

GENDER CLASSIFICATION FROM FACIAL IMAGES

By

Ng Epin

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF INFORMATION SYSTEMS (HONS)

INFORMATION SYSTEMS ENGINEERING

Faculty of Information and Communication Technology

(Perak Campus)

May 2015

DECLARATION OF ORIGINALITY

I declare that this report entitled “**GENDER CLASSIFICATION FROM FACIAL IMAGES**” is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : _____

Name : NG EPIN

Date : 14 September 2015

ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. Ng Hui Fuang for his support and guidance. I thank the Defense Advanced Research Products Agency (DARPA) for permission to use their facial database. I would also like to expand my deepest gratitude to my family and friends for putting up with me.

ABSTRACT

In this project, several classification techniques are evaluated in order to propose a technique that works well with principle component analysis (PCA) based gender classification system. Firstly, the facial database is divided by gender and is separated into training set and input set. Then PCA is applied to the training set to extract facial features. Lastly, classification techniques are used to classify input set into their respective categories. Based on the number of correct classification, accuracy, performance and ability to handle large database of each classification techniques will be evaluated and the results will be used to recommend suitable classification method for gender classification application.

TABLE OF CONTENTS

TITLE	i
DECLARATION OF ORIGINALITY	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	vi
LIST OF TABLES	viii
LIST OF SYMBOLS	ix
LIST OF ABBREVIATIONS	x
CHAPTER 1 INTRODUCTION	1
1-1 Project Background	1
1-2 Motivation and Problem Statement	2-3
1-3 Project Scope	3-4
1-4 Project Objectives	4
1-5 Impact, Significance and Contribution	4-5
CHAPTER 2 LITERATURE REVIEW	6
2-1 Face Detection Algorithms	6-7
2-2 Feature Extraction Technique	7-8
2-3 Recent Research on Gender Classification Methods	8-9
CHAPTER 3 PROPOSED METHOD/APPROACH	10
3-1 Methodologies	10
3-1-1 Steps of Gender Classification in Face Recognition	10
3-1-2 Face Database	11-13
3-1-3 Face Detection	14
3-1-4 Pre-processing	14-15
3-1-5 Features Extraction by PCA (Training)	15-17
3-1-6 Features Extraction by PCA (Recognize)	17
3-1-7 Classification Methods	18-21

3-1-8 Experimental Result & Analysis	22-30
3-2 Tools	30
3-2-1 MATLAB	30
3-3 Implementation Issue and Challenges	30
3-4 Timeline	31
CHAPTER 4 CONCLUSION	32
REFERENCE	33-34

LIST OF FIGURES

Figure Number	Title	Page
Figure 2.1	Skin Colour – Unusual Lightning.	7
Figure 2.2	Skin Colour – Normal Lightning.	7
Figure 3.1	Common Approach for Gender Classification in Face Recognition.	10
Figure 3.2	Male sample from Aberdeen database.	11
Figure 3.3	Female sample from Aberdeen database.	12
Figure 3.4	Male samples from CAS-PEAL database.	12
Figure 3.5	Female samples from CAS-PEAL database.	13
Figure 3.6	Male sample from FERET database.	13
Figure 3.7	Female sample from FERET database.	13
Figure 3.8	Multilayer neural network.	20
Figure 3.9	SVM – Without Margin	21
Figure 3.10	SVM – With Margin	21
Figure 3.11	Overall Accuracy Bar Chart	23
Figure 3.12	Neural Network Additional Runs Accuracy Line Chart	24
Figure 3.13	Average Classification Time Bar Chart	25
Figure 3.14	Aberdeen Dataset Line Chart	27
Figure 3.15	CAS-PEAL Dataset Line Chart	28
Figure 3.16	FERET Dataset Line Chart	29
Figure 3.17	Gantt Chart	31

LIST OF TABLES

Table Number	Title	Page
Table 3.1	Overall Accuracy	22-23
Table 3.2	Neural Network Additional Runs Accuracy	24
Table 3.3	Average Classification Time	25
Table 3.4	Aberdeen Dataset	26-27
Table 3.5	CAS-PEAL Dataset	27-28
Table 3.6	FERET Dataset	28-29

LIST OF SYMBOLS

Γ	Gamma
Ψ	Psi
Φ	Phi
Λ	Lambda
μ	Mu
Ω	Omega
θ	Theta
Θ	Theta
$\%$	Percent Sign

LIST OF ABBREVIATIONS

<i>PCA</i>	Principal Component Analysis
<i>K-NN</i>	K-Nearest Neighbors
<i>SVM</i>	Support Vector Machine
<i>NN</i>	Neural Network
<i>LUT</i>	Look Up Table
<i>LBP</i>	Local Binary Patterns
<i>FERET</i>	Facial Recognition Technology (Database)
<i>WWW</i>	World Wide Web

CHAPTER 1 INTRODUCTION

1-1 Project Background

Face recognition is one of many biometric identification techniques for recognizing and verifying faces from an image. By using face detection algorithm, a computer identifies the positions of faces in an image, then facial features are extracted from the part of the image and finally the facial features data extracted are then compared with the features stored in a face database to find the most alike face to the detected face. The first research done on face recognition is in the 1960s by Woodrow Wilson (Woody) Bledsoe, the pioneer of face recognition (University of Texas at Austin, 1998). The face recognition program created by Bledsoe was not able to identify position of facial features without the aid from human. Variabilities like pose, angle, aging, distance, lighting intensity and facial expression can cause difficulty in facial recognition. In May 1971, Goldstein, Harmon, Lesk at Bell Telephone Laboratories published a paper to evaluate how good computers can identify human faces by subjectively deduce facial features such as nose size and lip width (Goldstein, et al., 1971). Location of facial features still require input from user before recognition. In year 1987, Sirovich and Kirby published a paper indicate that principle component analysis (PCA) can be used to reduce database of faces to form a face with foundation of facial features (Sirovich & Kirby, 1987). PCA-based face recognition is a traditional method to recognize faces, despite its existence for a long time it is still a decent and widely used technique to recognize faces until now.

PCA-based face recognition not only can recognize a particular person from its training database, it can also be used to recognize expression, gender, race and age. In this proposal, we analyse different classification techniques to find the technique that offer the highest accuracy in gender recognition. One of the classification techniques is the k-nearest neighbors (k-NN) algorithm, it is a simple and relatively fast technique when compared with others classification techniques such as logistic regression, support vector machine (SVM), neural network and etc.

1-2 Motivation and Problem Statement

Many research studies have been conducted on face recognition to improve its accuracy since the first research by Woodrow Bledsoe but face recognition still far from achieving accuracy that on par with human, let alone accuracy of gender classification using face recognition.

According to Dautenhahn (Dautenhahn, 2007), social skills are essential requirements in robotic field when it comes to human-machine interaction. A socially competent robot can ease human-machine interaction and gain human acceptance (Dautenhahn, 2007). In order to improve social skill, a machine must know information about the person it interact with. For example, to start a conversation with human being, the machine must first addressing the person with Sir or Madam. One of the approaches for machines to gain access to such information is through gender classification using face recognition techniques.

NEC, a multinational information technology company, is developing face recognition system that is capable of estimating gender to help marketers gather key demographic data about those eyeing their products (DigInfo TV, 2012). With the data, marketers can know more about the consumers that are interested in their product and those who doesn't. These information are important to marketers, it helps them to improve their existing product for specific group of consumers in order to maintain the market and compete with competitors, or marketers can use the information to create new product to target new market. Additionally, the gathered data can be used in data mining to analyse market and gain access to useful information that gives the marketer a competitive edge (The New York Times, 2004).

The traditional classification method to perform gender classification is the k-NN technique which offers simplicity and solid performance. But it is not the best classification technique compared to other techniques in terms of accuracy.

Additionally, k-NN technique has a shortcoming, when number of sample are much greater than k (number of closest training samples in the feature space), the samples that aren't considered as closest training sample will be ignored and do not contribute to the final result of classification. For example, a face recognition system that classifies gender has 50 male samples and 50 female samples, and k is equal to 5.

When a new sample is classified by the system, 5 closest samples will be selected by

the k-NN algorithm and the rest are ignored. If 3 or more closest samples were male, then the new sample will be classified as male. In a nutshell, 5 out of 100 samples contribute to the final result and 95 samples are disregarded. K closest samples could consist some bad samples, sample that has some major similarity to the new sample but both of the samples are from different categories. If some of the samples are bad samples, they could greatly degrade the accuracy of the whole gender classification system. In this proposal, we analyse several classification techniques to find the technique that offer the highest accuracy in gender classification. By finding a better classification technique to replace the traditional k-NN technique, we believe it could improve gender classification in terms of accuracy and make it more reliable and feasible to use. Classification algorithms such as logistic regression, neural network and support vector machine does not share the same shortcoming with k-NN algorithm, however large amount of sample can cause generalization in feature space and makes decision boundary hard to project which can be a problem for these three algorithms. In this proposal, we compared the three algorithms to find out which algorithms can handle large database reliably. Moreover, this project also includes performance measure of different classification techniques to find out the best algorithms.

