University of KwaZulu-Natal
Collage of Agriculture, Engineering And Science

UNIVERSITY OF KWAZULU-NATAL™
INYUVESI YAKWAZULU-NATALI

# Component-based Ethnicity Identification from Facial Images

*by:*

Aimée Booysens

210501411

*Supervisor:*

Prof. Serestina Viriri

*Submitted in fulfilment of the requirements for the degree of Master of Science in Computer Science, School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal*

January, 2017

# Declaration of Authorship

I, Aimée Booysens, declare that this thesis titled, 'Component-based Ethnicity Identification from Facial Images' and the work presented in it are my own.
I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

---

Aimée Booysens

# Declaration - Supervisor

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

---

Prof. Serestina Viriri

# Declaration - Publications

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

Publication 1:

A. **Booysens, S. Viriri**, "Component-based Ethnicity Identification from Facial Images', Computer Vision and Graphics, Springer LNCS, vol. 9972, pp. 293-303, 2016. DOI: 10.1007/978-3-319-46418-3_26, (ISBN 978-3-319-46417-6).

# Abstract

This dissertation presents an alternative analysis of component-based ethinic identification, which identifies the ethnicity from facial images of six different groups, using computer vision and image processing techniques. The six different ethnic groups identified are Asian, African, African American, Asian Middle East, Caucasian and Other. A number of different machine learning algorithms are investigated using a normalized feature vector; these are Decision Tree, Random Forest, Naïve Bayes, K-Nearest Neighbor Classification and Support Vector Machine (SVM). The results were: Naïve Bayes achieved 98.7% for African ethnicity identification rate; Naïve Bayes achieved 91.5% for African American ethnicity identification rate; Naïve Bayes achieved 90.3% for Asian ethnicity identification rate; K-Nearest Neighbor Classification achieved 95.6% for Asian Middle East ethnicity identification rate; Random Forest achieved 94.6% for Caucasian ethnicity identification rate; and Support Vector Machine Classification achieved a 95.6% for Other ethnicity identification rate. This research work achieved a total ethnicity identification rate of 84.6%.

# Preface

The research discussed in this dissertation was conducted at the University of KwaZulu-Natal, Durban, from January 2015 to December 2016 by Aimée Booysens under the supervision of Prof. Serestina Viriri.

# Acknowledgments

I want to express my profound gratitude to my supervisor, Prof. Serestina Viriri, for his words of encouragement, dedication and commitment throughout the Master's degree program.

I am deeply grateful to my Daddy, Mommy, family and family friends for their love and support, and for always being there for me.

My thanks go to some amazing people that whom I know - they know who they are - who constantly kept me normal and sane during the process. Their words and questions always served as an inspiration to keep me achieving at my best during the course of the Master's degree program.

I am extremely grateful to God for the abundant graces and blessings He has been showering on me.

# Contents

# List of Figures

# List of Tables

# Abbreviations

**AAM** Active Appearance Model

**KNN** K-Nearest Neighbor

**LBP** Local Binary Patterns

**LF** Local Feature analysis

**OGM** Oriented Gradient Map

**PCA** Principal Component Analysis

**RGB** Red, Green and Blue

**SVM** Support Vector Machine

**TRP** True Positive Rate

# Chapter 1

# General Introduction

## 1.1 Introduction

Human beings are identified in their daily life by their biometrics or security mechanisms. The security mechanisms that identify human beings are items such as keys, security discs or tags to obtain entrance into buildings, gates and homes. Other security mechanisms are passwords and pins to obtain money from bank accounts, logging on to computers, cell phones and tablets, and social media or social networking such as Gmail, Facebook and Twitter. These security mechanisms are unreliable as they can be lost or stolen. Hence, the use of security mechanisms that are based on biometrics is increasing as these these features are impossible to forge.

Due to the fact that biometrics are gaining importance in security and general identification methods, the question is whether the face is able to identify a persons ethnicity. In previous works, ethnicity is identified by using the whole facial region [9,36,54], however this research work entails research on identifying a persons ethnicity by using components of the face. The research question is, will ethnicity extraction from components be enhanced by using a variety of feature extraction techniques. The ethnicity groups that are identified in this research work are Asian, African, African American, Asian Middle East, Caucasian and Other which is an extension of the number of ethnicity groups used in previous works. Once the best feature extraction techniques are obtained, a general framework for ethnicity identification will be proposed.

## 1.2 Motivation and Applications

This section discusses the motivation and application of the research work. In the first part of this section, the motivation is discussed and the second part explains the applications of ethnicity identification.

### 1.2.1 Motivation

A framework for facial component-based ethnicity identification is the desired goal in this research work. Ethnicity identification is dependent on certain criteria for grouping pixels into their intensity values, gradient information or textures. The domain of the applications is to determine ethnicity grouping using the whole facial region. As a result of this, a maximum grouping of two to four ethnic groupings will occur.

Could there be room for improvement of the existing methods by using components of the facial region, for example the mouth, nose, eyes, chin, forehead and cheeks, and enhancing the ethnicity precision by using a variety of feature extraction techniques. This increase in ethnicity precision will mean extending the number of ethnicity groups.

### 1.2.2 Applications

There are several applications of ethnicity recognition. These are surveillance systems, security, image tagging and general identity verification.

- **Surveillance Systems** - A human-computer interaction surveillance system can be built to identify human attributes such as gender, age and ethnicity. There are many reasons that surveillance systems can be used; namely, terror-related crimes, law-enforcement and security [22].

- **Security** - Facial security is slowly replacing password logins on certain applications and computer systems. At Manchester University, researchers are working on creating consumer-focused facial recognition technology for security applications. A number of well-known companies are interested in integrating the security facial recognition technique into their products [56].

- **Image Tagging** - Facebook's automatic tagging feature uses facial and ethnicity recognition for users to suggest people they might want to tag in their photos. This has been seen to save people time when posting images on social media. Image tagging is available on applications such as Apple's iPhoto, Google's Picasa and Facebook [27].

- **General Identity Verification** - Ethnicity and facial indetification using facial images has general uses, including electoral registration, banking, electronic commerce, identifying newborns, national IDs, passports and employee IDs [23, 46, 62] .

## 1.3 Problem Statement

The human face can be split up into different components or regions. Several methods have been proposed in the literature for ethnicity identification of humans through facial images. These methods are used to obtain the colour of the facial region using hue and saturation values of an image. These methods use the whole face in order to ascertain the ethnicity. Obtaining the components of the facial region, for example the mouth, eyes, nose, forehead, cheeks and chin, can improve the extraction of the ethnicity from that image. A component image is shown in Figure 1.1 for the purpose of subsequent discussion.

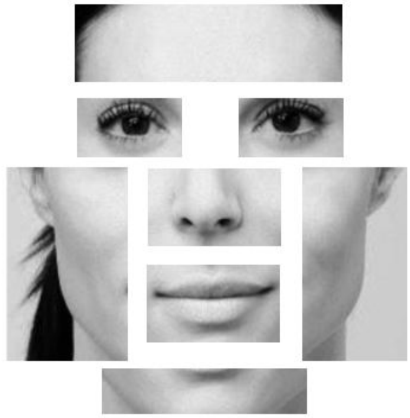


Figure 1.1: Extracted Facial Components

The methods proposed in literature classify the facial images into, at most, four ethnic groups [1, 3, 9, 32, 40, 49, 52, 54, 60]. These ethnic groups are classified as Asian or non-Asian [40, 49, 60], Myanmar or non-Myanmar [54] and East Asian, South Asian, White or Black [32]. However, this research work aims to expand the number of ethnic groups. Facial images are prone to variation in pose, illumination and expression [23, 46, 62]. Another issue observed with facial images is that they may also have sunglasses, glasses, hats and scarfs. These accessories make it difficult to ascertain the ethnicity of that image. These issues need to be resolved in order to ascertain the correct ethnicity [46, 62].

## 1.4 Dissertation Objectives

The objectives of this dissertation are to:

1. Identify human ethnicity using facial components.
2. Improve ethnicity identification accuracy rate using different feature extraction techniques.
3. Model a framework for facial component-based ethnicity identification.

## 1.5 Contributions of the Dissertation

In this dissertation, the main contributions are to improving the accuracy of characterising the ethnicity of facial images. According to the set objectives, the following are achieved:

- Automated identification of human facial components. This was achieved by the employment of pre-processing and segmentation techniques that automatically extract the facial components of the mouth, nose, chin, eyes, forehead and cheeks from the image. Many studies that were researched used the whole facial region to obtain the ethnicity of the facial image.

- Identification of ethnicity is automatically achieved using different feature extraction techniques. This is the main objective of this dissertation and was achieved by using a combination of Hu Moments, Zernike Moments, Gabor Filter, Haralick Texture Moments and Linear Binary Pattern feature extraction techniques in order obtain the ethnicity of the image.

- Lastly, from the above automations a model framework was proposed to improve ethnicity identification with component-based images. This framework, managed to improve ethnicity identification for a large number of ethnicity groups.

## 1.6 Dissertation Overview

The rest of the dissertation is structured as follows:

Chapter 2 presents a background study of facial and ethnicity recognition. In chapter 3, the components are outlined, feature extractions and classifications are extracted and basic system structures are presented. Chapter 4 describes the experiments and results. Lastly, chapter 5 concludes the dissertation and discusses future work.

# Chapter 2

# Background Study and Literature Review

## 2.1 Introduction

Ethnicity is a broadly defined topic of social science. One of its oldest definitions is that of Isaacs [30], Stack [51] and Geertz [17] who state that ethnicity is assumed at birth and is derived from the shared identity and structure of ones kin and clan in human society. Hence, ethnicity is a more fixed and permanent grouping than that of a social group. Whereas Hale [20] believes that the understanding of ethnicity is at the beginning stages and suggests that the definition given by Connor [12], Horowitz [26] and Shils [50] about ethnicity is the best definition. They state that it is an emotion educing word as it gives a person a sense of belonging and attachment to a particular kind of group. Green [18] defines ethnicity as a synonym for nation and race but proposes that ethnicity has three core elements; these are: common descent, history and homeland. This allows for a variety of interpretation which incorporates the pri-mordialism and constructivism.

Due to the variety of definitions of ethnicity, this research work has divided facial images into different ethnic groupings such as: Asian, African, African American, Asian Middle East, Caucasian and Other. This chapter will define methods of classifying facial recognition and ethnicity identification.

## 2.2 Feature Extractions and Classifications

### 2.2.1 Feature Extractions and Classifications to Facial Recognition

Many different techniques are used for facial recognition. The most common techniques that are used are: Eigenfaces, Local Feature Analysis, Elastic Graph Matching, Active Appearance Model and 3D Morphable Model. Eigenfaces [5,11,39,42,56] effectively represent pictures of faces using Principal Component Analysis (PCA). Different forms of Eigenfaces are used as a base for other face recognition. The Eigenfaces do not use the methods that involve normal human recognition but finding similarities between faces with minimal controlled environments achieves reasonable results.

Another technique is that of Local Feature analysis (LF) [21, 39] which is used for facial biometric technologies, such as eye analysis. These facial biometric technologies accommodate for changes in facial expressions and aging. Using facial images, Local Feature analysis extracts a set of geometric metrics and distances. The features used are eyes, mouth, nose, jaw line, eyebrows and cheeks. The performance of this technique is dependent on the environment and resolution of images.

Elastic Graph Matching Local Features [31,39] are extracted at certain locations on the face. The distances between these points are calculated and the more important points get higher weights than others. Elastic Graph Matching [39] is invariant to affine transformations and localised changes in facial expressions, after which the same images at different angles are then matched to the same point on the image. Each node is then put onto a graph which is called Elastic Bunch Graph Matching. This improves the recognition and this technique is more robust in finding the differences in posture and facial expressions.

On the other hand, statistical models such as Active Appearance Model (AAM) [39, 47] combine the shapes of the image with the appearance variations of a normalised frame. The images used in this model are in a gray level format in order to obtain the shape and appearance. The matching of the image to parts of another image involves finding the models parameters that have a minimal distance between the image and the part of the other image. This can be difficult as the large number of parameters used gives the model a robust nature. In Lu [40], AAM is used on the dataset in order to obtain the image in a vector format. Linear discrimination analysis is used to construct a subspace for the facial identification area. Facial recognition is then achieved by finding the best match between the image vector that had the AAM applied to it and the stored feature vectors in the database. Both theses vectors are then projected onto the discriminant subspace. The performance of the AAM had an 88% accuracy rate.

Another technique used was 3D Morphable Model [39, 63] which uses the face as a 3D area. It was seen that a 3D model of the face is a better representation as it can show facial variation, like pose and illumination. The 3D Morphable Model [39] encodes shape and texture as parameters which is achieved by using an algorithm that encodes the information from a single image. Due to the large variation in parameters that an image can have, generative image models are applied to the parameters. The goal is to separate the intrinsic model parameters from the extrinsic image parameters in order to make an identification. The separation of these parameters is obtained by simulating the process of image formation using 3D computer graphic technology. The similarity between the two images is defined and the points with the lowest similarity are used.

Lin et. al. [36, 41] used a technique for recognising gender, ethnicity and age using facial images. This technique combined Gabor filter, Adabost learning and SVM classification in order to do facial recognition. Facial features are extracted using Gabor filters and Adaboost learning and once features are extracted, SVM classifiers are applied in order to do facial recognition of soft feature biometrics. These are physical, behavioral technique or adhered human characteristics. The results that were achieved from this technique had high accuracy and good performance which was due to the fact that the application of the pre-processing step, the Gabor Filter, further improved the performance.

Futhermore, the technique of using a large set of low level features was developed by Kumar et al. [35], where they used the low level features to train the SMV classifier with RBF kernels in order to obtain facial features. Some low level features from original datasets that are related more to facial features were used as the input for the SVM classifier. This technique which uses low level features recorded reasonable facial recognition rates of 31.68%.

On the other hand, the techniques that were used by Zhang [62] in his experiments tried to improve facial recognition by using soft biometrics of the facial region. It was seen that soft biometrics cannot be treated independently without the use of an algorithm; in many cases soft biometrics are seen as redundant. The algorithm of SVM and Adaboost was used to encode the soft biometrics in order to do facial recognition. It was found that with few soft biometrics limits, the performance of these experiments increased by 11%.

Atharifard and Ghofrani [4, 44, 45] presented a robust component face detection algorithm based on the colour features. The algorithm used had huge benefits as it had a reduction in computational time . It was found that when applying component detectors on facial images it took a few minutes to detect faces. In order to apply the component detectors, extra pixels, including non-skin regions, were removed. Two models were used in order to detect eyes and mouth as the facial components. These models were colour space, which extracted the RBG colour from the image, and skin colour segmentation, using a Gaussian model in order to extract skin colour. It was presented that these two models performed well even if the eyes and mouths were obstructed or not detected. This proves that the models have the ability to localise missed or obstructed facial components and that the accuracy is higher than other techniques.

