MLP).

The results of this research came up with key findings first being that the use of Shape features with either LDP or LBP leads to a higher classification accuracy as shown TABLE 4.6. Hence futher research would need to be carried out to determine if Shape feature alone can outperform the other approaches individually. The use of hair for classsification is an important takeaway from this research using the technique put forward by [74], who found that it had higher classification accuracies than other facial components.

5.2 Future Work

5.2.1 Face Detection

The use of the Haar cascades for face detection recorded an accuracy of 91.2% on the FG-NET. However this can be improved with the use of better techniques which may include background subtraction as in some cases the background is included as part of the face which has affected training and the resultant classification. Hence an approach that extracts the facial area in its actual shape rather than drawing a rectangle would improve accuracy results.

5.2.2 Use of Convolutional Neural Networks

With the amount of data being generated worldwide, images for this research were limited to labelled data however since we used a supervised learning approach we needed labelled datasets. However Deep Convolutional Neural Networks have shown great results in unlabelled data, as shown by Levi et al [126] and this approach would not require labelled datasets which are usually small and often limit training resulting in overfitting.

Bibliography

[1] W. L. Perry, B. McInnis, C. C. Price, S. C. Smith, and J. S. Hollywood, "Predictive policing: The role of crime forecasting in law enforcement operations", *Rand Corporation*, 2013.

[2] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: Past, present and future.", *Computer Vision and Image Understanding, 138:1*, 2015.

[3] H. ALABBASI and F.Moldoveanu, "Human face detection from images, based on skin color", *Proceedings of the 18th International Conference on System Theory, Control and Computing, Sinaia, Romania, October 17-19, 2014*, 2014.

[4] D. Reid, S. Samangooei, C. Chen, M. Nixon, and A. Ross, "Soft biometrics for surveillance: An overview. machine learn- ing: Theory and applications", *Elsevier, pages 327352, 2013*, 2013.

[5] T. Bissoon and S. Viriri, "Gender classification using face recognition", *Adaptive Science and Technology (ICAST), 2013 International Conference*, 2013.

[6] X. Chen and B.-L. Lu, "Gender classfication by combining facial and hair information", *Advances in Neuro-information Processing:15th International Conference ,ICONIP*, 2008.

[7] A. Jaswante and D. A.U.K.D. B. Gour, "Gender classification technique based on facial features using neural network", *(IJCSIT) International Journal of Computer Science and Information Technologies*, 2013.

[8] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition", *BMVC*, 2015.

[9] M.Kosinski and Y.Yang, "Deep neural networks are more accuate at detrmining sexual orientatiojn from facial images", *Journal of Personality and Psychology*, 2017.

[10] S. writer. (Jun. 5, 2018). You can now open an fnb account by taking a selfie, [Online]. Available: `https://businesstech.co.za/news/banking/244887/you-can-now-open-an-fnb-account-by-taking-a-selfie/`.

[11] B. A.M and N. Bruce V. Dench, "Whatâăźs the difference between men and women? evidence from facial measurement. 22, 153âăş176 (1993)", *Perception*, 993.

[12] H. A. Wild, "Recognition and sex categorization of adultsâăź andbchildrenâăźs faces: Examining performance in the absence of sex-stereotyped cues," *.,J. Exp. Child Psychol. 77(4), 269âĂŞ291*, (2000).

[13] O. dictionary. (Jun. 4, 2018). K. Martin, Ed., [Online]. Available: `https://en.oxforddictionaries.com/definition/gender`.

[14] G.Herdt, "Third sex third gender", *New york Zone*, 1996.

[15] Y.Andreu and R. Mollineda, "The role of face parts in gender recognition", *ICIAR : Image Analysis and Recognition*, 2008.

[16] G. Bommagani and O. Singh, "Pattern based gender classification", *International Journal of Advanced Research in Computer Science and Software Engineering*, 2013.

[17] N. Nilsson, *Introduction to Machine Learning*. Stanford, 1998.

[18] V.UtkinaAnatoly, I.ChekhbYulia, and A.Zhukc, "Binary classification svm-based algorithms with interval-valued training data using triangular and epanechnikov kernel", *ELS*, 2007.

[19] J. T, K. H, and C. O, "Gender classification using ldp", *IEEE Int. Conf. on Pattern Recognition*, 2010.

[20] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions", *IEEE Transactions on Image Processing, vol., ,* 2010.

[21] T. OJALA, M. PIETIKAINEN, and D. HARWOOD, "A comparative study of texture measures with classification based on feature distributions", *Pattern Recognition*, 1995.