1-3 Project Scope

This project mainly focus on improving accuracy of gender classification by analysing different classification technique and compare their accuracy improvement on gender classification. A face recognition system was built with the purpose of measuring the algorithms performance and accuracy. Several face databases were retrieved from websites and use as training database for the face recognition system. Additionally, this project only focus on investigate and analyse the use of various classification algorithms, so only frontal facial images were used as training set. Different numbers of samples in the database were used to test the performance of the classification algorithms on sample size. Performance of the classification algorithms were measured and compared with each other. We also proposed a new classification technique to compare with other existing classification technique. The proposed method calculates sum of distance from new sample to training samples and divide by

number of samples for each category to find average distance, then categorize the new sample to the lowest average distance category. K-nearest neighbors (k-NN), logistic regression, support vector machine (SVM), neural network and the proposed method were used to compare with each other.

1-4 Project Objectives

- Build a gender classification system based on facial images and PCA features.
- Compare and find the algorithm that can handle large training samples while retaining reliable accuracy.
- Recommend suitable classification method for gender classification application.

1-5 Impact, Significance and Contribution

A face recognition system does not recognize faces in just one single step or procedure, it requires the use of many image processing techniques and statistical procedure to function properly such as normalization, principal component analysis, Euclidean distance and etc. Every part of the system can be drawback or contribute to the accuracy and reliability of the whole system. This makes every part of the system equally important and slight changes will affect the whole system directly. In this project, we explored different classification techniques to find out the technique that is most suitable for and works well with gender classification. This research suggest better ways for developers to implement a solid and reliable gender classification in face recognition system or to improve existing system.

The traditional gender classification method in face recognition, the k-nearest neighbours algorithm, has some shortcomings, most sample did not contribute to the final result of the classification when the total number of samples is much greater than k and it's vulnerable to bad samples, samples that are similar to the input sample but are from different categories. By using classification techniques that do not inherit the k-NN problem, the misclassification rate of a face recognition system can be reduced. Classification algorithms such as logistic regression, support vector machines (SVM),

neural network and the proposed method does not inherit k-NN shortcoming, makes them ideal candidates to replace k-NN algorithm.

CHAPTER 2 LITERATURE REVIEW

2-1 Face Detection Algorithms

Face detection methods are divided into roughly four categories, the feature invariant approaches, template matching approaches, knowledge-based approaches and appearance-based approaches (Yang & Ahuja, 2001). Feature invariant approaches locate facial features that are invariant to face angle, position, pose and lighting condition. Template matching approaches uses pre-selected faces as templates to compare with input image. The knowledge-based approaches uses rules and fact about human faces to model facial features, for example a face consist of pair of symmetric eyes, a nose underneath the eyes and the mouth at the bottom. Appearance-based approaches is similar to template matching approaches, it uses pre-labelled sets of images to train or derive pattern database which can be compare with input image.

Tathe and Narote (2012), Chai et. al. (2009), and Rahman et. al. (2013) proposed a face detection technique uses human skin colour models. The skin colour area in image is located using skin colour models and template matching is then perform within the area to locate the face actual location. Using skin colour to locate face position and its features do improve face recognition performance since skin colour required a little computation compared to appearance-based approach. In addition, detect faces using human skin colour could reduce chances animals or objects being detected as human face. However, skin colour can be vary based on the light reflected on human skin and colour that captured on camera is relies on illumination condition, different light temperature will makes skin colour looks difference, thus causing face detection sensitive to illumination condition and not suitable to use in certain condition. Moreover, the face detection will not works on grayscale images since the images do not contain human skin colour other than a scale of black and white. Or, if the light source cause skin colour to reflect differs from normal skin tone colour could cause the entire face ignore by face detection algorithm. Figure 2.1 & 2.2 show same person under different lightning condition, the skin colour is determined by the temperature of the light reflected on.

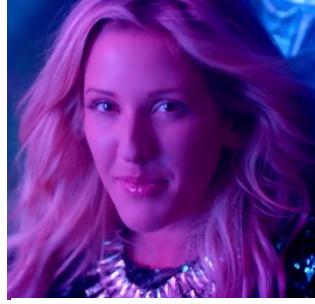


Figure 2.1 Skin Colour – Unusual Lighting.



Figure 2.2 Skin Colour – Normal Lighting.

Since the face detection method used in this paper, the Viola-Jones face detection algorithm, is an appearance-based face detection and it does not identify faces via human skin colour instead it normalizes both training and input samples to greyscale, so it does not expose to the problem. Viola and Jones (2001) proposed face detection framework offer reliable feature selection and real-time detection. By calculating the difference in pixel intensity between features, a knowledge base is formed. For example, in human face the lips region is darker than the nose bridge. The information are then used to train classifier that categorize face and non-face.

2-2 Feature Extraction Technique

Sirovich and Kirby (1987) published a paper indicated that principle component analysis (PCA) can be used to reduce database of faces to form a face with foundation of facial features. In other words, PCA finds the data that best describe the variance of facial feature among faces. In this project, PCA were used to extract features from training samples.

Chai et. al. (2009) suggested that by eliminating beard and moustache from image can improve face recognition performance. But in this project, beard and moustache were retained as it's a features that unique to men and could be an important feature that distinguish male and female.

Gottumukkal and Asari (2004) proposed that by splitting the face image into several images and applying PCA technique on the images, they believe that not all of these split images will be affected by illumination intensity or pose, thus increase in recognition rate and accuracy. It is agreeable that the part of the image that is not

affected by illumination intensity is easier to recognize but a face without all of its features can't be considered as a face. Recognizing a face with the lack of some features will remove certain useful information. For example, the distance between eyes and nose. Moreover, variability such as pose makes the face image harder to split due to the difficulties of finding the centre of the face. The method that used in this proposal does not have this kind of problem as it does not exclude any facial features, and does not require to find the centre of the face.

2-3 Recent Research on Gender Classification Methods

Rahman et. al. (2013) proposed gender classification using support vector machine (SVM), it classifies all training vectors by design a hyper plane that segregate classes into 2 or more. In support vector machine (SVM), samples were projected into high-dimensional feature space and efficiently separate them in a non-linear way using kernels. By using support vector machine (SVM), they achieve accuracy of 88%. In this project, support vector machine (SVM) was used to compare with other classification technique.

Mäkinen & Raisamo (2008) conduct an investigation on gender classification using multiple algorithms, neural network, SVM, threshold adaboost, LUT adaboost, mean adaboost, and LBP + SVM. Neural network and SVM show superior accuracy when compared to other algorithms. In some cases, neural network offers better accuracy compared to SVM, others cases SVM offers better accuracy. In short, there's no clear winner between the two algorithms. Two databases were used in the research, FERET database and WWW images. 760 images from FERET database and 3808 images from WWW were used in the experiment. Result for FERET database shows neural network offers better accuracy 92.22% while SVM offers 88.89%. On the other side, the result for WWW images shows SVM offers better accuracy 66.48% compared to neural network 65.95%. This phenomena is possibly caused by number of training samples. To prove this, same database and different number of samples were used as training sets to find out and compare changes in accuracy for all algorithms in this project.

According to Lu & Plataniotis (2002), using a large-scale database could decrease face recognition performance. They introduce a novel clustering method based on a

linear discriminant analysis methodology to solve degradation in face recognition performance. Faces in the database are divide into several simpler subset based on a novel two-stage hierarchical organization structure. By decomposed the database into a set of simpler subset, degradation in face recognition performance can be reduced. A large database could increase error-rate in face recognition due to generalization in trained data but divide database into several subset can solve the problem. This proposal uses similar idea, face database is divided into 2 categories, male and female.