Ahnoen, Hadid and Pietikainen [3, 10, 34] discussed another technique to perform facial recognition namely, Local Binary Patterns (LBP) which takes into account both shape and textural information to represent facial image. This model obtained a 98% accuracy for facial recognition. This technique is achieved by taking a gray scale image and dividing it into smaller areas then applying an LBP 3 x 3 mask to the area. The middle pixel value is then compared to the pixel values around it; if the pixel value is lower than the middle pixel, then the original value is changed to 0; if the pixel value is higher then the original value, it is changed to 1. Then, reading from left to right, a vector is obtained with only 1's and 0's and the binary values obtained are extracted onto a histogram. Each smaller area is then plotted onto a histogram with its binary value and all these are then concatenated together. The image recognition is performed by using Nearest Neighbour Classifier and dissimilarity measure of Chi Squared applied to the concatenated histogram value.

Shah [49] presented an approach to facial recognition by converting the image into different colour spaces; these colour spaces were RGB, YcbCr and HLS. The image is then binarized and this binary image is obtained from erosion and dilation which are morphological operators. From the binary image, the region of interest is extracted and the mouth, eye and nose regions are mapped. Once the above pre-processing stages are completed, they are passed to the function linear SVM to classify the facial region.

Facial recognition has been tested and processed by many different authors. These authors use many different techniques and models in order to achieve facial recognition.

### 2.2.2 Feature Extractions and Classifications to Ethnicity

The most used techniques for ethnicity recognition are Support Vector Machine, Principal Components Analysis, Linear Discriminant Analysis, Hue, Saturation and Value and Linear Binary Pattern. Lu et al. [40] used the technique of Support Vector Machine (SVM) which is applied to ranges. These ranges are obtained by taking 3D models of facial images and normalizing them. These normalized images are then grouped into ranges of intensities. Lu et al. obtained an ethnicity accuracy rate of 98%. Two classes of ethnicity were classified: Asian and non- Asian.

Another technique was used by Yang and Ai [60] - that of Linear Binary Pattern (LBP) which tries to extract the demographic classification features which include ethnicity and race. Two classes of ethnicity were classified; namely, Asian and non-Asian. The images used are first normalized then LBP is applied. This entails taking each pixel as a threshold in which a 3 x 3 mask neighbourhood is transferred to an 8-bit binary code. This LBP 8-bit binary code is plotted onto a histogram in order to obtain local texture patches. Weak patches are removed by Chi Square which is done by finding the smallest difference between LBP histogram and the reference histogram. The real AdaBoost is applied to the stronger patches in order to obtain the correct ethnicity classification.

The technique of using Principal Component Analysis (PCA) for identification of gender, ethnicity, age and identification for facial images was presented by Buchala et al. [9]. Principal Component Analysis (PCA) is the linear transformation of 2D images. This is achieved by obtaining the linear and orthogonal function that accounts for the maximum possible variances in the data. The above is best achieved by estimating the Eigenvectors and Eigenvalues of the covariance matrix. The facial images used were cropped so that little or no hair is visible in the image region and align the eyes in the centre, then the resolution of the image region is lowered and the equalization histogram is applied in order to reduce the effect of the light. Thereafter, a single vector is obtained and PCA is applied to the vector to extract each property, as shown in Table 2.1. The accuracy results obtained for each property, gender, ethnicity, age and identity, are shown in Table 2.1. Buchala et al. [9] tried to classify three different classes of ethnicity, Caucasian, African and East Asian, and achieved an 81.67% ethnicity accuracy rate.

Table 2.1: The results achieved for Gender, Ethnicity, Age and Identity [9]

| **Property** | **Classification Percentage** |
|---|---|
| Gender | 86.43% |
| Ethnicity | 81.67% |
| Age | 91.5% |
| Identity- a | 68.7% |
| Identity - b | 90% |
| Identity - c | 100% |

Tin and Sein [54] also used the technique of Principal Component Analysis (PCA) to classify two classes of ethnicity, Myanmar and non-Myanmar. In order to distinguish between these two classes, the image is normalised and scores are obtained. These normalised scores then have PCA applied to them. The results of PCA classifier easily identified between inter-ethnic sensitive features which allowed for a higher ethnicity identification rate.

An attempt to conduct ethnicity and gender identification by Traiq, Hu and Huang [52] used silhouetted face profiles with computer vision techniques. The facial images used were between the ages of 18 and 30 years for both male and female. Facial images of males who had beards and moustaches were removed. For each facial image that was used, the shape context was obtained for it. The shape context is a discrete set of sample points from both the internal and external contours of the object. After obtaining the shape context, the shape distance is calculated, which uses a weighted sum between the points chosen and the measure of bended energy. Once this is calculated for all points chosen on the test profile, K-nearest

Neighbour is applied to the points in order to do classifications. On four different classes of ethnicity, Black, East and South East Asians, South Asians and White, this achieved a below average ethnicity accuracy rate.

On the other hand, Salah [46] presented a fusion of two techniques for ethnicity identification. The techniques that were combined were Haar Wavelet-based global features and uniform LBP-based Local Features. This was achieved by converting the image into grayscale, cropping and then resizing it to the appropriate size. Haar Wavelet transform is applied first to the extracted region, thereafter Local Features are extracted by conducting Local Binary Pattern (LBP). The LBP values are then divided into 4 x 2 equal sizes of horizontal and vertical blocks. The fusion is then applied at the feature-level and is the concatenation of the Haar Wavelet and the uniform LBP histogram. This is done after values of the histograms are normalized by using the division-by-norm normalization techniques. Thereafter, the K-nearest neigbor classifier is used to find four nearest subjects, according to Euclidean distance, between sample region and training region. Three different classes of ethnicity that were distinguished between were European, Oriental and African.

Another technique that was used was Craniofacial features (eyes, mouth and nose). Colour variance of the skin is extracted from the facial image in order to identify ethnicity. This technique was presented by Abdullah and Abood [1]. This was achieved by using images that were cropped into a particular size. Thereafter, they extract the sink colour and geometric features that are used to determine the ethnicity of the facial image. Then only the face of the image is extracted and the single scale retinex and multi-scale retinex algorithm is applied in order to adjust the illumination, enhance the light and contrast the input colour of the images. Once completed, five ratios are used to classify the ethnicity. These ratios are computed by detecting well defined features on the face, for example the red, blue and green colour and the nose, mouth and the divide between the eyes. Four different classes of ethnicity were distinguished. These were: White, Asian, African and Middle. This technique of using Craniofacial features and colour variance of the skin achieved on average 82% for ethnicity identification.

Ding et al. [13] discusses a technique that distinguishes between seven different classes of ethnicity. These were White, Asian, Hispanic, Asian Middle Eastern, Asian Southern, Black and African American and Unknown. The technique is a fusion of both the boosted local textures and shape features which are obtained from 3D face models. The 3D model is put through a number of pre-processing techniques which include removal of spikes and filling holes, before Oriented Gradient Map (OGM) is applied. OGM is a biological vision-based representation for 3D facial recognition as it was seen to have good performance and was insensitive to illumination and geometric transformations. OGM is calculated by obtaining gradients over a certain range of texture points then applying a normalisation of Gaussian kernel to remove any abrupt changes. A response vector is then obtained by collecting all values of the convolved gradient map at a given point; the response vector is then normalized to a unit norm vector. The classification technique that was used is the AdaBoost technique which is from the Boosting group and is defined as a strong linear classifier. A number of tests were performed in order to obtain ethnicity. Ding et al. achieved a 98.3% accuracy rate using the fussion of boosted local textures and shape features. This achieved a high accuracy rate as it used 3D models of the face which is computationally expensive to use.

Lyle et al. [28, 59] used another technique of computing periocular texture which is the small region of skin around the eye. Periocular texture is computed from grayscale image using Local Binary Patterns. A Support Vector Machine (SVM) classifier is then trained to classify the texture features. These experiments were done on the visible spectrum of the images. This technique achieved for gender and ethnicity classification an accuracy rate of 93% and 91% with cross validation. This showed that the fusion of periocular data from both eyes with Local Binary Patterns over the face region achieved high accuracy for gender and ethnicity.

Another technique that was used by Du et al [15] is the multi-level fusion scheme for ethnicity identification. The two techniques that were fused were Local Binary Patterns (LBP) and HSV Binning in order to extract the lower level features. First the cheek area is cropped from the colour facial image that is inputted, then HSV binning is applied. This involves extracting a histogram of the image in which the hue, saturation and value is obtained and concatenated into multiple bins. Once HSV binning is applied, the image is converted to grayscale which is obtained by applying LBP, which is a 3 x 3 pixel window that is applied to each pixel in the facial image. The LBP image is divided into 4 x 2 blocks in which the histogram feature vector is extracted; this is the concatenation of colour and texture feature. Due to the uneven mixture of colour and texture features, the feature vector needs to have a normalisation technique applied; this was the division by range method. Once the feature vector is normalised, classification by the K-nearest neighbour and Support Vector Machine is applied. The ethnicity classes of European, Oriental and African were classified. The results of the multi-level fusion scheme achieved above average results for ethnicity identification.

On the other hand, Todric [55] used the technique of using a 3D mesh of the face in order to obtain the gender and race or ethnicity of the image. This was done by not obtaining the texture or the photographic information of a facial image. As there is no photographic information being obtained, skin tone was ignored. This paper tried to present the power that the facial structure has a means of obtaining gender, race or ethnicity. The images used are decomposed using both Haar wavelet decomposition and the Steerable Pyramid transform. Once the metadata from the decomposition is obtained, the distance function is obtained from the weighted sum of the two metadata. The classification technique that was used is K-Nearest Neighbors which gave better results for gender than that of race or ethnicity.

Table 2.2 presents the ethnicities that was detected and the database, features and classifier that were used in each article discussed in this section.

Table 2.2: Break down of the systems for previous works

| **Reference Article** | **Ethnicity Groups** | **Feature Extraction Techniques** | **Classification Techniques** |
|---|---|---|---|
| Abdullah et al. [1] | White, Asian, African and Middle | Craniofacial Features obtianed | Craniofacial Features obtianed |
| Buchala et al. [9] | Caucasian, African and East Asian | PCA | PCA |
| Ding et al. [13] | White, Asian, Hispanic, Asian Middle Eastern, Asian Southern, Black or African American and Unknown | OGM | AdaBoost |
| Du et al. [15] | European, Oriental and African | LBP and HSV | KNN and SVM |
| Lu et al. [40] | Asian, non-Asian | Grouped images with same intenstiy values | SVM |
| Lyle [59] | Asian , White, Indian | LBP | SVM |
| Salah [46] | European, Oriental and African | LBP | Haar Wavelet |
| Tin and Sein [54] | Myanmar and non-Myanmar | PCA | PCA |
| Todric et al. [55] | White and Asian | Haar Wavelet and Steerable Pyramid | KNN |
| Tariq, Hu and Huang [52] | Black, East and South Asian, South Asian and White | Shape context | KNN |
| Yang and Ai [60] | Asian, non-Asian | LBP | AdaBoost |

## 2.3 Component Analysis for Facial Recognition

The component-based facial recognition approach entails the extraction of parts of the face in order to do facial recognition. This approach is not sensitive to image variations, like facial rotations. The problem that is faced with this approach is the automatic extraction and validation of the facial components; to do this without any human interaction will cause this approach to be robust.

Heisele et al. [22–25, 29] proposed a method to automatically detect and recognise facial components. Firstly an object window of a certain size is slid over the input image. Thereafter, 14 points of reference are selected in the object window based on their 3D correspondences from a morphable model. The algorithm then draws small rectangles around the selected reference points. The detection of facial components is carried out by finding the maximum output of the smaller rectangular area with each component being classified using linear Support Vector Machine (SVM). The coordinates of the position of the maximum output of each component classifier is recorded with the position. The Haar transform is applied on the frontal faces to obtain the feature vectors.

The drawback of this approach is that it needs human intervention to detect and extract the components; this is done by using geometric consideration and assumption on the location on the facial components. This approach is preformed on only frontal view images. It was seen that using a large number of techniques to ascertain the location of facial components leads to high computational time.

## 2.4 Drawbacks of Current Methods and Contributions

Gutta [19], presented that a persons face will exhibit variations, such as the eyes being closed and the face smiling. This will affect the way the computer vision system recognizes gender, ethnicity and age in the image. The accessories worn by the subject, such as eye glasses, hat, scarf or sun glasses affect the accuracy of facial recognitions. Illumination, lighting and image quality (like blurring, noise and resolution) change the subject and affect recognition. Images where the orientation of the subjects head changes also affect the accuracy of recognition. Many studies [9, 54, 55] have shown that the the whole face is used when classifying a persons ethnicity. Lastly, it was observed that in most previous research works in ethnicity two ethnic groups were conducted.

## 2.5 Conclusion

This chapter included a comprehensive description of ethnicity and facial recognition. Information was reviewed on component-based extraction techniques which was discussed. Pathways of our research were presented with identification of the drawbacks in current methods and contributions.

In the next chapter, the algorithms and feature extraction techniques for ethnicity identification are defined and described.

# Chapter 3

# Methods and Techniques

## 3.1 Introduction

In order to obtain the ethnic identification of the facail image different techniques and methods are used in order to obtain the correct ethnicitiy for the image. In this chapter, the methods and techniques for ethnicity identification used in this dissertation are presented.

The structure of the system is introduced in section 3.2; thereafter, the setup of each system is explained and discussed. In section 3.3, facial image pre-processing is explained and in setion 3.4 the detection of the facial region is examined. The process by which the components are extracted from the facial region is explained in section 3.5. In sections 3.6 and 3.7, the Feature Extraction techniques are described and the feature fusion and normalization are examined. In seaction 3.8, the classification techniques are examined and explained.

## 3.2 Structure of the System

The general system framework analyses facial images and obtains the ethnicity of a facial image. The acquired facial images are pre-processed by converting them to greyscale then reducing each image to a size of 300 by 300 pixels and they are then normalised. Thereafter, the system obtains the different components of the facial region these components that are extracted are the left and right eye, nose, chin, mouth, forehead and left and right cheek. Feature extraction methods are applied to each component that is extracted. Once the best feature extraction method is found for each component, a feature vector is then extracted for facial region and then normalised. The normalised feature vector is then used in the classification method to obtain the facial images ethnicity. The proposed system is depicted in Figure 3.1.