[22] H. Zhang and S. Xu, "The face recognition algorithms based on weighted ltp", *Journal of Image and Graphics*, June.

[23] H.-C. LIAN and B.-L. LU, "International journal of neural systems", *MULTI-VIEW GENDER CLASSIFICATION USING MULTI-RESOLUTION LOCAL BINARY PATTERNS AND SUPPORT VECTOR MACHINES*, 2007.

[24] G.Guo, C.Dyer, Y.Fu, and T. Huang, "Is gender recognition affected by age?", *IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, 2009.

[25] G. B, L. D, and S. T, "Sexnet: A neural network identifies sex from human faces.", *Advances in neural information processing systems*, 1991.

[26] J.Buolamwini and T. Gebru, "Gender shades:interectional accuracy disparities in commercial gender classification", *Conference on Fairness, Accountability and Transparency*, 2018.

[27] S. Mozaffari, H. Behravan, and R. Akbari, "Gender classification using single frontal image per person: Combination of appearance and geometric based features", *2010 20th International Conference on Pattern Recognition*, 2010.

[28] A.Angulu, J. Tapamo, and A. Adewumi, "Human age estimation using multi-frequency biologically inspired features (mf-bif)", *IEEE AFRICON*, 2017.

[29] C. BenAbdelkader and P. Griffin, "A local region-based approach to gender classification from face images", *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPRâĂŹ05)*, 2005.

[30] G.Gua, C. Dyer, Y.Fu, and T. Huang, "Is gender recognition affected by age?", *IEEE 12th international conference on Computer Vision Workshops*, 2009.

[31] B. A and V. S, "Component-based ethnicity identification from facial images", *Computer Vision and Graphics. ICCVG 2016. Lecture Notes in Computer Science,*, 2016.

[32] Q. Tian, T. Arbel, and J. J. Clark, "Efficient gender classification using a deep lda-pruned net", *Computer Vision and Pattern Recognition (cs.CV)*, 2017.

[33] S. Khan, M. Ahmad, M.Nazir, and N.Riaz, "A comparative analysis of gender classification techniques", *International Journal of Bio-Science and Bio-Technology*, 2013.

[34] B.Heisele and V.Blanz, "Morphable models for training a component-based face recognition system", *Computer Vision and Pattern Recognition*, 2004.

[35] A.Dehghan, E.Ortiz, and G.Shu, "Dager: Deep age, gender and emotion recognition using convolutional neural networks", *article*, 2017.

[36] T. Cootes, C. Taylor, and D. C. J. Graham, "Cviu", *Active shape models-their training and application*, 1995.

[37] P.O'Hare and A. Fleischer, "Tobacco smoking smoking contributes little to facial wrinkling", *Journal of the European Academy of Dermatology and Venereology*, 1991.

[38] American, "Guidelines for psychological practice with lesbian, gay, and bisexual clients. american psychologist", *Guidelines for Psychological Practice with Lesbian, Gay, and Bisexual Clients. American Psychologist*, 2012.

[39] R. Jana, D. Datta, and Rituparna, "Age estimation from face images", *Procedia Computer Science*, 2015.

[40] D. Kim and J. Sung, *Automated Face Analysis Emerging Technologies and Research*, D. Kim, Ed. information science reference, 2009.

[41] C. B. Ng, Y. H. Tay, and B. M. Goi, "Vision-based human gender recognition: A survey", *Computer Vision and Pattern Recognition (cs.CV)*, 2012.

[42] E. MÃd'kinen and R. Raisamo, "An experimental comparison of gender classification methods", *Pattern Recognition Letters*, 2008.

[43] L. Wolin and P.Korgaonkar, "Web advertising: Gender differences in beliefs, attitudes, and behavior", 2005.

[44] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *IEEE Computer Society*, 2001.

[45] V. Gupta and D. Sharma, "A study of various face detection methods", *International Journal of Advanced Research in Computer and Communication Engineering*, 2014.

[46] C. Zhang and Z. Zhang, "A survey of recent advances in face detection", Microsoft Research Corporation, Tech. Rep., 2010.

[47] M.-H. Yang, D. J. Kriegman, and N. Ahuja., "Detecting faces in images: A survey", *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 2002.

[48] C. Kotropoulos and I. Pitas, "Rule-based face detection in frontal views", *IEEE Computer Society Washington, DC, USA Âl'1997*, 1997.

[49] B.Moghaddam and A.Pentland, "Probabilistic visual learning for object detection", *Proceedings of IEEE International Conference on Computer Vision*, 1995.