CHAPTER 3 PROPOSED METHOD/APPROACH

3-1 Methodologies

3-1-1 Steps of Gender Classification in Face Recognition

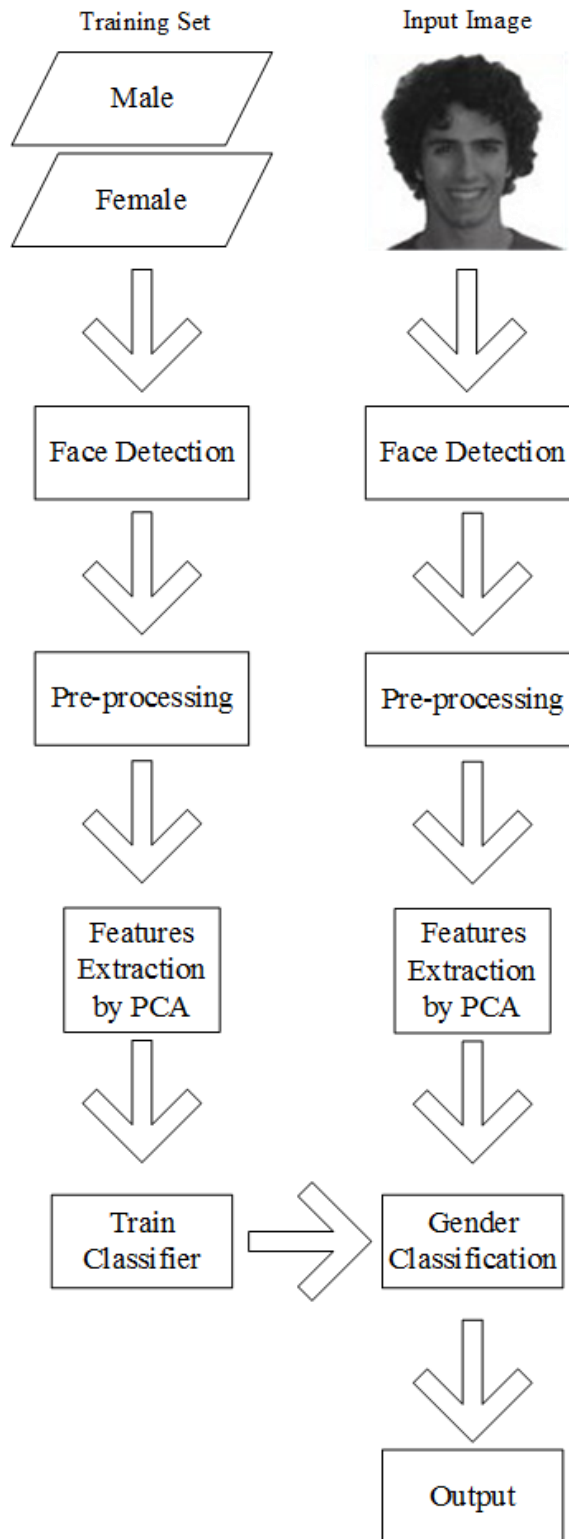


Figure 3.1 Common Approach for Gender Classification in Face Recognition.

3-1-2 Face Database

Three facial databases Aberdeen database, CAS-PEAL and FERET database were used to train the PCA algorithm. 70% of the database images were used as training samples and the remaining 30% were used as test samples.

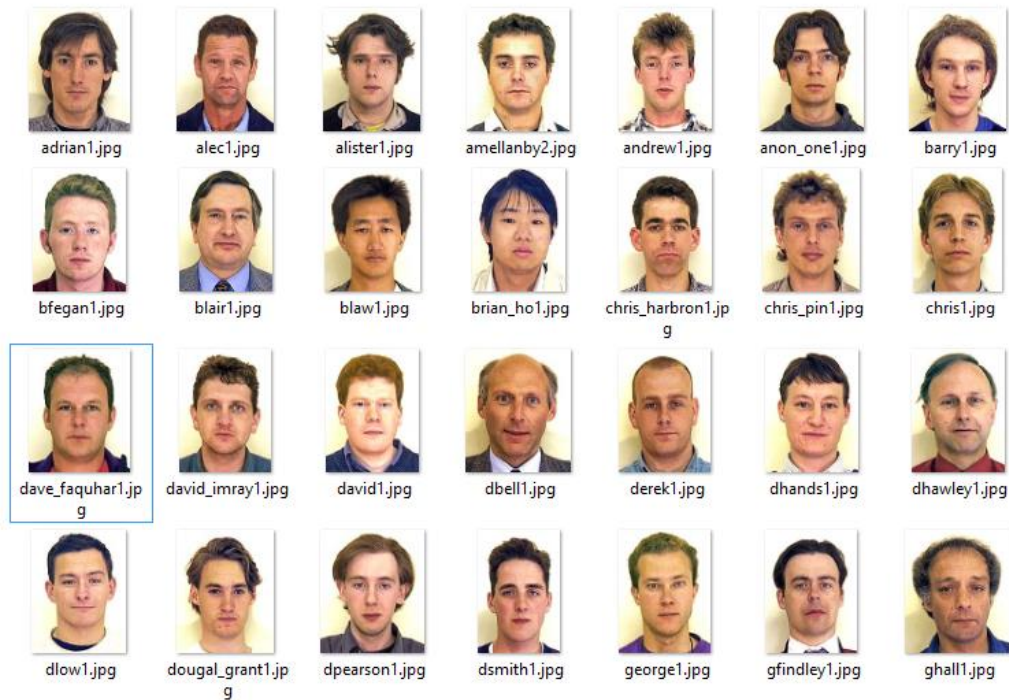


Figure 3.2 Male samples from Aberdeen database.

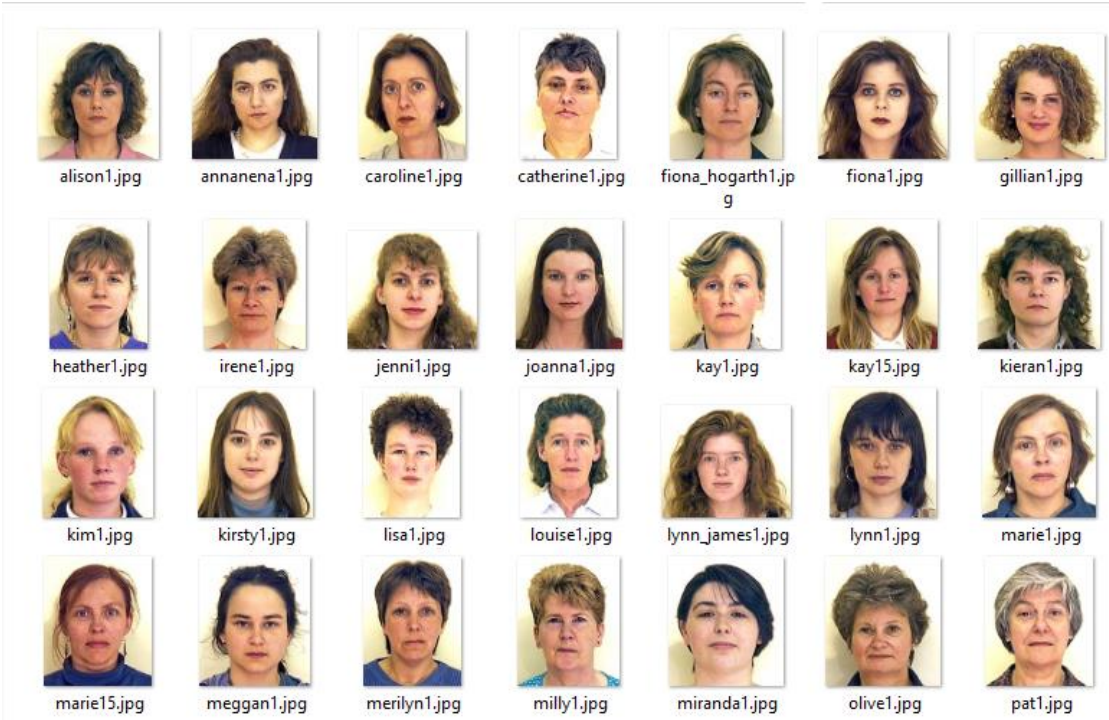


Figure 3.3 Female samples from Aberdeen database.