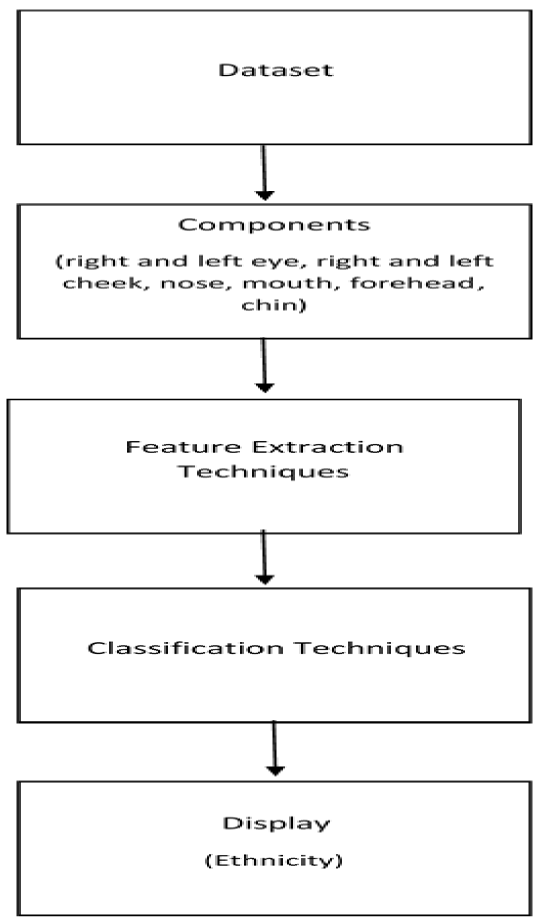



Figure 3.1: The General Framework of the component-based ethnicity identification system

## 3.3 Facial Image Pre-processing

The facial images are pre-processed by applying a greyscale conversion method to the facial image, resizing the facial image and then normalizing the image. The first process that was applied was the resizing of the facial image. This meant that the original image, shown in Figure 3.2 (a), was cropped to a size of 300 by 300 pixels. The resizing of the facial image was done to show a facial image with little or no hair as this can affect the way we obtain a facial images ethnicity. An example of the resized image is shown in Figure 3.2 (b).

The resized colour image is then converted to greyscale using the openCV [8] colour to greyscale method. This is done by taking each pixel of the image and obtaining that pixels Red, Green and Blue (RGB) values. These RGB values have a constant applied to them. As shown in equation (3.1), it is used to obtain the greyscale value of that pixel. The reason for converting the image to greyscale is to make each image uniform when it comes to comparisonss of the images.

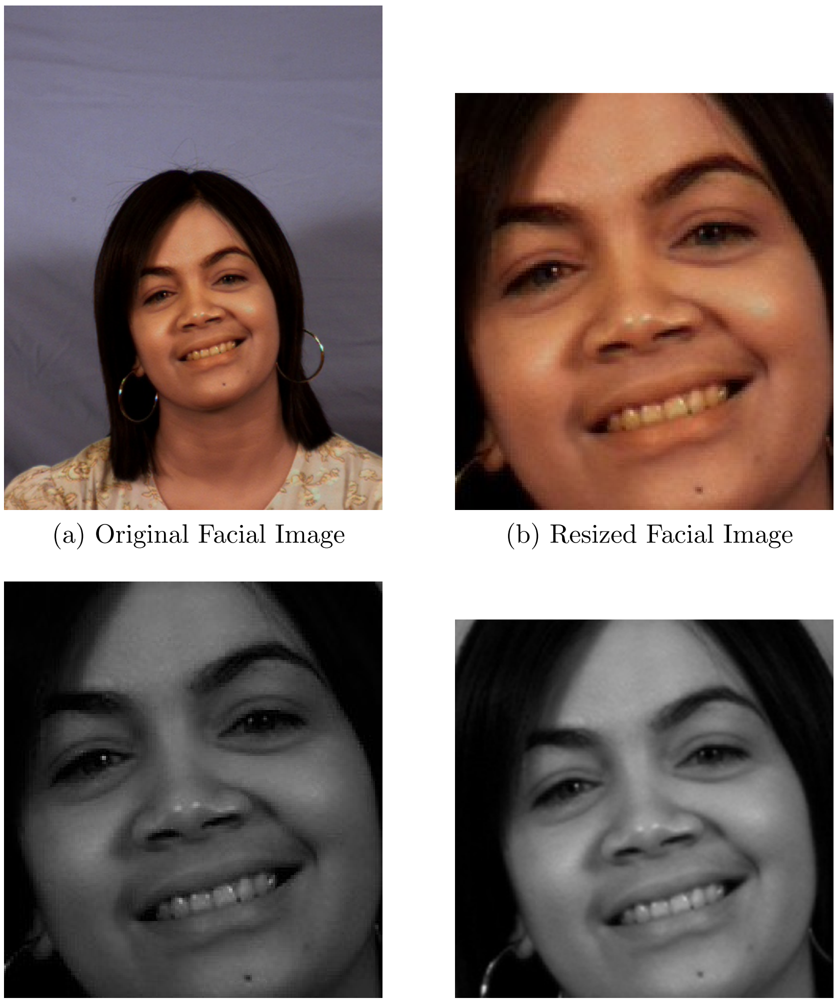
$$RGBtoGrey : Y \leftarrow 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B \tag{3.1}$$

Where $Y$ is the new greyscale pixel value, $R$ is the red pixel value, $B$ is the blue pixel value and $G$ is the green pixel value. An example of the converted greyscale image is shown in Figure 3.2 (c).

After the facial image has been converted to greyscale, normalization is applied to the image. Normalisation is the process of changing the range of pixels intensity values in order to equalize the image. This is obtained by remapping the image such that the maximum intensity value is 255. This is calculated using equation (3.2).

$$H'(i) = \sum_{0 \leq j \leq i} H(j) \tag{3.2}$$

Where $H'(i)$ is the remapped intensity value, $H(j)$ is the original intensity value, and $i$ and $j$ are the maximum $i$ and minimum $j$ intensity value. An example of the histogram image is in Figure 3.2 (d).



(a) Original Facial Image (b) Resized Facial Image

(c) Greyscale Facial Image (d) Normalized Facial Image

Figure 3.2: Facial Image Pre-Processed

## 3.4 Detection of Regions

The pre-processed facial image then has the facial region extracted. This is to include only the skin part of the face and has little or no hair and glasses and sunglasses. The extraction method that is used is Haar Wavelet Transformation. The Haar Wavelet Transformation [43] method that is used in this research work is stated in the matrix below (3.3).

$$HT^{n}(f) = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix} \tag{3.3}$$

Where $n$ is the number of element in the function, $HT$ is the Haar Wavelet Transformation and $f$ the function.

Haar Wavelet Transformation is a generalization of a two dimensional matrix and applying a one dimension transformation to the matrix. The two dimensional matrix is then obtained from the facial images. Once this two dimensional matrix is obtained, a one dimensional Haar Wavelet Transformation is applied to each pixel value in the two dimension matrix row. Once this matrix is obtained, it is used as a coefficient value for each row of the two dimensional matrix. Thereafter, the above matrix is applied to each column of the new two dimensional matrix. After the application of the Haar Wavelet Transformation, the facial area of the image is left.

## 3.5 Components of the Facial Region

In this research work, analysis is through components being extracted from the facial region of the image. These components are the left and right eyes, mouth, nose, chin, forehead and left and right cheek and are extracted using Haar Cascade. Haar Cascade is built in to openCV [8] in which different masks are used that identify different facial components. These masks check and contrast the values between the adjacent pixels group to determine if they need to be included or not. If the masks are not found in the pixels, they are then moved to the next set of pixel values. If the image is the same or similar then the Haar Cascade-like feature is found. Examples of all the components that are extracted can be seen in Figure 3.3.

(a) Normalized Facial Image

Figure 3.3: Facial Components

## 3.6 Feature Extraction Techniques

A feature vector was extracted from facial component. Analysis of each component was done using different textural, structural and geometric feature extraction techniques in order to obtain the most accurate and distinctive feature extraction techniques for that component. The textural feature extraction used were 7 Hu Moments, Haralick texture moments and Gabor Filter. The geometric and structural feature extraction techniques were Zernike moments and Local Binary Pattern. Once each component feature vector is obtained, they are fused together and normalised.

### 3.6.1 7 Hu Moments

The 7 Hu Moments [53] is a textural feature extraction technique which obtains a set of invariant moments that are characterized regardless of an images scale, position, size and orientation. These are computed by normalizing the central moment through the order of 3 in conjunction with the central moments. The 7 Hu Moments are define with the below set of equations (3.4).

$$
\begin{aligned}
I_1 &= n_{20} + n_{02} \\
I_2 &= (n_{20} - n_{02})^2 + 4n_{11}^2 \\
I_3 &= (n_{30} - 3n_{12})^2 + (3n_{21} - n_{03})^2 \\
I_4 &= (n_{30} - n_{12})^2 + (n_{21} + n_{03})^2 \\
I_5 &= (n_{30} - 3n_{12})(n_{30} + n_{12})[(n_{30} + n_{12})^2 \\
&\quad - 3(n_{21} + n_{03})^2](3n_{21} - n_{03})(n_{21} + n_{03}) \\
&\quad [3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] \\
I_6 &= (n_{20} - n_{02})[(n_{30} + n_{21})^2 - (n_{21} + n_{03})^2 + \\
&\quad 4n_{11}(n_{30} + n_{12})(n_{21} + n_{03})] \\
I_7 &= (3n_{21} - n_{03})(n_{30} + n_{12})[(n_{30} + n_{12})^2 \\
&\quad - 3(n_{21} + n_{03})^2] + (n_{30} - 3n_{12})(n_{21} + n_{03}) \\
&\quad [3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2]
\end{aligned}
\tag{3.4}
$$

where $n_{pq}$ and $n_{qp}$ are normalized central moments of the order $(p+q)$.

### 3.6.2 Zernike Moments

Zernike Moments are used to overcome the redundancy that other geometric moments obtain [53]. Zernike Moments are a class of orthogonal moments which are rotational-invariant and effective in image representation. Zernike Moments are a set of complex, orthogonal polynomials defined as the interior of the unit circle. The general form of Zernike Moments is defined in equation (3.5):

$$Z_{nm}(x,y) = Z_{nm}(p,\theta) = R_{nm}(p)\epsilon^{jm\theta} \tag{3.5}$$

where $x$, $y$, $p$ and $\theta$ correspond to Cartesian and Polar coordinates respectively, $n \in Z^{+}$ and $m \in Z$, constrained to $n-m$ even, $m \leq n$.

$$R_{nm}(p) = \sum_{k=0}^{\frac{n-m}{2}} \frac{(-1)^k (n-k)!}{k!(\frac{n+m}{2}-k)!(\frac{n-m}{2}-k)!} p^{n-2k} \tag{3.6}$$

Where $R_{nm}(p)$ is a radial polynomial and $k$ is the order.

### 3.6.3 Haralick Texture Moments

Haralick texture moments are textural features extraction technique that can be used to analyze the spatial distribution of the image's textural features [53] with different spatial positions and angles. Four most commonly used Haralick texture moments are used in this disatation as they defined the texture of an image in the best possible way as they are concerned with the properties of the individual pixel in relation to the surrounding pixels. The Haralick texture moments are sensitive to the choice of direction and supply an indication of the dominant values. The Haralick texture moments that are computed are: Energy, Entropy, Correlation and Homogeneity.

### Entropy

Entropy is the reflection of the disorder and complexity of the images texture. Entropy is defined with equation (3.7):

$$Entropy = \sum_{ij} \hat{f}(i,j) \log \hat{f}(i,j) \tag{3.7}$$

where $\hat{f}(i,j)$ is the $[i,j]$ entry if the grey level value of image matrix and $i$ and $j$ are points on the image matrix.

### Energy

Energy is the measure of uniformity and is computed using equation (3.8). This calculation is the opposite of the Entropy calculation as the Energy calculation has its main concentration on the texture of the image.

$$Energy = \sum_{ij} \hat{f}(i,j)^2 \tag{3.8}$$

where $\hat{f}(i,j)$ is the $[i,j]$ entry of the grey level value of the image matrix and $i$ and $j$ are points on the image matrix.

### Homogeneity

Homogeneity is the reflection of the uniformity of the images texture. This also represents the scale in which the texture of the image changes. An image that has high homogeneity has little or no changes between the regions textures. This can be defined using equation (3.9):

$$Homogeneity = \sum_{i} \sum_{j} \frac{1}{1+(i-j)^2} \hat{f}_{i,j} \tag{3.9}$$

where $\hat{f}(i,j)$ is the $[i,j]$ entry of the grey level value of the image matrix and $i$ and $j$ are points on the image matrix.

**Correlation**

Correlation is a measure of how correlated the pixel is to the neighbouring pixels. Correlation is measured in a range from -1 to 1 which defines whether the value of the image is perfectly correlated either negatively or positively. This is described in equations (3.10):

$$Correlation = \sum_{ij} \frac{(i-\mu_i)(j-\mu_j)\hat{f}(i,j)}{\sigma_i \sigma_j} \tag{3.10}$$

in which $\mu_j$, $\mu_i$, $\sigma_i$ and $\sigma_j$ are described as:

$$\mu_i = \sum_{i=1}^{n}\sum_{j=1}^{n} i\hat{f}(i,j) \qquad \mu_j = \sum_{i=1}^{n}\sum_{j=1}^{n} j\hat{f}(i,j) \tag{3.11}$$

$$\sigma_i = \sqrt{\textstyle\sum_{i=1}^{n}\sum_{j=1}^{n}(i-\mu_i)^2\hat{f}(i,j)} \qquad \sigma_j = \sqrt{\textstyle\sum_{i=1}^{n}\sum_{j=1}^{n}(i-\mu_j)^2\hat{f}(i,j)} \tag{3.12}$$

where $\hat{f}(i,j)$ is the $[i,j]$ entry of the grey level value of the image matrix and $i$ and $j$ are points on the image matrix.

### 3.6.4 Gabor Filter

A Gabor Filter [6]an orientation-sensitive filter meaning that the image needs to be in a certain orientation before this filter will work. It also best used for edge and texture analysis of an image. The filter works by applying a Gausian Modular onto a particular image which allows the image to be viewed as a sinusoidal plane with a particular frequency and orientation. Gabor Filter can be calculated using equation(3.13):

$$f(x,y,w,\Theta,\sigma_x,\sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\left[\frac{-1}{2}\left(\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2\right)\right] + jw(xcos\Theta + ysin\theta) \tag{3.13}$$

where $\sigma$ is spatial spread, $w$ is the frquency and $\theta$ is the orientation.

### 3.6.5 Local Binary Pattern (LBP)

Local Binary Pattern [46] is a non-parametric descriptor that efficiently summarizes the structure of the image. Local Binary Pattern is considered to be tolerant of monotonic illumination changes and simple to compute. It is a texture analysis method but can also define local structures. Local Binary Pattern is computed by obtaining a decimal number for each pixel. Each pixel is then compared with the eight neighbouring pixels to the central pixel. The comparison is done by subtracting the central pixel from the neighbouring pixel, shown in equation (3.14). If the obtained value is negative then the pixel is assigned a 0; if a positive value is obtained, then that pixel is given a 1. The binary value is then read in a clockwise direction starting in the top left hand corner. The binary value that is obtained from each of theses pixels is then converted to a decimal value. This decimal value that is obtained is then the pixel value, this value is between 0 and 255 pixel intensity. This can be calculated with the below equation (3.14)

$$LBP_{P,R}(x_c,y_c) = \sum_{P=0}^{P-1} s(i_p - i_c)2^P \tag{3.14}$$

$$s(x) = \begin{cases} 1 \text{ if } x \geq 0 \\ \\ 0 \text{ if } x < 0 \end{cases}$$

where $i_c$ and $i_p$ are gray-level values of a central pixel, $P$ surrounding pixels are the circle neighbours with radius of $R$ and $s(x)$ is the defining 0 and 1 values.