[50] I. C. David and T. Bennett, "Finding face features", *European Conference on Computer Vision*, 1992.

[51] C.Zhang and Z.Zhang, "A survey of recent advances in face detection", Microsoft research, Tech. Rep., 2010.

[52] G.Yang and T. Huang, "Human face detection in complex background", *Pattern recognition*, 1994.

[53] Q. Rizvi, B. Agarwal, and D. R.Beg, "A review on face detection methods", *ICP*, 2011.

[54] Q.Liu and G.ZhengPeng, "A robust skin colour based approach", *2nd International Asian Conference of Informatics in Control Automation and robotics*, 2010.

[55] M. F. Augusteijn and T. L. Skufca, "Identification of human faces through texture-based feature recognition and neural network technology", *IEEE International Conference on Neural Networks, San Francisco, CA,*, 1993.

[56] R. Haralick, K.Shanmugam, and I.Dinstein, "Texture features for image classifcation", *IEEE Trans.Systems,Man,Cyernetics*, 1973.

[57] S. M. Nagao and S. Fujibayashi, "Line extraction and pattern detection in a photograph", *Pattern Recognition*, 1969.

[58] N. Ahmed, N. T, and Rao, "Discrete cosine transform", *article*, 1974.

[59] D. P. Lale, Kailash.J, and Karande, "Gender classification using facial features", *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)*, 2016.

[60] H. H. K. Tin, “An effective age dependent face recognition system”, *Conference: International Conference on Multidisciplinary Research & Practice (ICMRP-2015) December 2015, At India*, 2015.

[61] F. Lin, Y. Wu, and Y. Zhuang, “Human gender classification: A review”, *Artificial Intelligence (cs.AI)*, 2012.

[62] A. F. Basha and G. S. B. Jahangeer, “Face gender image classification using various wavelet transform and support vector machine with various kernels.”, *International Journal of Computer Science Issues*, 2012.

[63] L. Cao, M. Dikmen, Y. Fu, and T. S. Huang, “Gender recognition from body.”, *In Proceedings of the 16th ACM international conference on Multimedia, pages 725âĂŞ728. ACM, 2008.*, 2008.

[64] H.-C. Kim, Z. G. Daijin Kim, and S. Y. Bang, “Gender classification with bayesian kernel methods”, *2006 International Joint Conference on Neural Networks Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada July 16-21, 2006*, 2006.

[65] D. G. Lowe, “Object recognition from local scale-invariant features”, *Proc. of the International Conference on Computer Vision, Corfu (Sept. 1999)*, 1999.

[66] O. Oguz, “The proportion of the face in younger adults in using the thumb rule of leonardo da vinci”, *Surgical and Radiological Anatomy*, 1996.

[67] F. Nainia, J. Mossb, and D. Gille, “The enigma of facial beauty :esthetics,proportions ,deformity and controversy”, *American Journal of Orthodontics and Dentofacial Orthopedics*, 2006.

[68] L. J. Chmielewski, A. Datta, R. Kozera, and K. Wojciechowski, “Computer vision and graphics”, *International Conference, ICCVG 2016, Warsaw*, 2016.

[69] R.Brunelli and T.Poggio, “Hyperbf networks for gender classification”, *DARPA Image Understanding Workshop*, 1992.

[70] S.Ashok, V.Subramania, and D.Marx, “Analysis of sexual dimorphismin human face”, *J.Vis.Commun*, 2008.

[71] Z. Xu, L. Lu, and Pengfei, “A hybrid approach to gender classification from face images”, *2008 19th International Conference on Pattern Recognition*, 2008.

[72] P.Viola and M. Jones, “Robust real-time face detection”, *International Journal of Computer Vision*, 2004.

[73] R.Lienhart, A.Kuranov, and V.Pisarevsky, “Empirical analysis of detection cascades of boosted classifiers for rapid object detection”, *Microprocessor Research Lab,Intel Labs*, 2002.

[74] A.Lapedriza, D.Masip, and J.Vitria, “Are external face features useful for automatic face classification ?”, *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPRâĂŹ05)*, 2005.

[75] V. W. Parate and P. Patel, “Pca, dct and dwt based face recognition system using random forest classifier”, *International Journal of Digital Application & Contemporary research*, 2015.

[76] L Mdakane and V. den Bergh, “Extended local binary pattern features for improving settlement type classification of quickbird images”, *IEEE Prasa*, 2012.