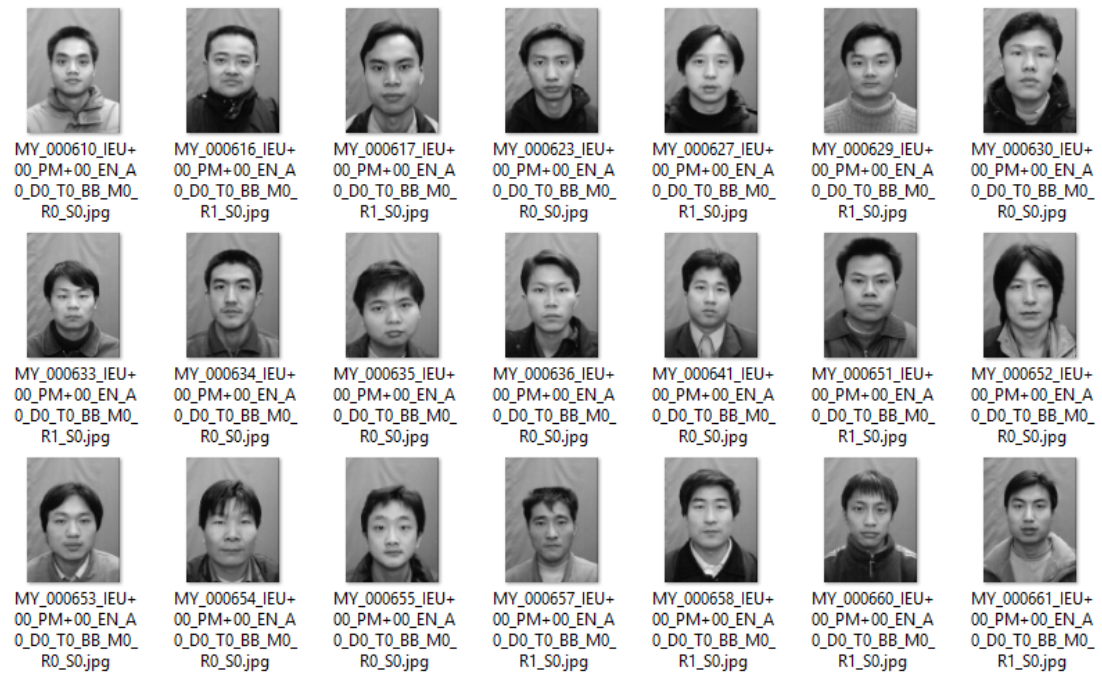


Figure 3.4 Male samples from CAS-PEAL database.



Figure 3.5 Female samples from CAS-PEAL database.



Figure 3.6 Male samples from FERET database.



Figure 3.7 Female samples from FERET database.

3-1-3 Face Detection

Viola-Jones face detection is one of examples of appearance-based face detection. Viola-Jones face detection is an automatic face detection algorithm that offer reliable feature selection and real-time detection because of its fast feature computation. In the proposed system, the Viola-Jones face detection algorithm was used to detect and locate faces from input images and training images.

3-1-4 Pre-processing

Pre-processing is a step where both input and training sets of images will be normalized and had its noise reduced. Pre-processing will go through several image processing technique mention below in following order.

1. Greyscale

Both input and training images will convert to greyscale images where saturation and hue are eliminated and retain only intensity information. Greyscale images has a scale of 0 to 255, where 0 represents black and 255 represents white. Any value in between 0 and 255 will be grey. The purpose of this conversion is to reduce computational cost for up to two-thirds of original cost and normalize images to remove unnecessary variation that can be noise such as image colour temperature. The formula to convert RGB to greyscale image are showed below.

$$Greyscale = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B \quad (3.1)$$

Where R, G and B represents red, green and blue intensities respectively (MathWorks, 2015).

2. Histogram Equalization

Histogram equalization were used to enhance contrast of both input and training images. The purpose of using histogram equalization is to fix the images that were taken in poor lighting condition. Histogram Equalization is an image processing technique that enhances the contrast of images by flatten the intensity of pixels (MathWorks, 2015).

3. Image Resize

Principle component analysis (PCA) has a strict restriction that require training samples to have the same dimensions. Different image size will result different dimensions of data size and therefore unable project into face space. To make sure both input and training image has the same resolution, all image will be resize to 64×64 pixels.

3-1-5 Features Extraction by PCA (Training)

PCA-based face recognition known as eigenface, is one of the most effective techniques to represent faces using statistical method. PCA can lessen dimensions of original data while holding data that best describe the variance of the data. PCA is a traditional method to represent faces, despite its existence for a long time it is still a decent and widely used technique to recognize faces until now. In this step principle component analysis (PCA) will be applied to facial images to extract facial features of faces.

1. Suppose there are M number of training samples, after the pre-processing, the image's pixels that in matrix form will then be convert into a vector for every sample. Each row is transpose and concatenated with its upper and lower row like the equation showed below.

$$I_i = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix}_{N \times N} \xrightarrow{\text{Concatenation}} \begin{bmatrix} a_{11} \\ \vdots \\ a_{1N} \\ \vdots \\ a_{2N} \\ \vdots \\ a_{NN} \end{bmatrix}_{N^2 \times 1} = \Gamma_i \quad (3.2)$$

2. The second step of features extraction is calculate the mean face. The mean face finds the common facial features among all of the training samples. Sum all the samples vectors Γ_i row-by-row and then divide by the total number of sample, its result should be a vector of the mean face.

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (3.3)$$

3. Then, find the unique facial features between all of the samples by subtracting the mean face from all of the sample vectors. This step should remove all of the common features and retain distinguishing features from all samples.

$$\Phi_i = \Gamma_i - \Psi \quad (3.4)$$

4. In this step, the vectors from last step are then combined where each vector is column of the matrix to form matrix A , the covariance matrix C is obtain by simply multiply A by its transpose.

$$C = \frac{1}{M} \sum_{i=1}^M \Phi_i \Phi_i^T = AA^T, \text{ where } A = [\Phi_1, \Phi_2 \dots \Phi_M] \quad (3.5)$$

5. Covariance matrix C , eigenvalues λ_i and eigenvector u_i is expressed as the equation showed below.

$$Cu_i = \lambda_i u_i \quad (3.6)$$

6. In this step, eigenvectors u_i of C need to be compute, but it require large amount of memory to perform cause A is a $N^2 \times M$ matrix, with N represent number of pixels and M represent number of sample. The outcome of AA^T ($N^2 \times M$ multiply $M \times N^2$) would be $N^2 \times N^2$. If sample images have resolution of 64×64 , it will have 4,096 pixels total, thus $N^2 \times N^2$ matrix will have 16,777,216 elements which require large amount of memory to process. However, there is a workaround for this problem, instead of multiply A by A^T , consider multiply A^T by A ($M \times N^2$ multiply $N^2 \times M$) which output is $M \times M$ matrix. After that, find the reduced eigenvectors v_i using the reduced covariance matrix $A^T A$ as equation showed below.

$$A^T A v_i = \mu_i v_i \quad (3.7)$$

Because $C = AA^T$, multiply equation above with A will gives equation below.

$$AA^T A v_i = \mu_i A v_i \quad (3.8)$$

Compare both of the equation 3.7 & 3.8 gives the equation below. Then, use the equation to find the reduced eigenvectors v_i .

$$\begin{cases} u_i = Av_i \\ \lambda_i = \mu_i \end{cases} \quad (3.9)$$

Lastly, find the original eigenvectors u_i using the reduced eigenvectors v_i as showed below.

$$u_i = \sum_{j=1}^M \Phi_j v_{ij} = Av_i \quad (3.10)$$

Since not all eigenvectors u_i contain useful variance, only a portion of eigenvectors u_i with high variance will be used to train classification algorithms.

7. The final step of features extraction is to calculate weights for all of the samples. The weights of all samples will be used by classification methods. To find the weight of all samples, multiply the transpose of eigenvector with adjusted face Φ that calculated in step 3 as the equation showed below.

$$\Omega_i = u_i^T \Phi_i = u_i^T (\Gamma_i - \Psi) \quad (3.11)$$

Combine weight Ω_i of each sample column by column to form a weight matrix W .

$$W = [\Omega_1, \Omega_2, \dots, \Omega_M] \quad (3.12)$$

3-1-6 Features Extraction by PCA (Recognize)

1. After pre-processing, the input image's pixels that in matrix form will then be convert into a vector. Each row is transpose and concatenated with its upper and lower row like the equation 3.2.
2. The input image is then subtract by mean face that calculated in feature extraction training phase step 2.

$$\Phi_{\text{input}} = \Gamma_{\text{input}} - \Psi \quad (3.13)$$

3. To find the input image's weight, multiply eigenvectors u_i from feature extraction training step 6 by the adjusted input image from last step.

$$w_{\text{input}} = u_i^T \Phi_{\text{input}} \quad (3.14)$$

3-1-7 Classification Methods

The weights calculated from features extraction are then used by classification methods. Weights from features extraction in training phase will be used to train the classification algorithms and weights from recognition phase will use as input to the algorithms.

1. Euclidean Distance

Euclidean distance is the most common distance measure metric, it used to find shortest distance between input image and training images in database. In this study, k-nearest neighbors and average distance will use Euclidean distance to measure distance.

$$\|x - y\|_e = \sqrt{|x_i - y_i|^2} \quad (3.15)$$

2. K-Nearest Neighbors (Matlab fitcknn)

K-nearest neighbors is the simplest and fastest learning algorithm when compared with others classification techniques such as logistic regression, support vector machines (SVM), neural network and etc. In classification phase, all of the training samples are projected into multidimensional feature space, distance between input and training samples are then measure by Euclidean distance, k number of sample with shortest distance are selected, the input sample is then assign to most frequent among the selected samples.