## 3.7 Feature Fusion and Normalization

Feature fusion [16] is the combination of all the feature vectors obtained in the feature vector process for each component of the same image. This then generates a single feature vector which has various feature vectors. This single vector needs to normalized [16]; the normalized technique that is used is the weighted sum rule (3.15) which ensures that the mean and variance of the feature vector are similar. In doing this, it allows all the images feature vectors to be comparable and removes outliers in the feature vector.

$$\sum_{i=1}^{N} \alpha_i P(\omega_j|x_1, x_2) = max_{k=1,2} \sum_{i=1}^{N} \alpha_i P(\omega_j|x_1, x_2), j = 1, 2 \tag{3.15}$$

Where $\omega_j, j = 1, 2$ the orginal index and the compared index, $N$ is the total number of matching results, $P(\omega_j|x_1, x_2)$ represents the posteriori probability. $\alpha_i$ is the weight assignto the classifier $i$

## 3.8 Classification

Classification involves taking the feature vectors extracted from the image and using them to automatically classify an images ethnicity. This is done by using different machine learning algorithms. In this research work, a number of supervised and unsupervised machine learning algorithms were used to obtain the images ethnicity. The supervised machine learning algorithm involves training the feature vector in order to compare an unknown feature vector with the trained data. The supervised machine learning algorithms used were Decision Tree, Nave Bayes, K-Nearest Neighbor (KNN) and Support Vector Machine (SVM). The unsupervised machine learning algorithm has the task of trying to define structures from an unknown data source. The unsupervised machine learning algorithms used were Random Forest and K-Means.

### 3.8.1 Decision Tree

Decision Tree [14, 38] is a hierarchically based classifier which compares data using a range of defined features. The features that are defined in this research work will be the classes of ethnicity. The advantages of the Decision Tree classifier is that it has a lower computation time than other machine learning algorithms and also avoids statistical errors. Decision Tree uses recursive partitioning to separate the dataset by finding rules to split and define the data. Using Entropy, equations (3.16) and (3.17), to calculate if the data will make a difference to the variable. If the Entropy is 0 then the variable is considered in that rule, otherwise a new rule is defined to incorporate the data.

$$H(D) = -\sum_{i=1}^{k} P(C_i|D) \log_k(P(C_i|D)) \tag{3.16}$$

where the entropy of a sample $D$ with respect to the target variable of $k$ possible classes $C_i$.

$$P(C_i|D) = \frac{numberofcorrectobservationforthatclass}{totalobservationforthatclass} \tag{3.17}$$

where the probability of class $C_i$ in $D$ is obtained directly from the dataset.

### 3.8.2 Naïve Bayes

The Naïve Bayes [14, 38] algorithm is based on conditional probability. It uses Bayes Theorem to calculate the probability of the class occurring given the probability that another event occurred. The advantages of Naïve Bayes is that it is fast, can be scalable and can be used for both binary class and multi-class classification problems. Naïve Bayes is defined in equation (3.18).

$$P(x_1, ......., x_n|y) = \frac{\prod_{i=1}^{n} P(y)P(x_n|y)}{P(x_1, ......., x_n)} \tag{3.18}$$

where case $y$ is a class value, attributes are $x_1, ........, x_n$ and $n$ is the sample size.

### 3.8.3 K-Nearest Neighbor (KNN)

K-Nearest Neighbour [14] is an algorithm that stores all the cases and classifies the new case based on the similarity measure. The K-Nearest Neighbour method is used both in statistical estimation and pattern recognition. The whole dataset is classified into either the training or the testing sample data. From the training sample point, the distance is calculated using the Euclidean Distance. This equation is shown in equation (3.19). If the result is less than the neighbours around that data, then it is considered the neighbour.

$$d = \sqrt{\sum_{i=1}^{n} (x_i - q_i)^2} \tag{3.19}$$

where $n$ is the size of the data, $x_i$ is an element in the dataset and $q_i$ is a central point.

### 3.8.4 Support Vector Machine (SVM)

The Support Vector Machine [33] is a trained algorithm for learning classification and regression rules from data. Support Vector Machine is based on structural risk minimisation and is related to regularisation theory. The implementation of the Support Vector Machine is a mathematical programing and kernel function (3.20). The parameters are found by solving a quadratic programing problem with linear equality and inequality constraints. Using the kernel function allows for flexibility and can search for a wide variety in the dataset. The defined algorithm searches for an optimal separating surface, known as the hyperplane. All the data is then separated using the hyperplane. If there are too many outliers using the calculated hyperplane, a new hyperplane is calculated until the simplest hyperplane is formulated. Where there exists a pair $(w, b)$ such that:

$$w^T x_i + b \geq 1, forall x_i \varepsilon P \\ w^T x_i + b < 1, forall x_i \varepsilon N \tag{3.20}$$

with the decision rule of the below in equation (3.21)

$$f_{w,b}(x) = sgn(x^T x + b) \\ y_i(w^T x_i + b) \geq 1, forall x_i \varepsilon P \cup N \tag{3.21}$$

where $w$ is the weighted vector and $b$ is the bias of the gievn points $P$ and $N$.

### 3.8.5 Random Forest

Random Forest [33] is an unsupervised machine learning algorithm. It is a group of unpruned classification or regression trees. The trees that are created are made from randomly selected samples of the training database. In order to create the tree, random features are selected in the beginning and are thereafter changed according to the presence or absence of that data point. Every tree is grown to the largest possible extent with no pruning; examples of these trees can be seen in Figure 3.4 below. With the use of Random Forest, the trees that are created are more robust to noise.

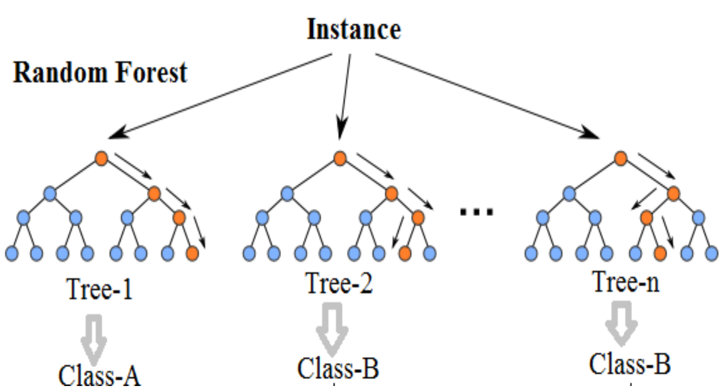



Figure 3.4: Random Forest representation of trees

### 3.8.6 K-Means

The K-Means [14, 33] algorithm is an unsupervised machine learning algorithm that solves the problem of clustering. The main idea is to have a certain number of clusters or classes. Thereafter, each data point is taken and associated to the nearest cluster. Each time a data point is calculated (3.22), the cluster recalculates the central cluster. This recalculation will either move the central clusters closer together or further apart. This continually recalculates the central cluster until no more changes can be made with the dataset. The advantages of K-Means is that it is fast, robust and easy to understand.

$$v_i = (1/c_i) \sum_{c_i}^{j=1} x_i \tag{3.22}$$

where $x$ be the set of data points and $v$ be the set of centers. $c_i$ represents the number of data points in the $i^{th}$ cluster.

## 3.9 Conclusion

In this chapter, the different methods and techniques used in this research work have been discussed, including the formulation of the ethnicity identification framework. This framework included feature extraction techniques are classified to obtain the resulting ethnicity. In the next chapter, the results and discussions of this research work are presented.

# Chapter 4

# Results and Discussions

## 4.1 Introduction

In this chapter, the performance of the facial images are analysed with the different classification techniques. The results achieved in this research project are also shown and discussed.

The programming environment is explained in section 4.2 and the dataset is examined in section 4.3. Thereafter, the Feature Extraction Technique is discussed in section 4.4 followed by the classification techniques for ethnic identification in section 4.5.

## 4.2 Programming Environment

The system was implemented and run with Intel Core(TM) i7-4770S @ 3.10 Ghz and 8.00GB RAM. The system was implemented in Java using different forms of plugins: ImageJ [48], openCV [8] and Catalono [2]. All these plugins are in the public domain and are used for Java image processing. They have also developed separate image processing operations, some of which are region of interest detection, bounding box drawing, Gabor Filters, Zernike Moments and many more. The plugins were used in sequence such that the output of one image will serve as the input to another image or plugin. A statistical Java package of Javaml [2] was used for the Support Vector Machine, Decision Tree, Naïve Bayes, K-Nearest Neighbor, Random Forest and K-Means classification.

## 4.3 Datasets

This investigation used a union of four different facial image databases: Yale [5], ORL [47], FERET [45] and MUCT [41]. The total dataset contained 1304 facial images and 900 subjects. These subjects were divided into six different ethnic groups. These ethnic groups: Asian, African, African American, Asian Middle East, Caucasian and Other.

Each ethnic group was composed with the following datasets: the Asian dataset was composed of Yale and FERET; the African dataset was composed of MUCT; the African American dataset was composed of Yale, ORL and FERET; the Caucasian dataset was composed of Yale, ORL, FERET and MUCT; the Asian Middle East dataset was composed of ORL, Yale, FERET and MUCT; and the Other dataset was composed of FERET.

The breakdown of the datasets and facial images used for each ethnic group is displayed in table 4.1.

Table 4.1: Breakdown of the dataset used

| **Ethnic Group** | **Dataset** | | | | |
|---|---|---|---|---|---|
| | **Yale [5]** | **ORL [47]** | **FERET [45]** | **MUCT [41]** | **Total** |
| African | - | - | - | 91 | **91** |
| African American | 1 | 1 | 77 | - | **79** |
| Asian | 8 | - | 171 | - | **179** |
| Asian Middle East | 6 | - | 53 | 46 | **105** |
| Caucasian | 17 | 39 | 619 | 102 | **777** |
| Other | - | - | 73 | - | **73** |
| **Total** | **32** | **40** | **993** | **239** | **1304** |

## 4.4 Composition of the Extraction Feature

Analysis was done to ascertain which feature extraction technique would achieve the most accurate True Positive Rate (TRP) for ethnic identification. True Positive Rate (TRP) is calculated using equation (4.1) and the False Positive Rate (FRP) is calculated using equation (4.2). The feature extraction techniques used in this study are: 7 Hu moments; Zernike Moments; Linear Binary Pattern; Gabor Filter; and

Haralick texture moments. In order to test that the feature extraction techniques were correct in classifying the components K-Nearest Neighbour classifier was used.

$$TPR = \frac{TP}{TP + FN} \tag{4.1}$$

Where $TRP$ is the True Positive Rate, $TP$ is the True Positive and $FN$ is the False Negative.

$$FPR = \frac{FP}{FP + TN} \tag{4.2}$$

Where $FRP$ is the False Positive Rate, $FP$ is the False Positive and $TN$ is the True Negative.

### Zernike Moments

Zernike Moments is a geometric feature extraction technique which is embedded in the Catalano [2] toolbox which was used to implement the Zernike Moments. In this research work Zernike Moments Feature Extraction Technique was applied to the same facial images that were used for the training of the other feature extraction techniques. The dataset was used as training, validation and testing for the Zernike Moments. The results of the experiment are shown in Table 4.2:

Table 4.2: Results achived by Zernike Moments

| **Facial Component** | **True Positive Rate** | **False Positive Rate** | **True Positive %** | **False Positive %** |
|---|---|---|---|---|
| **Nose** | 1045 | 259 | 80.10% | 19.90% |
| **Left Eye** | 997 | 307 | 76.40% | 23.60% |
| **Right Eye** | 996 | 308 | 76.30% | 23.70% |
| **Mouth** | 941 | 363 | 72.10% | 27.90% |
| **Forehead** | 1037 | 267 | 79.50% | 20.50% |
| **Chin** | 947 | 357 | 72.60% | 27.40% |
| **Left Cheek** | 686 | 618 | 52.60% | 47.40% |
| **Right Cheek** | 933 | 371 | 71.50% | 28.50% |

Zernike Moments was tested using the complete dataset which consisted of 1304 images. Each image all components are extracted - eyes, nose, chin, forehead, mouth and cheeks. From the results in Table 4.2, 363 instances of the mouth, 259 instances of the nose, 307 instances of the left eyes, 308 instances of the right eyes, 267 instances of the forehead, 618 instances of the left cheek and 371 instances of the right cheek were misclassified. Zernike Moments has an average True Positive classification rate of 72.6%.

### Gabor Filter

The Gabor Filter is a Textural Feature Extraction Technique which is embedded in the Catalano [2] toolbox and was used to implement the Gabor Filter. In this research work Gabor Filter Feature Extraction Technique was applied to the same facial images used for the training of the other feature extraction techniques. The dataset was used as training, validation and testing for the Gabor Filter. The results of the experiment are shown in Table 4.3:

Table 4.3: Results achived by Gabor Filter

| **Facial Component** | **True Positive Rate** | **False Positive Rate** | **True Positive %** | **False Positive %** |
|---|---|---|---|---|
| **Nose** | 1076 | 228 | 82.50% | 17.50% |
| **Left Eye** | 957 | 347 | 73.40% | 26.60% |
| **Right Eye** | 963 | 368 | 71.80% | 28.20% |
| **Mouth** | 1150 | 154 | 88.20% | 11.80% |
| **Forehead** | 921 | 383 | 70.60% | 29.40% |
| **Chin** | 872 | 432 | 66.80% | 33.2% |
| **Left Cheek** | 871 | 433 | 66.70% | 33.30% |
| **Right Cheek** | 1067 | 237 | 81.80% | 18.20% |

Gabor Filter was tested using the dataset which consisted of 1304 images. Each image all components are extracted - eyes, nose, chin, forehead, mouth and cheeks. From the results in Table 4.3, 154 instances of the mouth, 154 instances of the nose, 347 instances of the left eyes, 368 instances of the right eyes, 432 instances of the chin, 383 instances of the forehead, 433 instances of the left cheek and 237 instances of the right cheek were misclassified. Gabor Filter has an average True Positive classification rate of 75.23%.