[77] A. Rosenfeld, *Multiresolution Image Processing and Analysis*, A. E. Rosenfeld, Ed. Springer Series in Information Sciences, 1982.

[78] A. R. Ardakany and M. Nicolescu, “Improving gender classification using an extended set of local binary patterns”, *Journal International Journal of Multimedia Data Engineering & Management*, 2014.

[79] Y. Fang and Z. Wang, “Improving lbp features for gender classification”, *2008 International Conference on Wavelet Analysis and Pattern Recognition*, 2008.

[80] E. Hadjidemetriou, M. Grossberg, and S. Nayar, “Histograms and their use for recognition”, *IEEE PAMI, 26(7):831âĂŞ847*, 2004.

[81] D. V.Jadhao and R. S.Holambe, “Feature extraction and dimensionality reduction using radon and fourier transforms with application to face recognition”, *Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on*, 2007.

[82] G. Kukharev, P. l, and ForczmaÂť nski, “Data dimensionality reduction for face recognition”, *Machine GRAPHICS & VISION vol. 2, no. 4, 1993, pp*, 1993.

[83] D.Gering, “Linear and nonlinear data dimensionality reduction”, 2002.

[84] S.Buchala, N.Davey, R.Frank, and T.Gale, “Dimensionality reduction of face images for gender classification”, *In Intelligent Systems, 2004. Proceedings. 2004 2nd International IEEE Conference (Vol. 1, pp. 88-93). IEEE.*, 2004.

[85] S. L and K. M, “Low-dimensional procedure for characterization of human faces”, *J. Optical Soc. Am. 4, 519âĂŞ524*, 1987.

[86] HuangWeilin, H. Y. Huang, Weilin, and H. Yin, “"linear and nonlinear dimensionality reduction for face recognition."”, *6th IEEE International Conference on. IEEE, 2009. Image Processing (ICIP),*, 2009.

[87] R. Vyas and D. S. Shah, “Comparision of pca and lda techniques for face recognition feature based extraction with accuracy enhancement”, *International Research Journal of Engineering and Technology (IRJET)*, 2017.

[88] L. L. Zhicheng, “A complete subspace analysis of linear discriminate analysis and its robust implementation”, 2016.

[89] R. Azahmer, “Real-time embedded age and gender classification in unconstrained video”, *ICVRP*, 2015.

[90] S.Goel, A.Verma, S.Goel, and K.Juneja, “Ica in image processing:a survey”, *2015 IEEE International Conference on Computational Intelligence & Communica*, 2015.

[91] S. S. S, K. N. B. Murthy, and S Natarajan, “Dimensionality reduction techniques for face recognition”, *INTECH open minds*, 2001.

[92] D. I. Processing, *R. C. Gonzalez and R. E. Woods*. Prentice Hall, 2008.

[93] A. Maina, B. Ugail, and D. Connah, “Automatic age and gender classification using supervised appearance model”, *Journal of Electronic Imaging 25(6), 061605 (NovâĹŢDec 2016)*, 2017.

[94] G. B Huang, Ramesh, M Berg.T, and Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments (vol. 1, no. 2, p. 3).”, University of Massachusetts, Amherst., Tech. Rep., 2007.

[95] H. Han and A. K. Jain, “Age, gender and race estimation from unconstrained face images”, 2014.

[96] D. Mery, Y. Zhao, and K. Bowyer, “On accuracy estimation and comparison of results in biometric research”, *Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on*, 2016.

[97] O.Bousquet and A .Elisseeff, “Stability and generalization”, *Journal of Machine Learning Research 2 (2002) 499-526*, 2001.

[98] M Budka and B Gabrys, “Density-preserving sampling: Robust and efficient alternative to cross-validation for error estimation”, *IEEE Transactions on Neural Network and Learning Systems*, 2013.

[99] L Refaeilzadeh P andTang and L. H, “Cross-validation. in encyclopedia of database systems”, *Springer US.*, 2009.

[100] D. M. Allen, “The relationship between variable selection and data augmentation and a method for prediction”, *Technometrics*, 1974.

[101] E. B and J. AM, “Estimating the error rate of a prediction rule: Improvement on cross-validation”, *Stat. Assoc., 78:316âĂŞ331,*, 1983.

[102] M.Ngan and P.Grother, “Face recognition vendor test (frvt) performance of automated gender classification algorithms”, 2015.

[103] A Majid, A.Khan, and A.Mirza, “Gender classification using discrete cosine transformation: A comparison of different classifiers”, *Proceedings IEEE INMIC*, 2003.