3. Average Distance

This classification method is proposed by us to compare with other classification method. The purpose of the algorithm is to find out if it necessary to use advanced classification algorithm such as support vector machine and neural network to solve the gender classification problem. This method uses distance calculated by Euclidean distance to classify the unlabelled vector. The distances between input and male and female samples are calculated separately, the distance to each gender's samples are then sum up and divide by number of sample in the particular gender. The unlabelled vector is then assign to the lowest average distance class.

4. Multinomial Logistic Regression (Matlab mnrfit)

Logistic regression are supervised learning models and it uses probabilistic statistical model to predict classes. It can be binomial which only deals with situations that has two possible outcome or it can be multinomial where more than 2 outcome are allowed. Logistic regression uses hypothesis function to separates the classes and which also known as decision boundary.

$$h_{\theta}(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \quad (3.16)$$

This algorithm require the use of cost function to fit the parameter θ in hypothesis function. Optimization algorithms such as gradient descent is often used to find the optimal parameter θ .

5. Neural Network (NN) with Bayesian Regularization Backpropagation (Matlab Neural Network Toolbox)

The neural network also known as artificial neural network is a supervised learning algorithm that inspired by how biological brain works. Neural network is nothing new and has been exist for years. It research begin in 1943 by Warren McCulloch and Walter Pitts a neurophysiologist and mathematician. It was widely used back in the 1980s and early 1990s, but it fall out of favour since the late 90s. Recently it has becomes the state of the art machine learning technique for many applications. A neural network is form by layer of nodes (or neural network), all of the nodes are interconnected layers by layers.

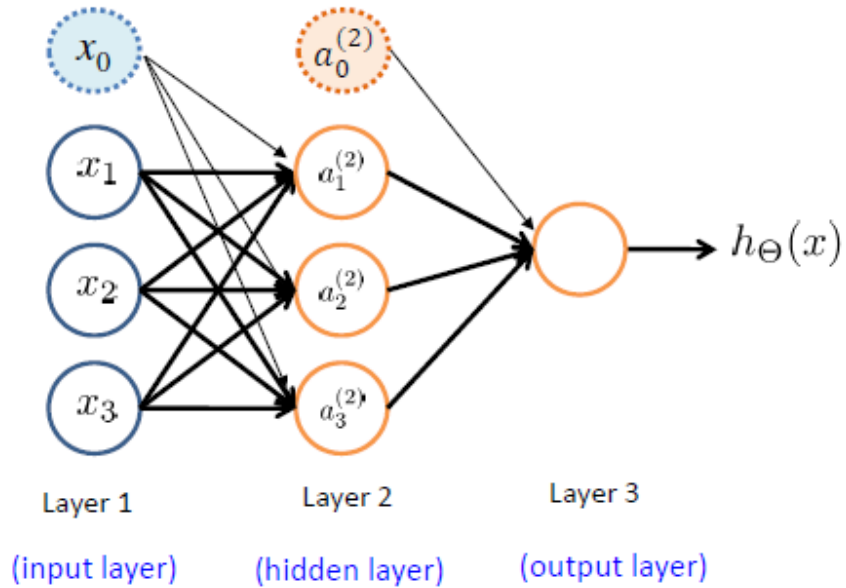


Figure 3.8 Multilayer neural network.

Despite its structural dissimilarity to other classification techniques such as linear and logistic regression, it uses a similar idea to obtain final output. In logistic regression, it uses the sigmoid function to find the hypothesis $h_{\theta}(x)$ output, whereas a neural network uses the sigmoid function to forward propagate to the next layer.

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}} = g(z), \text{ where } z = \theta^T x \quad (3.17)$$

Similarly, a neural network has its cost function to minimize through backward propagation. The purpose of backpropagation is to optimize the weights Θ and minimize error in each node.

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{k=1}^K y_k^{(i)} \log(h_{\theta}(x^{(i)}))_k + (1 - y_k^{(i)}) \log(1 - (h_{\theta}(x^{(i)}))_k) \right] + \frac{\lambda}{2m} \sum_{l=1}^{L-1} \sum_{i=1}^{S_l} \sum_{j=1}^{S_{l+1}} (\theta_{ji}^{(l)})^2 \quad (3.18)$$

6. Support Vector Machines (SVM) using Least Squares Method (Matlab svmtrain)

SVM machine learning algorithm is one of the best supervised learning algorithms. SVM classifies all training vectors by design a hyper plane that segregate classes into 2 or more. In SVM, samples were projected into high-dimensional feature space and efficiently separate them in a non-linear way using least squares. The samples that close to the hyper plane make the classification task difficult to perform. SVM expand the margin of these samples by classes to form a gap. By measuring the shortest distance between input sample and these gap makes classification task easier and less error.

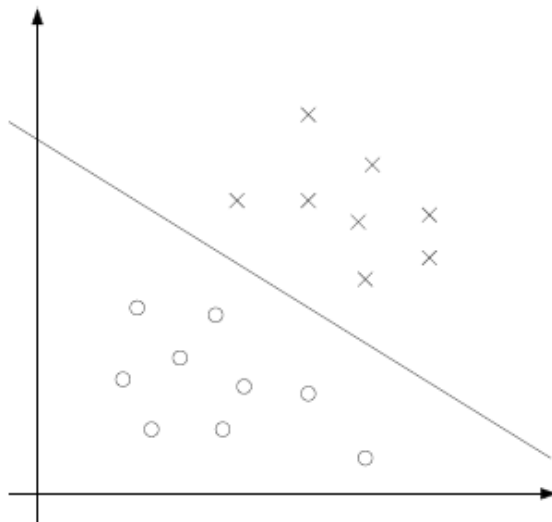


Figure 3.9 SVM – Without Margin

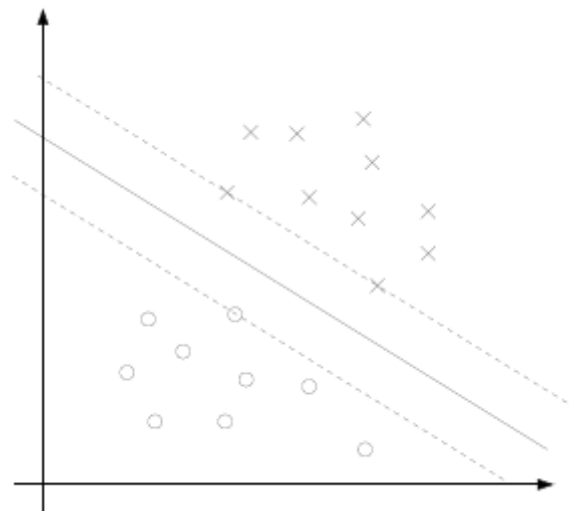


Figure 3.10 SVM – With Margin

3-1-8 Experimental Result & Analysis

The result that we aim to find are the accuracies of each classification algorithm given same and different number of training images used in each test. Another result to gather is the average time required to classify each sample for each algorithm. Each facial database used in this experiment will split into 2 sets, training set and testing sets. 70% of the database images were used as training samples and the remaining 30% used as test samples. In this project, three experiment were performed. The result in this section are produced using algorithms most optimal parameters and implementation in order to compare the 5 algorithms properly. In order to find the most optimal parameters for each algorithm, a range of parameters was used for all of the database used in this experiment. For K-NN algorithm, range from 1 to 5 k sample was used and the best k sample we found is 3. A range of hidden nodes 5,10,15 and 20 was tested on neural network. Result from multiple run show that 10 hidden nodes often yield good accuracy and additional nodes does not affect accuracy. SVM however uses least squares method to separate hyperplane and maximum number of iteration allowed is set to 100,000.

1. The first experiment is to find the highest accuracy among all classification algorithm given multiple facial database. All of the samples in database were used. Accuracy was measured using formula below:

$$\frac{\text{Number of Correct Classification}}{\text{Number of Testing Samples}} \times 100 = \text{Accuracy (\%)} \quad (3.19)$$

The measured accuracy for each algorithm is shown in table below:

Algorithms Database	K-NN's Accuracy (%)	Average Distance's Accuracy (%)	Logistic Regression's Accuracy (%)	Neural Network's Accuracy (%) (*)	Support Vector Machine's Accuracy (%)
Aberdeen	75.00	83.33	91.67	95.83	91.67
CAS-PEAL	83.01	69.87	89.74	90.71	91.35
FERET	79.87	78.86	86.58	88.26	89.60

Average Accuracy (%)	79.29	77.35	89.33	91.60	90.87
-----------------------------	--------------	--------------	--------------	--------------	--------------

Table 3.1 Overall Accuracy

*Due to the inconsistent result produced by neural network, additional runs were recorded. The neural network accuracies shown above are the highest out of all run. Result of neural network additional run can be found in table 3.2.