### Linear Binary Pattern

The Linear Binary Pattern is a Textural Feature Extraction Technique which is embedded in the ImageJ [48] toolbox and was used to implement the Linear Binary Pattern. In this research work the Linear Binary Pattern Feature Extraction Technique was applied to the same facial images used for the training of the other feature extraction techniques. The dataset was used as training, validation and testing for the Linear Binary Pattern. The results of the experiment are shown in Table 4.4:

Table 4.4: Results achived by Linear Binary Pattern

| **Facial Component** | **True Identification** | **False Identification** | **True Percentage** | **False Percentage** |
|---|---|---|---|---|
| **Nose** | 818 | 486 | 62.70% | 37.30% |
| **Left Eye** | 1092 | 212 | 83.70% | 16.30% |
| **Right Eye** | 491 | 813 | 37.60% | 62.40% |
| **Mouth** | 1003 | 301 | 76.90% | 23.10% |
| **Forehead** | 1057 | 247 | 81.00% | 19.00% |
| **Chin** | 620 | 684 | 47.50% | 52.50% |
| **Left Cheek** | 398 | 906 | 30.5% | 69.50% |
| **Right Cheek** | 295 | 1009 | 22.70% | 77.30% |

The Linear Binary Pattern was tested using the complete dataset which consisted of 1304 images. Each image all components are extracted - eyes, nose, chin, forehead, mouth and cheeks. From the results in Table 4.4, 301 instances of the mouth, 486 instances of the nose, 212 instances of the left eyes, 813 instances of the right eyes, 684 instances of the chin, 247 instances of the forehead, 906 instances of the left cheek and 1009 instances of the right cheek were misclassified. Gabor Filter has an average True Positive classification rate of 55.33%.

### 7 Hu Moments

7 Hu Moments is a Geometric Feature Extraction Technique which is embedded in the Catalano [2] toolbox and was used to implement the 7 Hu Moments. In this research work 7 Hu Moments Feature Extraction Technique was applied to the same facial images used for the training of the other feature extraction techniques. The dataset was used as training, validation and testing for the 7 Hu Moments. The results of the experiment are shown in Table 4.5:

Table 4.5: Results achived by 7 Hu Moments

| **Facial Component** | **True Identification** | **False Identification** | **True Percentage** | **False Percentage** |
|---|---|---|---|---|
| **Nose** | 1109 | 195 | 85.00% | 15.00% |
| **Left Eye** | 749 | 555 | 57.20% | 42.80% |
| **Right Eye** | 818 | 486 | 62.70% | 37.30% |
| **Mouth** | 1168 | 136 | 89.50% | 10.50% |
| **Forehead** | 590 | 714 | 45.20% | 54.80% |
| **Chin** | 1049 | 255 | 80.50% | 19.50% |
| **Left Cheek** | 788 | 516 | 60.40% | 39.60% |
| **Right Cheek** | 815 | 489 | 62.50% | 37.50% |

7 Hu Moments was tested using the dataset which consisted of 1304 images. Each image all components are extracted - eyes, nose, chin, forehead, mouth and cheeks. From the results in Table 4.5, 136 instances of the mouth, 195 instances of the nose, 555 instances of the left eyes, 486 instances of the right eyes, 255 instances of the chin, 714 instances of the forehead, 516 instances of the left cheek and 489 instances of the right cheek were misclassified. 7 Hu Moments has an average True Positive classification rate of 67.88%.

### Haralick Texture Moments

Haralick texture moments is a Textural Feature Extraction Technique which is embedded in the Catalano [2] toolbox and was used to implement the Haralick texture moments. In this research work the Haralick texture moments Feature Extraction Technique was applied to the same facial images used for the training of the other feature extraction techniques. The dataset was used as training, validation and testing for the Haralick texture moments. The results of the experiment are shown in Table 4.6:

Table 4.6: Results achived by Haralick Texture Moments

| **Facial Component** | **True Identification** | **False Identification** | **True Percentage** | **False Percentage** |
|---|---|---|---|---|
| **Nose** | 868 | 436 | 66.50% | 33.50% |
| **Left Eye** | 1005 | 299 | 77.10% | 22.90% |
| **Right Eye** | 827 | 477 | 63.40% | 36.60% |
| **Mouth** | 915 | 389 | 70.10% | 29.90% |
| **Forehead** | 1073 | 231 | 82.30% | 17.70% |
| **Chin** | 910 | 394 | 69.80% | 30.20% |
| **Left Cheek** | 951 | 353 | 72.80% | 27.20% |
| **Right Cheek** | 1098 | 206 | 84.20% | 15.80% |

The Haralick texture moments was tested using the dataset which consisted of 1304 images. Each image all components are extracted - eyes, nose, chin, forehead, mouth and cheeks. From the results in Table 4.6, 389 instances of the mouth, 436 instances of the nose, 299 instances of the left eyes, 477 instances of the right eyes, 394 instances of the chin, 231 instances of the forehead, 353 instance of the left cheek and 206 instances of the right cheek were misclassified. Haralick texture moments has an average True Positive classification rate of 73.28%.

### General Disscusion of the Feature Extraction Techniques

The results of the True Positive Rate that each Feature Extraction Technique achieved presented that certain Feature Extraction Techniques achieve a higher result at distigusihing the ethnicity of a facial image for that particular component than others did. The results for each component with each Feature Extraction Technique are summarised in Table 4.7.

Table 4.7: True Positive Rates per Component per Feature Extraction Technique

| | **7 Hu Moments** | **Zernike Moments** | **LBP** | **Gabor Filter** | **Haralick Texture Moments** |
|---|---|---|---|---|---|
| **Nose** | 85.0% | 80.1% | 62.7% | 82.5% | 66.5% |
| **Left Eye** | 57.2% | 76.4% | 83.7% | 73.4% | 77.1% |
| **Right Eye** | 62.7% | 76.3% | 37.6% | 71.8% | 63.4% |
| **Mouth** | 89.5% | 72.1% | 76.9% | 88.2% | 70.1% |
| **Forehead** | 45.2% | 79.5% | 81.0% | 70.6% | 82.3% |
| **Chin** | 80.5% | 72.6% | 47.5% | 66.8% | 69.8% |
| **Left Cheek** | 60.4% | 52.6% | 30.5% | 66.7% | 72.8% |
| **Right Cheek** | 62.5% | 71.5% | 22.7% | 81.8% | 84.2% |

For the components of the eyes, it can be seen that Zernike Moments, which is a Geometric Feature Extraction Technique, achieved a 76.3% True Positive Rate for the right eye. As no two eyes are the same, the left eye achieved better results using the Linear Binary Pattern, which is also a Geometric Feature Extraction Technique. The Geometric Feature Extraction Techniques achieve high results for the eyes due to the fact that eyes are structurally different in size and shape between ethnic groups.

The mouth achieved better with two different Feature Extraction Techniques. These Feature Extraction Techniques, were the Gabor Filter with a True Positive Rate of 88.2% and 7 Hu Moments with a True Positive Rate of 89.3%. Both these techniques are texture-based Feature Extraction Techniques and achieved well for the mouth as the colour, shape and coarseness varies for each ethnic group. Due to both Feature Extraction Techniques achieving high results, the weighted average between the two Feature Extraction Techniques is calculated and used as the Feature Extraction Technique for the component of the mouth.

The chin and nose components achieved well with the two Feature Extraction Technique: 7 Hu Moments. This textural Feature Extraction Technique achieved an 80.5% True Positive Rate for the chin and an 85.0% True Positive Rate for the nose. As both components are skin-based components, it is the texture of the skin that is different for each ethnic group. Therefore, for the chin and the nose, 7 Hu Moments is the best Feature Extraction Technique.

The forehead and left and right cheeks are also skin-based components and can define ethnicity for the different groups by the texture, colour and gradients of the components. These components achieved well with the Haralick texture moments as this uses the correlation and homogeneity of the image. The forehead achieved a True Positive Rate of 82.3%, the left cheek had a 72.8% True Positive Rate and the right cheek had a 84.2% True Positive Rate.

Therefore, each image that is processed by the ethnic identification framework, shown in Figure 4.1, has a feature vector that is a fusion of each of these components feature vectors. The components of the nose and chin compute the feature vector using the 7 Hu Moments. The forehead, left cheek and right cheek implement the feature vector using the Haralick texture moments. The eyes compute the feature vector using the Zernike Moments and the Gabor Filter. Lastly, the mouth uses the weighted average of the Gabor Filter and the 7 Hu Moments Feature Extraction Technique. Once all the feature vectors are obtained for each component, they are then fused together and normalised to be used in the classification process.

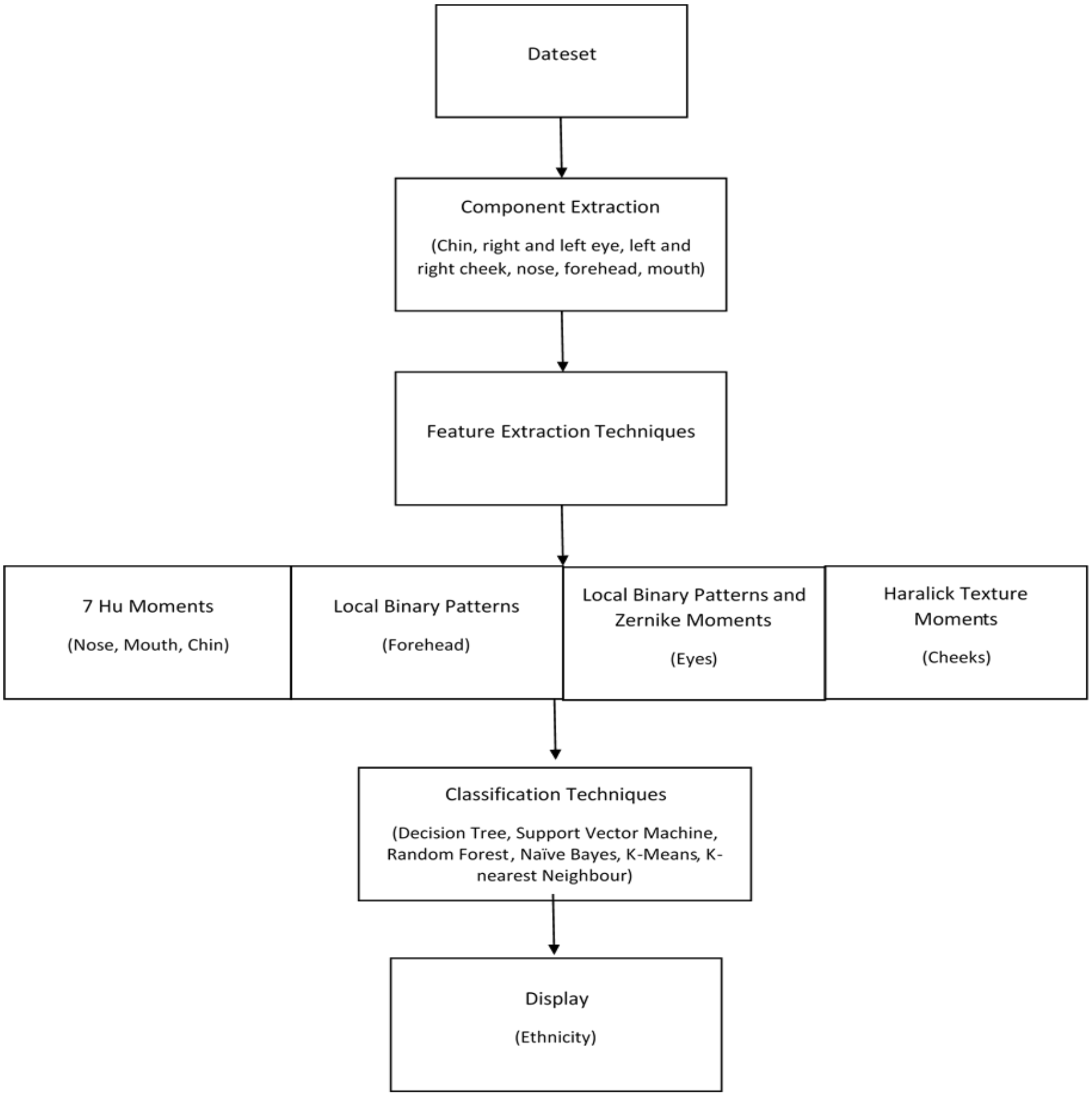



Figure 4.1: The proposed framework of the component-based ethnicity identification system

## 4.5 Feature Extraction and Identification

Once the composite fused and normalised feature vector is obtained, it is then used to identify the ethnicity of an image. This is done by using different Machine Learning Algorithms to obtain the ethnicity of the facial image. These Machine Learning Algorithms that were used were Support Vector Machine, K-Nearest Neighbour, Decision Tree, Random Forest, K-Means and Naïve Bayes.

### 4.5.1 Support Vector Machine

A Support Vector Machine is a supervised machine learning algorithm which was embedded in the javaml [2] toolbox and used to implement the Support Vector Machine. The study took the same dataset as for the other machine learning algorithm to test and train this machine learning algorithm. The results obtained for this experiment are shown in the Figure 4.2.

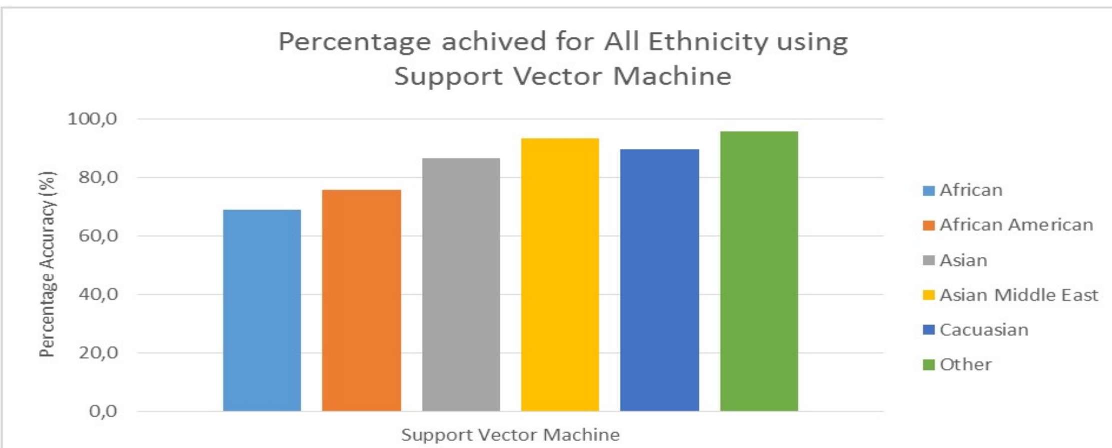



Figure 4.2: Accuracy Achieved for All Ethnicity using Support Vector Machine

The Support Vector Machine was tested using the dataset which presented that Support Vector Machine correctly identified the ethnic group of Other and was the best with 95.6%. Support Vector Machine preformed badly with African Ethnicity as it only achieved 68.9%. This is because SVM could not correctly indentify the African as these images were consered close to the African American images.

### 4.5.2 K-Nearest Neighbour

K-Nearest Neighbour is a supervised machine learning algorithm which was embedded in the javaml [2] toolbox and was used to implement K-Nearest Neighbour. The study took the same dataset as for the other machine learning algorithm to test and train this machine learning algorithm. The results obtained for this experiment are shown in the Figure 4.3.