[104] J. A. Swets, R. M. Dawes, and J. Monahan, “Bener decisions through science”, *Scientic American*, 2000.

[105] M. Hu, Y. Zheng, and F. Ren, “Age estimation and gender classification of facial images based on local directional pattern”, *Cloud Computing and Intelligence Systems (CCIS), 2014 IEEE 3rd International Conference on*, 2014.

[106] H. M. Ebied, “Feature extraction using pca and kernel-pca for face recognition”, *The 8th International Conference on INFOrmatics and Systems (INFOS2012) Computational Intelligence and Multimedia Computing Track*, 2014.

[107] R. A. Vyas and S. Shah, “International research journal of engineering and technology (irjet)”, *Comparision of PCA and LDA Techniques for Face Recognition Feature Based Extraction With Accuracy Enhancement*, 2017.

[108] D. Cao, C. Chen, M. Piccirilli, D. Adjeroh, T. Bourlai, and A. Ross, “Can facial metrology predict gender?”, *Proc. of International Joint Conference on Biometrics (IJCB), (Washington DC, USA), October 2011.*, 2011.

[109] N. S. Lakshmiprabha, “Face image analysis using aam, gabor, lbp and wd features for gender, age, expression and ethnicity classification”, *Computer Vision and Pattern Recognition*, 2016.

[110] M. Zhou and H. Wei, “Facial feature extraction and selection by gabor wavelets and boosting”, *2009 2nd International Congress on Image and Signal Processing*, 2009.

[111] M. Saraswathi and D. K. Muthiah, “Improving gender recognition rate in face recognition system based on linear discriminant analysis technique”, *IJARCSMS*, 2015.

[112] R. J. Tom and dr.Arulkumaran, “Fingerprint based gender classification using 2d discrete wavelet transforms and principal component analysis”, *International Journal of Engineering Trends and Technology- Volume4Issue2- 2013*, 2013.

[113] S. Tamura, H Kawai, and H.Mitsumoto, “Male/female identification from 8x6 very low resolution face images by neural network”, *Pattern Recognition*, 1996.

[114] K. D. Kaur, P. Rai, and P. Khanna, “Gender classification system for half face images using multi manifold discriminant analysis”, *2017 7th International Conference on Cloud Computing, Data Science & Engineering âĂŞ Conflu*, 2017.

[115] K. H. Ghazali, M. F. Mansor, M. M. Mustafa, and A. Hussain, “Feature extraction technique using discrete wavelet transform for image classification”, *Other2007 5th Student Conference on Research and Development, SCORED*, 2007.

[116] P.Mohanaiah, P.Sathyanarayana, and L Gurukumar, “Image texture feature extraction using glcm approach”, 2011.

[117] G. Bradski and A. Kaehler, *Learning OpenCV:Computer vision with the OpenCv lbrary*. O’Reilly Media Inc, 2008.

[118] R. V. Reddy and K. Raju, “Grey level to rgb using ycbcr color space technique”, *International Journal of Computer Applications (0975 âĂŞ 8887)*, 2016.

[119] E. Hadjidemetriou and M. D. Grossberg, “Multiresolution histograms and their use for recognition”, *IEEE TRANSACTIONS ON PATTERN ANALYSIS*

AND MACHINE INTELLIGENCE, VOL. 26, NO. 7, JULY 2004, NO. 7, JULY 2004.

[120] P. Belhumeur, J. Hespanha, and D. Kriegman, “Eigenfaces versus fisherfaces: Recognition using class specific linear projection”, *IEEE Trans. Pattern Anal. Mach. Intell, 19 (1997), pp. 711âĂŞ720.*, 1997.

[121] A.Kaehler and G.Bradski, *Learning Opencv 3:Computer vision in C++ with the opencvLibrary*. O’Reilly Media Inc, 2016.

[122] G. Panis, A. Lanitis, N. Tsapatsoulis, and T. F. Cootes, “Overview of research on facial ageing using the fg-net ageing database”, *Institutiomn of Engineering and technology*, 2015.

[123] R.Azahmer, “Real-time embedded age and gender classification in unconstrained video”, *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW*, 2015.

[124] J. D.Silver T.Hubert, “Mastering chess and shogi by self-play with a general reinforcement learning algorithm”, *arXiv:1712.01815v1 [cs.AI]*, 2017.

[125] E.BrownU and E Perrett, “What gives a face its gender?”, *Perception 22, 829âĂŞ840*, 1993.

[126] G. Levi and T. Hassner, “Age and gender classification using convolutional neural networks”, *2015 IEEE*, 2015.