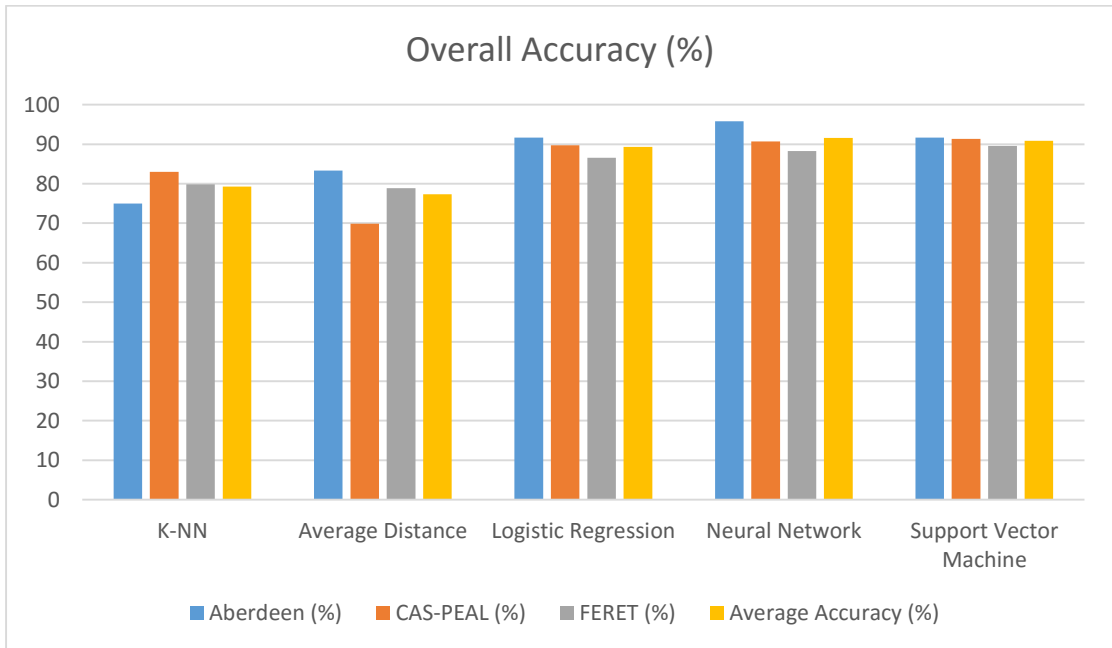


Figure 3.11 Overall Accuracy Bar Chart

Based on the data gathered, we can conclude that

- SVM offer the highest accuracy for all of the large database CAS-PEAL and FERET (1,040 and 994 samples respectively)
- The highest accuracy of all test belong to neural network with accuracy of 95.83% using Aberdeen database (80 samples)
- SVM is superior in term of handling large database compare to other algorithms
- Neural network is superior in term of handling small database compare to other algorithms but result produced by neural network are inconsistent. The reason of neural network result inconsistencies is the algorithm uses random initialized weight. Different initial weight will cause the algorithm to converge to different local or global minima.

Additionally, Bayesian regularization backpropagation neural network is vulnerable to getting stuck at local minima rather than finding the global minima. Result of neural network can be found on table 3.2 below.

Algorithms Database	Neural Network's First Run (%)	Neural Network's Second Run (%)	Neural Network's Third Run (%)	Neural Network's Forth Run (%) (*)	Neural Network's Fifth Run (%)
Aberdeen	95.83	83.33	87.50	83.33	87.50
CAS-PEAL	90.71	89.74	89.74	89.10	89.42
FERET	87.25	87.58	85.24	86.91	88.26

Table 3.2 Neural Network Additional Runs Accuracy

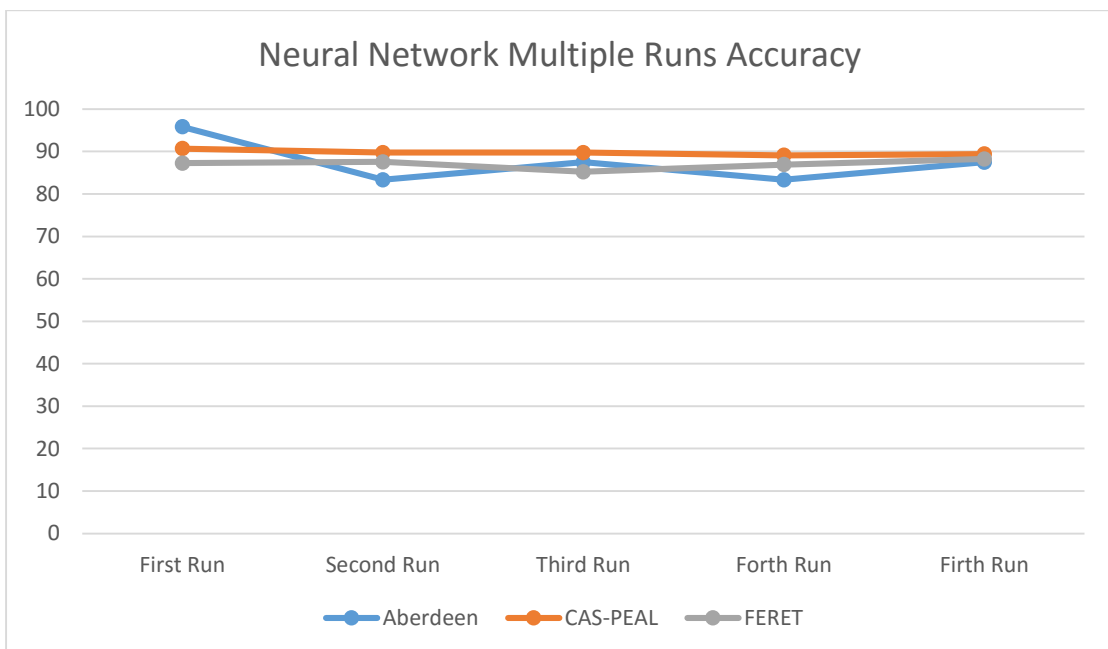


Figure 3.12 Neural Network Additional Runs Accuracy Line Chart

2. The second experiment is to find the fastest classification algorithm. In this experiment, 5 samples from Aberdeen database were used and the average time (millisecond) required to classify a sample is measured and recorded into table below:

Algorithms Database	K-NN's Time (ms)	Average Distance's Time (ms)	Logistic Regression's Time (ms)	Neural Network's Time (ms)	Support Vector Machine's Time (ms)
First Run	1.997	5.838	0.856	7.145	0.416
Second Run	1.636	5.243	0.477	6.340	0.366
Third Run	1.647	5.467	0.508	6.397	0.369
Forth Run	1.602	5.217	0.546	6.418	0.399
Firth Run	1.644	5.195	0.532	6.676	0.373
Average Time (ms)	1.7052	5.3919	0.5838	6.5952	0.3846

Table 3.3 Average Classification Time

*The algorithm with lowest average time will be consider as the best in term of performance.

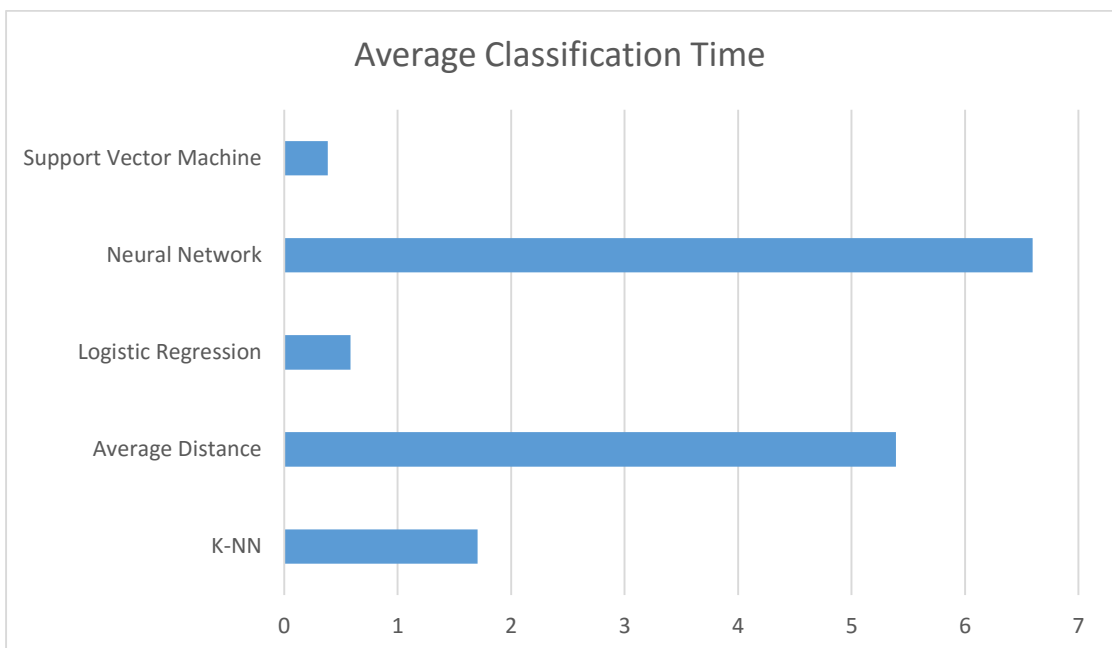


Figure 3.13 Average Classification Time Bar Chart

Based on the result, we can conclude that SVM require the least time to classify samples and neural network took the longest. The second fastest classification algorithm is logistic regression, this is expected because logistic regression and support vector machine uses alike technique to classify samples.