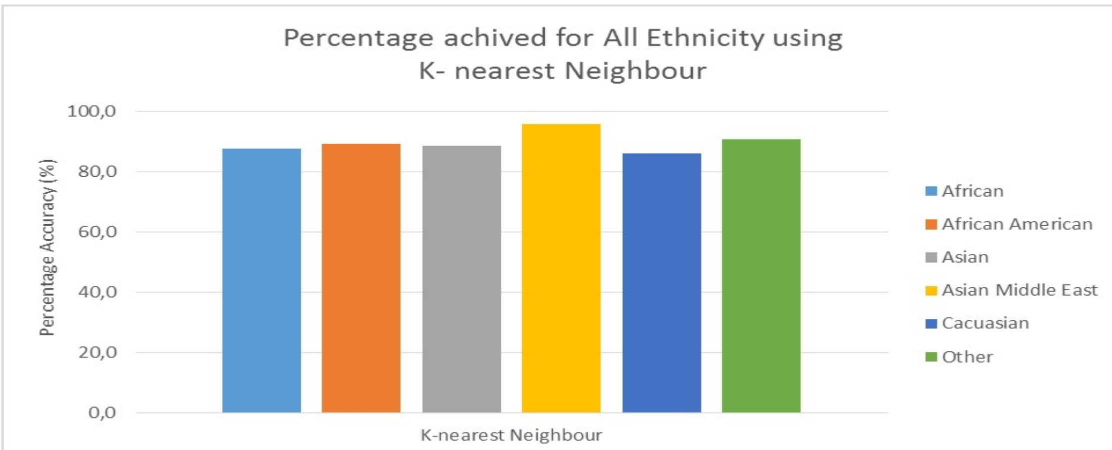



Figure 4.3: Accuracy Achieved for All Ethnicity using K-Nearest Neighbour

K-Nearest Neighbour was tested using the full dataset which presented that K-Nearest Neighbour correctly identified the ethnic group of Asian Middle East and was the best with 95.6%. The worst preforming ethnic group was African, with 87.5%. This is because K-nearest Neighbour could not correctly indentify the African as these images were consered close to the African American images.

### 4.5.3 Decision Tree

The Decision Tree is a supervised machine learning algorithm which was embedded in the javaml [2] toolbox and was used to implement the Decision Tree. The study took the same dataset as for the other machine learning algorithm to test and train this machine learning algorithm. The results that are obtained for this experiment are shown in the Figure 4.4.


Figure 4.4: Accuracy Achieved for All Ethnicity using Decision Tree

Decision Tree was tested using the full dataset which presented that Decision Tree preformed the best for the ethnic group of African American with 87.2%.Decision Tree also could not correctly identify the Other ethnicity, which achieved 69.3%. This was because the Other image feature vectors were extremly close to the Cacuasian and Decision Tree could not find the tree that would correctly define the Other ethnic group.

### 4.5.4 Random Forest

Random Forest is a supervised machine learning algorithm which was embedded in the javaml [2] toolbox and was used to implement the Random Forest. The study took the same dataset as for the other machine learning algorithm to test and train this machine learning algorithm. The results of the experiment are shown in the Figure 4.5.

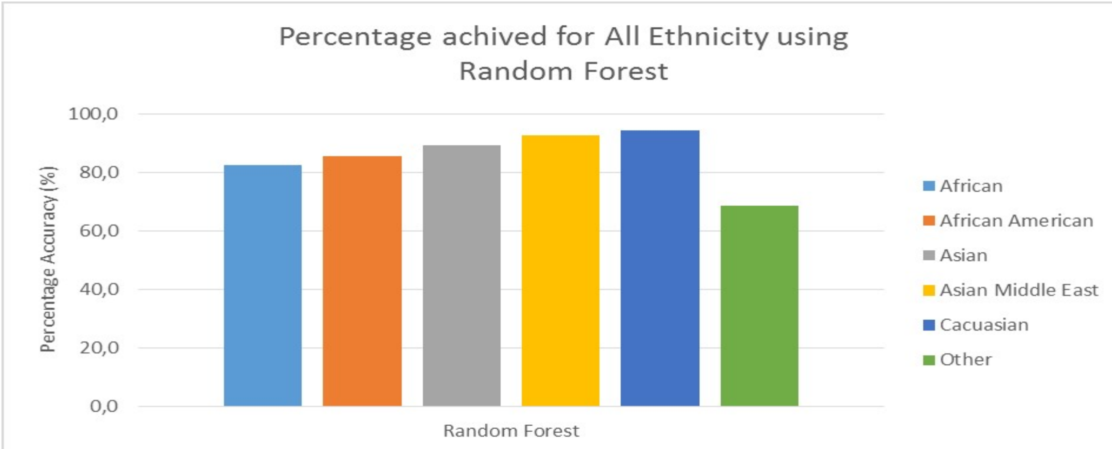



Figure 4.5: Accuracy Achieved for All Ethnicity using Random Forest

Random Forest was tested using the full dataset which presented that Random Forest correctly identified the ethnicity group of Caucasian, which achieved 94.6%. Random Forest achieved poorly for the Other ethnicity, which achieved 68.6%. This was because the Other image feature vectors were extremly close to the Cacuasian and Random Forest could not find the tree that would correctly define the Other ethnic group.

### 4.5.5 Naïve Bayes

The Naïve Bayes is a supervised machine learning algorithm which was embedded in the javaml [2] toolbox and was used to implement the Naïve Bayes. The study took the same dataset as for the other machine learning algorithm to test and train this machine learning algorithm. The results of the experiment are shown in the Figure 4.6.

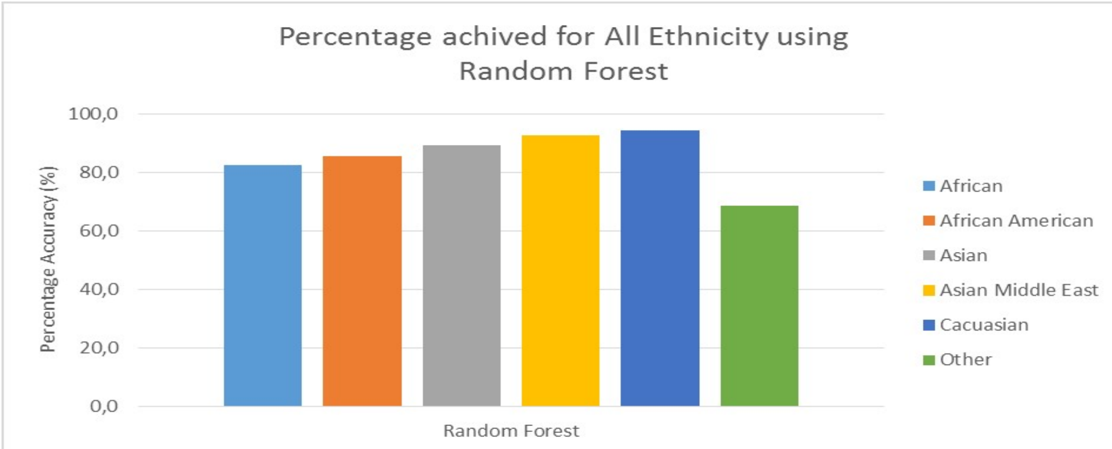



Figure 4.6: Accuracy Achieved for All Ethnicity using Naïve Bayes

Naïve Bayes was tested using the full dataset which presented that Naïve Bayes identified the African ethnicity group, which achieved the result of 98.7% accuracy rate. Naïve Bayes achieved 88.3% for Caucasian, which was the worst performing ethnic group.

### 4.5.6 K-Means

K-Means is an unsupervised machine learning algorithm which was embedded in the javaml [2] toolbox and was used to implement the K-Means. The study took the same dataset as for the other machine learning algorithm to test and train this machine learning algorithm. The results of the experiment are shown in the Figure 4.7.

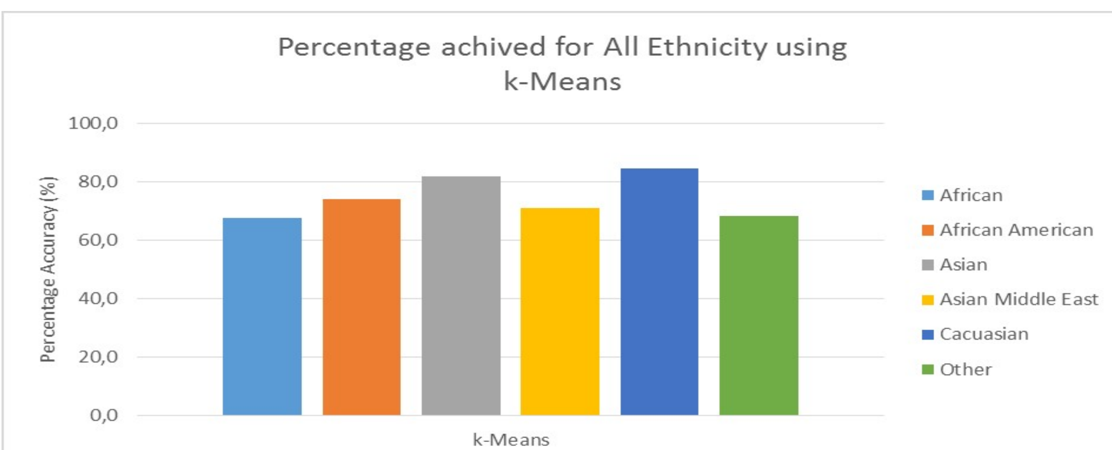



Figure 4.7: Accuracy Achieved for All Ethnicity using K-Means

The K-Means was tested using the full dataset which preesented that K-Means achieved the worse results over all the machine learning algorithms. The African ethnicity group achieved the worse results of 67.9% and Caucasian achieved the best results, with 84.7%.

## 4.6 Analysis of each Ethnic Group

This section will discuss the results achieved for each ethnic group. These ethnic groups were African, African American, Asian, Asian Middle East, Caucasian and Other. These results discussed here are being compared to similar works done in ethnic identification.

### 4.6.1 African Ethnicity Results

In Buchala et al. [9] it was shown that for the African ethnicity they achieved a percentage of 80%. Buchala et al used Principal Component Analysis as there classifier. They used a dataset size of 40 African images from the FERET dataset [45]. For this research work 91 images from the MUCT dataset [41] was used and achieved 98.7% with the Naïve Bayes. This is shown in Figure 4.8. The reason that Naïve Bayes achieved higher results is that there are more images where used hence achieving a more accurate result.

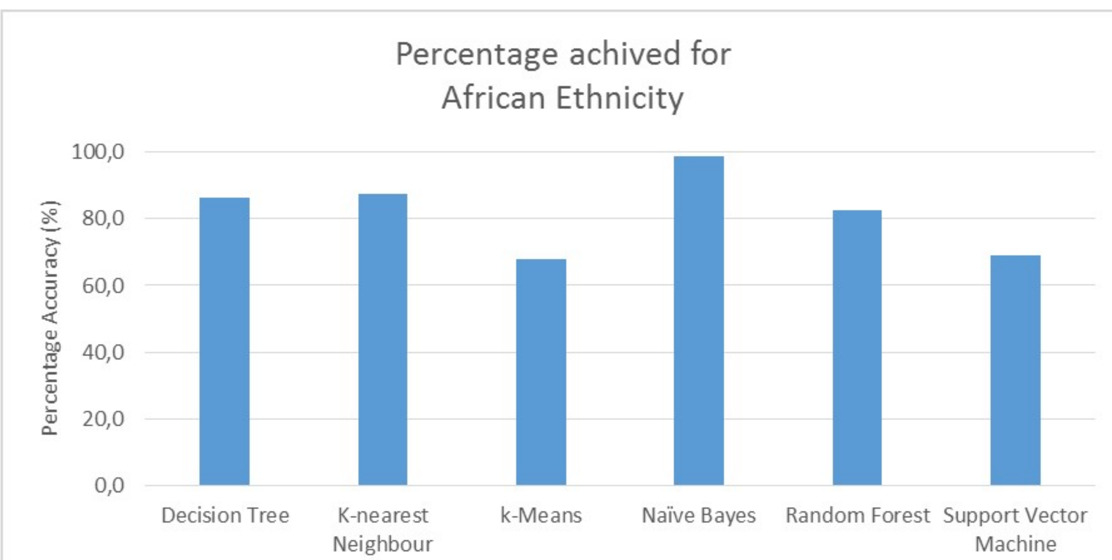



Figure 4.8: Accuracy Achieved for African Ethnicity

### 4.6.2 African American Ethnicity Results

Buchala et al [9] it was shown that African American Ethnicity achieved 82% for this ethnic group. Here they used 320 African American facial images from FERET and used Principal Component Analysis classification. The results achieved in this research work are shown in Figure 4.9. The dataset used contained 79 African American images from different datasets. Naïve Bayes achieved higher results than Buchala et al. for the reason that Naïve Bayes is a more accurate classifier than Principal Component Analysis.

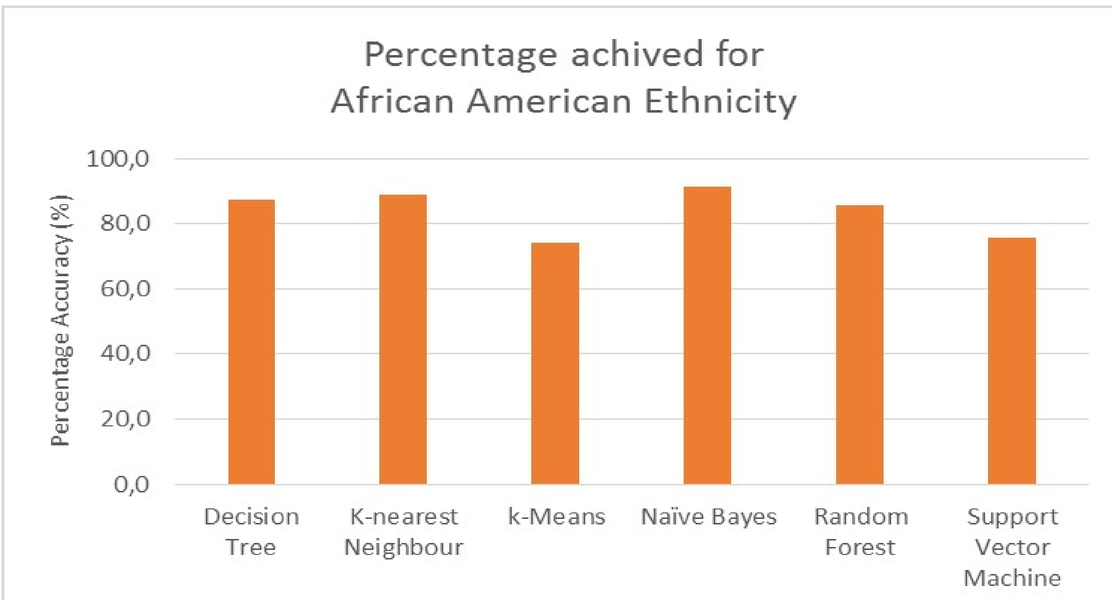



Figure 4.9: Accuracy Achieved for African American Ethnicity

### 4.6.3 Asian Ethnicity Results

In Lu and Juain [40] used two different classification techniques. These were Nearest Neighbour and Linear Discriminant Analysis, which achieved 97% and 95% respectively. They used 200 Asian facial images from the FERET [45] dataset. The results achieved in this research work are shown in Figure 4.10. In this research work, the dataset used contained 179 Asian facial images from the Yale [5] and FERET [45] datasets. Naïve Bayeswas the best preforming classification method and achieved 90.3%. This ethnic group did achieve less than the research work. This could be because of the number of images used and the mixing of different datasets.

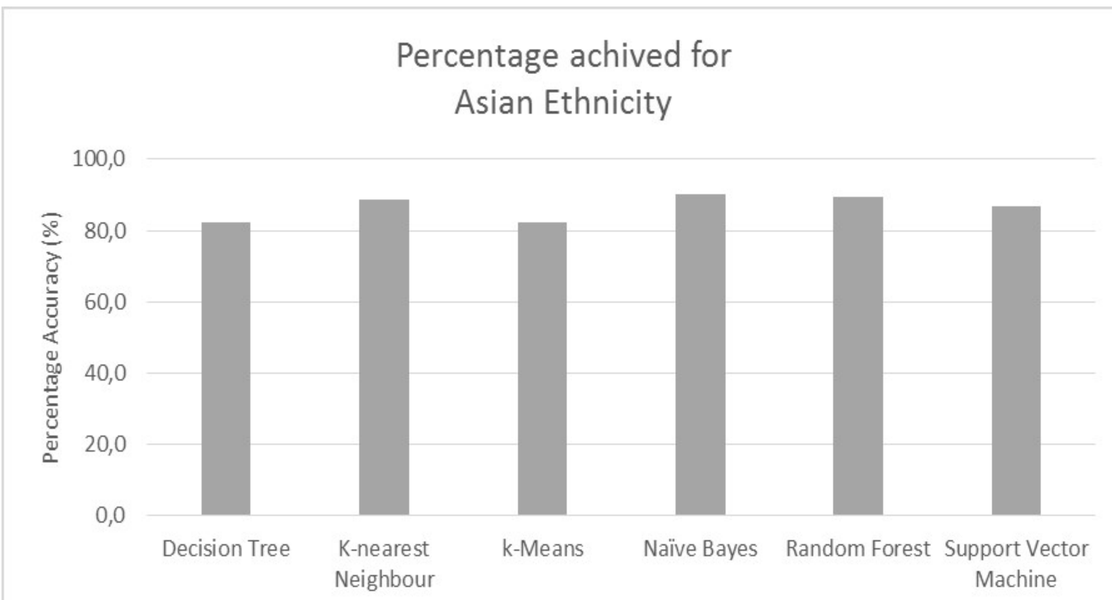



Figure 4.10: Accuracy Achieved for Asian Ethnicity

### 4.6.4 Asian Middle East Ethnicity Results

Buchala et al [9] results for Asian Middle East ethnicity were an average of 83% using Principal Component Analysis classification. This was achieved with a dataset size of 363 Asian Middle East images from random datasets. K-nearest Neighbor was the best achieving Machine Learning Algorithm for Asian Middle East in this study. It achieved a result of 95.6% accuracy rate, which is shown in Figure 4.11. The dataset used 105 Asian Middle East facial images. Here we obtained a higher accuracy than that of Buchala et al.

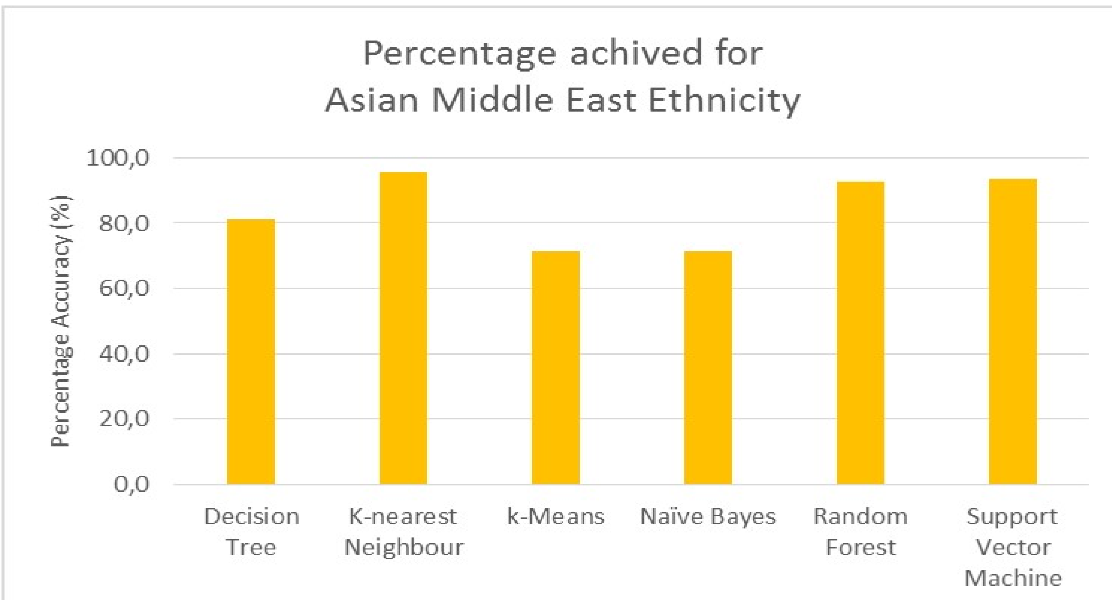



Figure 4.11: Accuracy Achieved for Asian Middle East Ethnicity

### 4.6.5 Caucasian Ethnicity Results

Buchala et al [9] used 1758 images from the FERET dataset. Their results showed that Caucasian ethnicity achieved 82% accuracy by using Principal Component Analysis classification. In this study, 777 Caucasian facial images were used in the dataset. This achieved a result of 94.6% using Random Forest, which is shown in Figure 4.12. This research work achieved a higher accuracy rate for this ethnic group than that of Buchala et al. This is due to the fact that Buchala et al used the same subjects multiple times in their dataset, whereas this research work used different images which helped to distinguish variations in facial images.

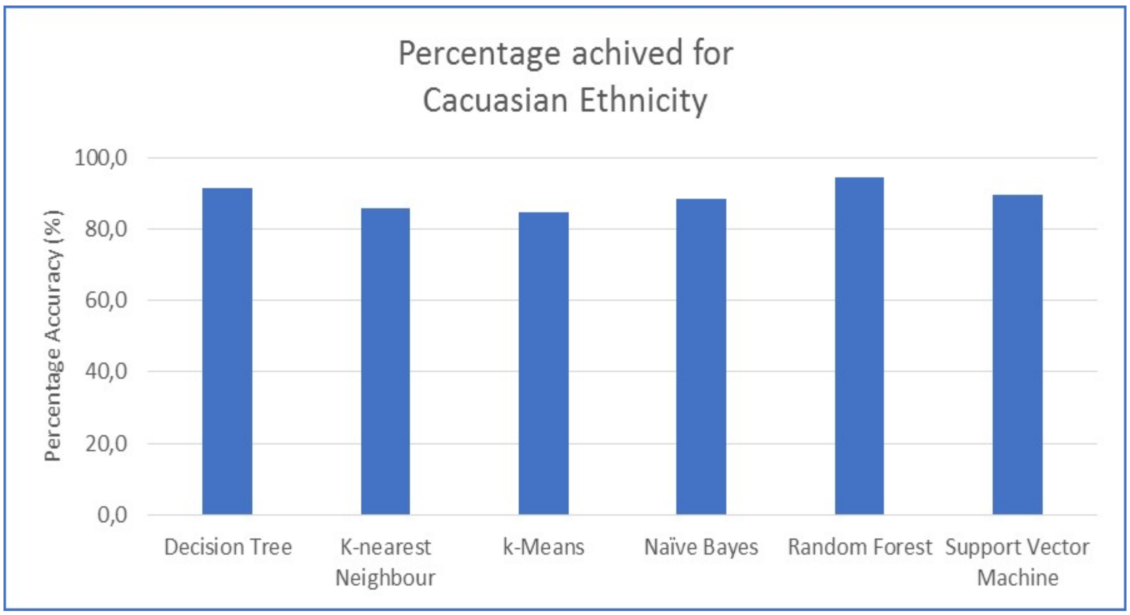



Figure 4.12: Accuracy Achieved for Caucasian Ethnicity

### 4.6.6 Other Ethnicity Results

Tin and Sein [54] used two different classification techniques. On average, the result achieved for Other ethnicity was 93% for Nearest Neighbor and 96% for Principal Component Analysis classification. Tin and Sein used 250 Other facial images which were obtained from the Internet. In this study, Support Vector Machine achieved a result of 95.6% and is shown in Figure 4.13. Here this study used only 73 images from the FERET [45] dataset. On average, this achieved the same as Tin and Sein.

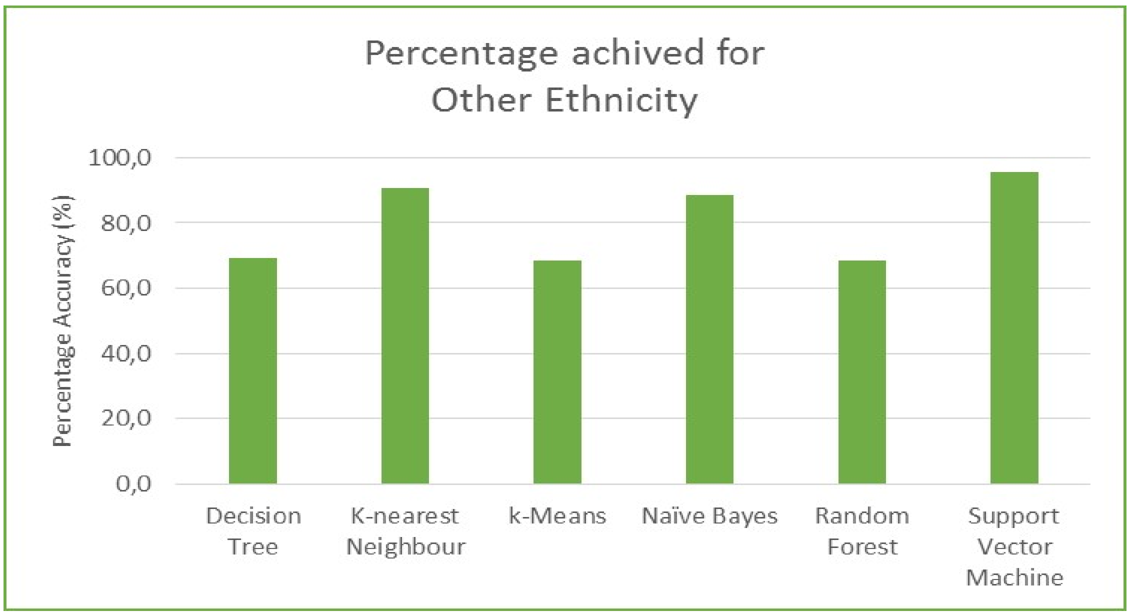



Figure 4.13: Accuracy Achieved for Other Ethnicity

### 4.6.7 Results of all Six Ethnicities

A number of empirical experiments were carried out to investigate whether or not the Fused and Normalised Fusion Vector is suitable to determine ethnicity. The dataset used for the training comprised 10% of each ethnic group and was tested against different Machine Learning Algorithms.

Experiments presented that K-Nearest Neighbour Machine Learning Algorithm achieved an 89.6% ethnic distinction rate with the lowest achieving Machine Learning Algorithm being K-Means, which achieved 74.8%. The variation between Machine Learning Algorithms 15% Ethnic Identification Rate was 15%; this is shown in Figure 4.14.

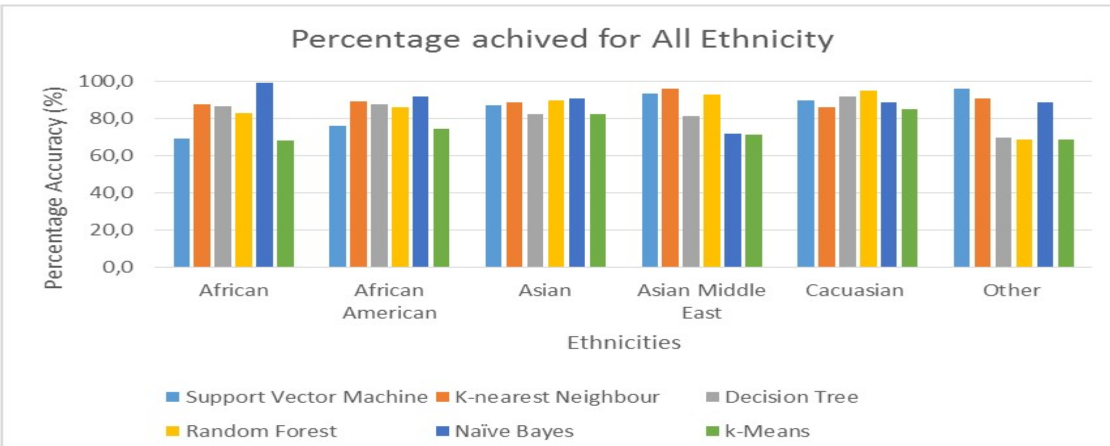



Figure 4.14: Accuracy Achieved for All Six Ethnicity

Table 4.8 presents the comparison between the results achieved by related research studies on ethnic identification and our average ethnic identification rate. From the results, we can see that this study achieved better results than those of related works.

Table 4.8: Comparison of Related Studies Results with Our Results for Ethnicity Identification

| | Asian | African | African American | Asian Middle East | Caucasian | Other |
|---|---|---|---|---|---|---|
| Lu and Jain [40] | 97.7% | - | - | - | - | - |
| Buchala et al [9] | - | 80.2% | 80.0% | 83.1% | 82.0% | - |
| Tin and Sein [54] | - | - | - | - | - | 96.0% |
| **Our Work** | **85.6%** | **84.7%** | **85.8%** | **86.8%** | **90.8%** | **82.3%** |

## 4.7 Conclusion

In this chapter, the proposed framework is discussed and explained. The experimental results have been presented and discussed and presents that the proposed framework gives a better performance and produces a more accurate ethnicity identification rate for an image than other research studies.

The next chapter concludes the research work and provides direction for future work.