3. The third experiment is to find if increase in number of training samples affect the accuracy, positively or negatively. Secondly, the experiment also finds the algorithm that can handle large database while retain reliable accuracy.

Accuracy for each facial database were measured and recorded in table below:

# of Samples Algorithms	25% of Training Set (Aberdeen)	50% of Training Set (Aberdeen)	All Samples (Aberdeen)
K-NN's Accuracy (%)	79.17	79.17	75.00
Average Distance's Accuracy (%)	79.17	83.33	83.33
Logistic Regression's Accuracy (%)	79.17	79.17	91.67
Neural Network's Accuracy (%)	75	58.33	95.83
Support Vector Machine's	79.17	79.17	91.67

Accuracy (%)			
--------------	--	--	--

Table 3.4 Aberdeen Dataset

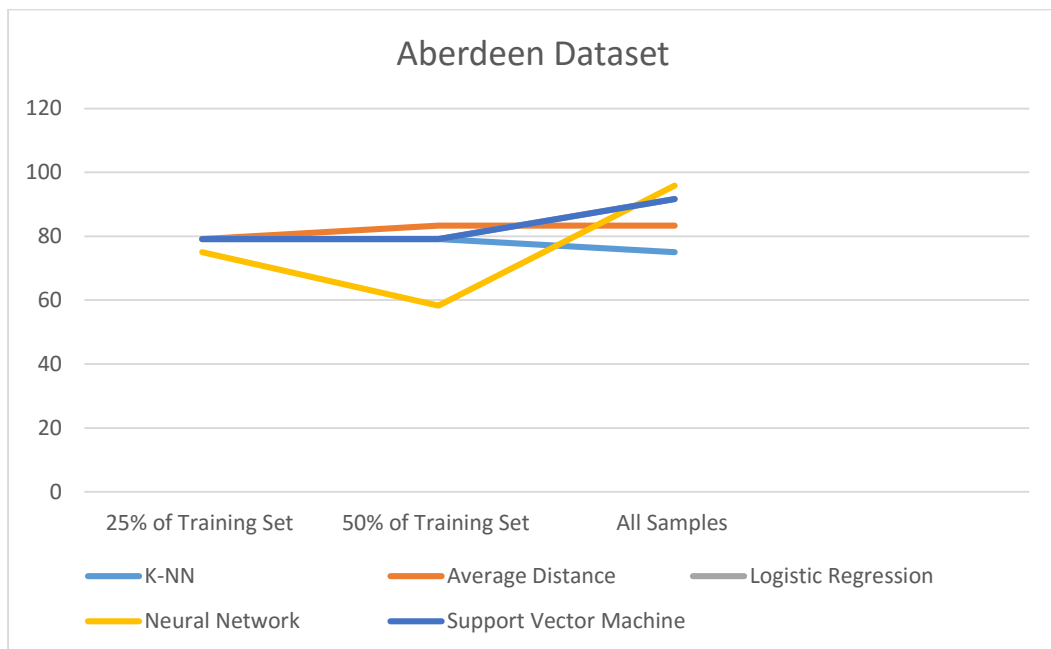


Figure 3.14 Aberdeen Dataset Line Chart

# of Samples Algorithms	25% of Training Set (CAS-PEAL)	50% of Training Set (CAS-PEAL)	All Samples (CAS-PEAL)
K-NN's Accuracy (%)	81.41	84.62	83.01
Average Distance's Accuracy (%)	69.87	68.91	69.87
Logistic Regression's Accuracy (%)	88.14	86.54	89.74

Neural Network's Accuracy (%)	86.85	89.10	90.71
Support Vector Machine's Accuracy (%)	88.14	91.35	91.35

Table 3.5 CAS-PEAL Dataset

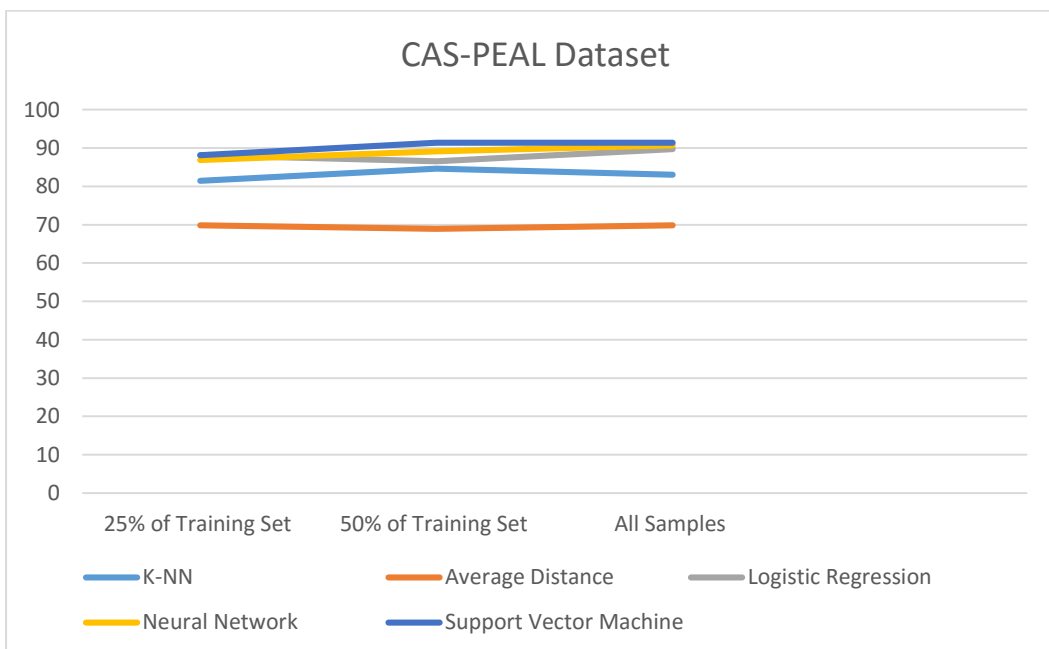


Figure 3.15 CAS-PEAL Dataset Line Chart

Algorithms	# of Samples	25% of Training Set (FERET)	50% of Training Set (FERET)	All Samples (FERET)
K-NN's Accuracy (%)		79.53	79.87	79.87

Average Distance's Accuracy (%)	78.86	78.19	78.86
Logistic Regression's Accuracy (%)	83.22	83.56	86.58
Neural Network's Accuracy (%)	83.89	84.90	88.26
Support Vector Machine's Accuracy (%)	83.89	87.25	89.60

Table 3.6 FERET Dataset

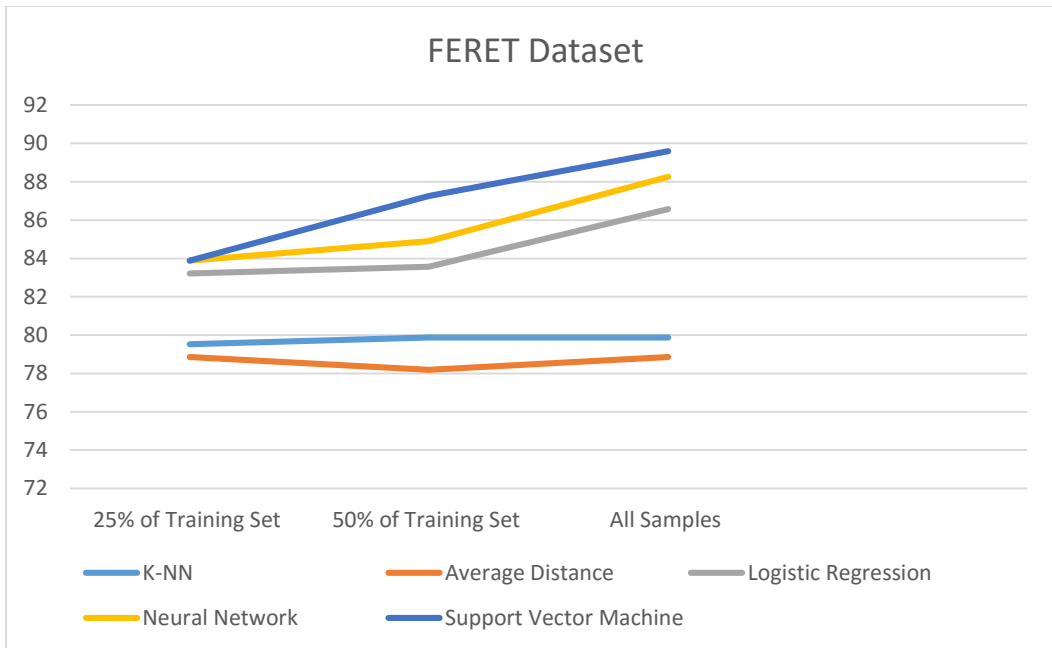


Figure 3.16 FERET Dataset Line Chart

Logistic regression, neural network and support vector machine accuracies increases as number of training sample in dataset increases. This occur because increase in number of samples improve the information the algorithms had on both of the classes male and female which allow better fitting of training data. All classification algorithms accuracies increase as more samples are used in training indicates that the algorithm handle large database reliably. K-NN algorithm accuracy is very consistent when different number of dataset is used, the reason to that is it only requires k number of samples to classify test sample rather than all samples in database.

3-2 Tools

3-2-1 MATLAB

In this project, a face recognition that can classify gender was built along with at least 4 of the classification method, the k-NN, logistic regression, support vector machine (SVM), neural network (NN) and the proposed method. The code was written using MATLAB, a high-level language for technical computing. The main reason for choosing MATLAB is its vast and well written libraries that offer variety of computational algorithms that suit the needs of this project such as the Viola-Jones face detection algorithm (MathWorks, 2015).

3-3 Implementation Issue and Challenges

Logistic regression, neural network and support vector machine share the same shortcoming, they all vulnerable to overfitting in some condition. Overfitting occurs when there are too many features to learn and not enough samples. An overfit model usually have poor prediction accuracy when it try to make predictions on data that not in training set. Most of the time overfitting can be avoided by adding regulation techniques to the classification algorithms but it does not guarantee to work every time.

3-4 Timeline

The gender classification system prototype with k-NN and average distance algorithms has been built before the end of last semester. Implementation of the whole actual system, data analysis has finish as shown in the Gantt chart below.

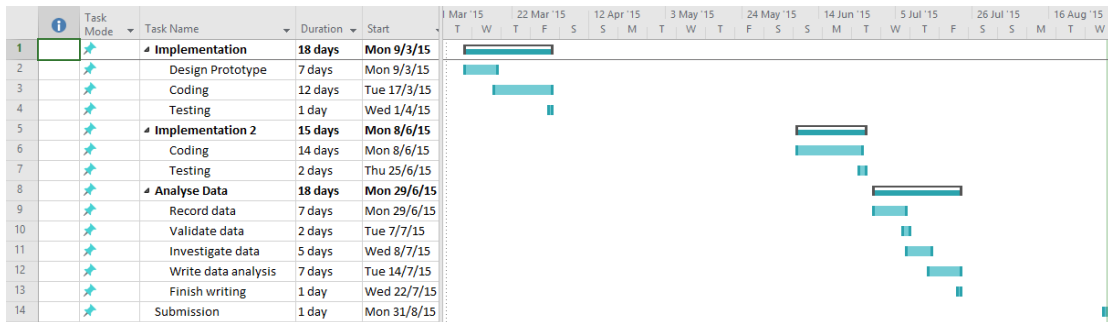


Figure 3.17 Gantt chart

CHAPTER 4 CONCLUSION

In this paper, we proposed to perform analysis on multiple classification techniques, the k-nearest neighbors, average distance, logistic regression, neural network and support vector machine. The intention of this research is to find out which technique works best with large amount of training samples. This research also finds the classification techniques that offer highest accuracy or performance or both. In order to achieve objectives mention in this paper, a gender classification system is built. In the system, standard PCA-based face recognition procedure is applied. The only different is the classification technique used. Classification accuracy for each technique are measured, recorded and compare with each other to find the algorithm that offer highest accuracy, or best performance, or best at handle large training samples. Based on the data gathered, we conclude that support vector machine is superior in terms of accuracy and performance comparing with other algorithms. Neural network also shown excellent accuracy in some cases but not always. Lastly, all of the algorithms handle large database reliably and accuracy increase as number of training samples increase.

REFERENCE

- Chai, T. Y., Rizon, M., Woo, S. S. & Tan, C. S., 2009. Facial Features for Template Matching Based Face Recognition. *American Journal of Applied Sciences*, vol. 6, no. 11, pp. 1897-1901.
- Dautenhahn, K., 2007. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1480, pp. 679-704.
- Goldstein, A. J., Harmon, L. D. & Lesk, A. B., 1971, ‘Identification of human faces’, *Proceedings of the IEEE*, vol. 59, no. 5, p. 748.
- Gottumukkal, R. & Asari, V. K., 2004, ‘An improved face recognition technique based’, *Pattern Recognition Letters*, vol. 25, no. 4, p. 429.
- Lu, J. & Plataniotis, K. N., 2002, ‘Boosting Face Recognition On a Large-Scale Database’, *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol 2, p. 109.
- Mäkinen, E. & Raisamo, R., 2008, ‘An experimental comparison of gender classification methods’, *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1544-1556.
- DigInfo TV, 2012. *Marketing service uses facial recognition tech to estimate gender, age, and visiting frequency*. Available from: <<http://www.diginfo.tv/v/12-0209-ren.php>>. [25 January 2015].
- MathWorks, computer software 2015. Available from: <<http://www.mathworks.com>>. [25 January 2015].
- MathWorks 2015, *Convert RGB image or colormap to grayscale - MATLAB rgb2gray*. Available from: <<http://www.mathworks.com/help/matlab/ref/rgb2gray.html#buiz8mj-7>>. [25 January 2015].
- MathWorks 2015, *Enhance contrast using histogram equalization - MATLAB histeq*. Available from: <<http://www.mathworks.com/help/images/ref/histeq.html>>. [5 August 2015].

- MathWorks 2015, *Fit k-nearest neighbor classifier - MATLAB fitcknn*. Available from: < <http://www.mathworks.com/help/stats/fitcknn.html>>. [5 August 2015].
- MathWorks 2015, *Multinomial logistic regression - MATLAB mnrfit*. Available from: < <http://www.mathworks.com/help/stats/mnrfit.html>>. [5 August 2015].
- MathWorks 2015, *Train support vector machine classifier - MATLAB svmtrain*. Available from: < <http://www.mathworks.com/help/stats/svmtrain.html> >. [5 August 2015].
- MathWorks 2015, *Bayesian regularization backpropagation - MATLAB trainbr*. Available from: < <http://www.mathworks.com/help/nnet/ref/trainbr.html> >. [5 August 2015].
- Rahman, H., Chowdhury, S., & Bashar, A., 2013, 'An Automatic Face Detection and Gender Classification from Color Images using Support Vector Machine', *Journal of Emerging Trends in Computing and Information Sciences*, vol. 4, no. 1, pp. 5-11.
- Sirovich, L. & Kirby, M., 1987, 'Low-dimensional procedure for the characterization of human faces', *Journal of Optical Society of America*, vol. 4, no. 3, p. 519.
- Tathe, S. V. & Narote, S. P., 2012, 'Face detection using color models', *World Journal of Science and Technology*, vol. 2, no. 4, p. 182.
- The New York Times 2004, *What Wal-Mart Knows About Customers' Habits*. Available from: < http://www.nytimes.com/2004/11/14/business/yourmoney/14wal.html?_r=0>. [21 January 2015].
- The University of Texas at Austin, 1998. *IN MEMORIAM WOODROW W. BLEDSOE, The University of Texas at Austin*. Available from: <<http://www.utexas.edu/faculty/council/1998-1999/memorials/Bledsoe/bledsoe.html>> [20 November 2013].
- Viola, J. & Jones, M., 2001, 'Rapid Object Detection using a Boosted Cascade of Simple Features', *Conference on Computer Vision and Pattern Recognition*.
- Yang, MH & Ahuja, N, 2001, *Face Detection and Gesture Recognition for Human-Computer Interaction*, Springer Science & Business Media, Boston.