# Chapter 5

# Conclusion and Future Works

## 5.1 Summary of work

In this study, a component-based ethnicity identification framework was proposed. The study provided a comprehensive, detailed review of the most popular facial recognition systems, as well as an investigation into ethnicity identification recognition and component-based analysis for facial recognition. These investigations revealed the shortcomings in previous studies. Some of these shortcomings included the fact that the whole face was used to identify ethnicity and in some cases the number of ethnic groups was small.

This study set out to address some of the above shortcomings. This was done by applying Haar Transform to each facial image to automatically extract the components of the left and right eye, left and right cheek, nose, mouth, chin and forehead. Once each component was extracted, analysis was carried out to ascertain which feature vector technique is best to identify certain ethnicities. It was presented that Linear Binary Pattern and Zernike Moments are the best feature extractions for the eyes, Haralick texture moments was best for the left cheek and forehead, Gabor Filter was best for the right cheek and 7 Hu Moments was best for the nose, chin and mouth. Once all the components of the feature vector are obtained, they are then concatenated and normalized. As a result, this approach detects ethnicity in facial images with high accuracy.

The facial component-based ethnicity identification framework takes in an image and components are extracted automatically. These components use different feature extraction techniques in order to obtain the feature vector. The feature vectors are the classified using different classification techniques in order to obtain the Ethnicity of the facial image.

Experiments were carried out on all feature extraction and ethnicity classes. Compared with previous studies, they presented that ethnicities were more accurately detected for all classes and that the feature extractions achieved an average ethnicity identification rate of 84.4%. African ethnicity identification rate was 98.7% with Naïve Bayes, while a 91.5% rate was achieved with Nave Bayes for African American ethnicity identification. Naïve Bayes achieved a rate of 90.3% for Asian ethnicity identification and K-Nearest Neighbor Classification obtained a 95.6% rate for Asian Middle East ethnicity identification. A rate of 94.6% was achieved for Caucasian ethnicity identification with Random Forest and Support Vector Machine Classification achieved a 95.6% rate for Other ethnicity identification.

Although this study provides solutions to some of the current problems in component-based ethnicity identification, there is still some room for improvement. These limitations are discussed in the next section.

## 5.2 Limitations of the System and Future work

The most important steps in the development of an ethnicity identification framework to test the shortcomings are here revealed and reported:

- **Sample Images:** In this study, grey level images of the facial components images are used in the identification of the ethnicity. If there is bad illumination during the capturing of the facial image, this affects the performance of the framework with that image. As a short term solution, this study removed the images that were not suitable. In future, at the time of image capture, more precaution can be taken to control the illumination of the image.
- **Component Detection:** With some facial images, it is still difficult to identify the components with this framework. The reason for this is that some areas of the face were covered with facial hair or sun glasses were worn. As a short term solution, the images with these issues were removed. In

future, auto-detection of unidentifiable components needs to be done and the components removed from the dataset.

## 5.3 Conclusion

This study has presented a component-based ethnicity identification framework. Facial components are firstly detected (left and right eye, left and right cheek, nose, mouth, chin and forehead), and the feature vectors are obtained for each component. These feature vectors are then concatenated and normalized. The feature vectors are validated with numerous classification techniques. In terms of application, this framework can be used as a solution for Security, Surveillance Systems and General Identity Verification.

# Bibliography


[1] Abdullah Abdul-Husain and Enas Abood. Race classification using craniofacial features from colored face images. *International Journal of Engineering and Advanced Technology (IJEAT)*, 3:138–143, 2014. 4, 12, 15

[2] Thomas Abeel, Yves Van de Peer, and Yvan Saeys. Java-ml: A machine learning library. *Journal of Machine Learning Research*, 10(Apr):931–934, 2009. 33, 35, 36, 38, 39, 43, 44, 45, 46, 47

[3] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face recognition with local binary patterns. In *European conference on computer vision*, pages 469–481. Springer, 2004. 4, 9

[4] Ali Atharifard and Sedigheh Ghofrani. Robust component-based face detection using color feature. In *Proceedings of the World Congress on Engineering*, volume 2, pages 6–8, 2011. 9

[5] Peter N Belhumeur, João P Hespanha, and David J Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, 1997. 7, 34, 50

[6] Sebastian Berisha. Image classification using gabor filters and machine learning. 2009. 26

[7] Kathryn Bonnen, Brendan F Klare, and Anubhav K Jain. Component-based representation in automated face recognition. *Information Forensics and Security, IEEE Transactions on*, 8(1):239–253, 2013.

[8] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library.* " O'Reilly Media, Inc.", 2008. 19, 22, 33

[9] Samarasena Buchala, Neil Davey, Tim M Gale, and Ray J Frank. Principal component analysis of gender, ethnicity, age, and identity of face images. *Proc. IEEE ICMI*, 2005. xiii, 1, 4, 10, 11, 15, 16, 48, 49, 51, 52, 54

[10] Bor-Chun Chen, Yan-Ying Chen, Yin-Hsi Kuo, and Winston H Hsu. Scalable face image retrieval using attribute-enhanced sparse codewords. *Multimedia, IEEE Transactions on*, 15(5):1163–1173, 2013. 9

[11] Matthew Collins, Jianguo Zhang, Paul Miller, Hongbin Wang, and Huiyu Zhou. Eigenbody: analysis of body shape for gender from noisy images. In *International machine vision and image processing conference*, 2010. 7

[12] Walker Connor. Beyond reason: the nature of the ethnonational bond. *Ethnic and racial studies*, 16(3):373–389, 1993. 6

[13] Huaxiong Ding, Di Huang, Yunhong Wang, and Liming Chen. Facial ethnicity classification based on boosted local texture and shape descriptions. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–6. IEEE, 2013. 13, 15

[14] Pedro Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, 2012. 29, 30, 32

[15] Hongbo Du, Sheerko Hma Salah, and Hawkar O Ahmed. A color and texture based multi-level fusion scheme for ethnicity identification. In *SPIE Sensing Technology+ Applications*, pages 91200B–91200B. International Society for Optics and Photonics, 2014. 13, 15

[16] Yun Fu, Liangliang Cao, Guodong Guo, and Thomas S Huang. Multiple feature fusion by subspace learning. In *Proceedings of the 2008 international conference on Content-based image and video retrieval*, pages 127–134. ACM, 2008. 28

[17] Clifford Geertz. The integrative revolution: Primordial sentiments and civil politics in the new states. *Old Societies and New States*, page 150, 1967. 6

[18] Elliott D Green. Redefining ethnicity. In *47th Annual International Studies Association Convention, March, San Diego, CA, available at http: personal. lse. ac. uk/green/ISA. pdf (accessed 30 May 2011)*, 2006. 6

[19] Srinivas Gutta and Harry Wechsler. Face recognition using hybrid classifier systems. In *Neural Networks, 1996., IEEE International Conference on*, volume 2, pages 1017–1022. IEEE, 1996. 16

[20] Henry E Hale. Explaining ethnicity. *Comparative Political Studies*, 37(4):458–485, 2004. 6

[21] Hu Han, Christina Otto, and Anubhav K Jain. Age estimation from face images: Human vs. machine performance. In *Biometrics (ICB), 2013 International Conference on*, pages 1–8. IEEE, 2013. 7

[22] Bernd Heisele and Volker Blanz. Morphable models for training a component-based face recognition system. *Computer Vision and Patern Recognition*, page 1055, 2004. 2, 15

[23] Bernd Heisele and Volker Blanz. Morphable models for training a component-based face recognition system, 2006. 3, 4, 15

[24] Bernd Heisele, Purdy Ho, Jane Wu, and Tomaso Poggio. Face recognition: component-based versus global approaches. *Computer vision and image understanding*, 91(1):6–21, 2003. 15

[25] Bernd Heisele, Thomas Serre, Massimiliano Pontil, and Tomaso Poggio. Component-based face detection. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–657. IEEE, 2001. 15

[26] Donald L Horowitz. *Ethnic groups in conflict*. Univ of California Press, 1985. 6

[27] Gee-Sern Hsu, Hsiao-Chia Peng, and Kai-Hsiang Chang. Landmark based facial component reconstruction for recognition across pose. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*, pages 34–39. IEEE, 2014. 3

[28] Di Huang, Caifeng Shan, Mohsen Ardebilian, and Liming Chen. Facial image analysis based on local binary patterns: a survey. *IEEE Transactions on Image Processing*, 2011. 13

[29] Jennifer Huang, Volker Blanz, and Bernd Heisele. Face recognition using component-based svm classification and morphable models. In *Pattern Recognition with Support Vector Machines*, pages 334–341. Springer, 2002. 15

[30] Harold Robert Isaacs. *Idols of the tribe: Group identity and political change.* Harvard University Press, 1975. 6

[31] Rabia Jafri and Hamid R Arabnia. A survey of face recognition techniques. *Journal of information processing systems*, 5(2):41–68, 2009. 7

[32] Felix Juefei-Xu and Marios Savvides. Multi-class fukunaga koontz discriminant analysis for enhanced face recognition. *Pattern Recognition*, 52:186–205, 2016. 4

[33] Tapas Kanungo, David M Mount, Nathan S Netanyahu, Christine D Piatko, Ruth Silverman, and Angela Y Wu. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE transactions on pattern analysis and machine intelligence*, 24(7):881–892, 2002. 30, 31, 32

[34] Brendan F. Klare, Mark J. Burge, Joshua C. Klontz, Richard W. Vorder Bruegge, and Anil K. Jain. Face recognition performance: Role of demographic information. *IEEE Transactions on Information Forensics and Security*, 7(6):1789–1801, 2012. 9

[35] Neeraj Kumar, Alexander C Berg, Peter N Belhumeur, and Shree K Nayar. Attribute and simile classifiers for face verification. In *2009 IEEE 12th International Conference on Computer Vision*, pages 365–372. IEEE, 2009. 8

[36] Hui Lin, Huchuan Lu, and Lihe Zhang. A new automatic recognition system of gender, age and ethnicity. In *2006 6th World Congress on Intelligent Control and Automation*, volume 2, pages 9988–9991. IEEE, 2006. 1, 8

[37] Chengjun Liu. Enhanced independent component analysis and its application to content based face image retrieval. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(2):1117–1127, 2004.

[38] Daniel Lowd and Pedro Domingos. Naive bayes models for probability estimation. In *Proceedings of the 22nd international conference on Machine learning*, pages 529–536. ACM, 2005. 29

[39] Xiaoguang Lu. Image analysis for face recognition. *Personal notes, May*, 5, 2003. 7, 8

[40] Xiaoguang Lu and Anil K Jain. Ethnicity identification from face images. In *Defense and Security*, pages 114–123. International Society for Optics and Photonics, 2004. 4, 8, 10, 15, 50, 54

[41] Stephen Milborrow, John Morkel, and Fred Nicolls. The muct landmarked face database. *Pattern Recognition Association of South Africa*, 201(0), 2010. 8, 34, 48

[42] Baback Moghaddam and Alexander P Pentland. Face recognition using view-based and modular eigenspaces. In *SPIE's 1994 International Symposium on Optics, Imaging, and Instrumentation*, pages 12–21. International Society for Optics and Photonics, 1994. 7

[43] Colm Mulcahy. Image compression using the haar wavelet transform. *Spelman Science and Mathematics Journal*, 1(1):22–31, 1997. 21

[44] Choon Boon Ng, Yong Haur Tay, and Bok Min Goi. Vision-based human gender recognition: A survey. *arXiv preprint arXiv:1204.1611*, 2012. 9

[45] P Jonathon Phillips, Harry Wechsler, Jeffery Huang, and Patrick J Rauss. The feret database and evaluation procedure for face-recognition algorithms. *Image and vision computing*, 16(5):295–306, 1998. 9, 34, 48, 50, 53

[46] S Hma Salah, H Du, and N Al-Jawad. Fusing local binary patterns with wavelet features for ethnicity identification. In *Proc. IEEE Int. Conf. Signal Image Process*, volume 21, pages 416–422, 2013. 3, 4, 12, 15, 27

[47] Ferdinando S Samaria and Andy C Harter. Parameterisation of a stochastic model for human face identification. In *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, pages 138–142. IEEE, 1994. 7, 34

[48] Johannes Schindelin, Curtis T Rueden, Mark C Hiner, and Kevin W Eliceiri. The imagej ecosystem: An open platform for biomedical image analysis. *Molecular reproduction and development*, 82(7-8):518–529, 2015. 33, 37

[49] Parin M Shah. Face detection from images using support vector machine. 2012. 4, 10

[50] Edward Shils. Primordial, personal, sacred and civil ties: Some particular observations on the relationships of sociological research and theory. *The British Journal of Sociology*, 8(2):130–145, 1957. 6

[51] John F Stack. *The primordial challenge: Ethnicity in the contemporary world*, volume 154. Praeger Pub Text, 1986. 6

[52] Usman Tariq, Yuxiao Hu, and Thomas S Huang. Gender and ethnicity identification from silhouetted face profiles. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 2441–2444. IEEE, 2009. 4, 11, 15

[53] Michael Reed Teague. Image analysis via the general theory of moments*. *JOSA*, 70(8):920–930, 1980. 23, 24

[54] Hlaing Htake Khaung Tin and Myint Myint Sein. Race identification for face images. *ACEEE Int. J. Inform. Tech*, 1(02), 2011. 1, 4, 11, 15, 16, 53, 54

[55] George Toderici, Sean M Omalley, George Passalis, Theoharis Theoharis, and Ioannis A Kakadiaris. Ethnicity-and gender-based subject retrieval using 3-d face-recognition techniques. *International Journal of Computer Vision*, 89(2-3):382–391, 2010. 14, 15, 16

[56] Matthew Turk. Eigenfaces and beyond. *Face Processing: Advanced Modeling and Methods*, pages 55–86, 2005. 3, 7

[57] Benjamin Weyrauch, Bernd Heisele, J Huang, and Volker Blanz. Component-based face recognition with 3d morphable models. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*, pages 85–85. IEEE, 2004.

[58] Phillip Ian Wilson and John Fernandez. Facial feature detection using haar classifiers. *Journal of Computing Sciences in Colleges*, 21(4):127–133, 2006.

[59] Damon L Woodard, Shrinivas J Pundlik, Jamie R Lyle, and Philip E Miller. Periocular region appearance cues for biometric identification. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 162–169. IEEE, 2010. 13, 15

[60] Zhiguang Yang and Haizhou Ai. Demographic classification with local binary patterns. In *Advances in Biometrics*, pages 464–473. Springer, 2007. 4, 10, 15

[61] Naser Zaeri. Component-based thermal face recognition. *British Journal of Applied Science & Technology*, 4(6):945–966, 2014.

[62] Zhengyou Zhang. Microsoft kinect sensor and its effect. *MultiMedia, IEEE*, 19(2):4–10, 2012. 3, 4, 9

[63] Wenyi Zhao and Rama Chellappa. *Face Processing: Advanced Modeling and Methods: Advanced Modeling and Methods*. Academic Press, 2011. 8