CHAPTER 1

INTRODUCTION

Face recognition is a basic human characteristic that connects people by communication and interactions with each other. This project is mainly working on video based face recognition which is the extension of the static image face recognition based on video frame. In this research, an automated video based face recognition helps to find and count the appearance of the selected identity within the video surveillance. However, to realize such system is not a simple task as there are many factors need to be taken into account. Therefore, in this research, the focus is given to develop a basic similiar system which is famous people recognition system. This system can be expanded or improved later to become a real-world system.

Purpose of this research is to come out with an initial basic system for an automated video face recognition which is able to identify and count the appearance of the chosen person and display the location of the identity in video. Advantages of the system is able to minimize the mistakes of the manual works and increase the speed of the video processing recognition. This chapter will briefly describe the background of face recognition, problem statement, objectives and scopes of the project, and the thesis outline.

1.1 Background

Face recognition is a biometric method that is direct and simple for human beings to recognize a person. Recognition is easy done by human being which is different from the recognition by computer. Computational models of the face recognition such as Eigenfaces model includes both theoretical insights and practical applications. Although these models of face recognition are not as strong as the human ability, the techniques are quite remarkable (Turk and Pentland, 1991a).

Variety of issues such as criminal identification, image and film processing, security systems, and human-computer interaction can be implemented together with these computation models. For example, the ability to distinguish a modeled face of criminal person from a large number of stored face models of criminal identity. Police are able to identify the criminal person if the face is found in the storage of face models. Despite the complexity, multidimensional images, and meaningful visual stimuli, computational approach of face recognition can be fast and accurate (Turk and Pentland, 1991a).

Computer recognition of faces is biologically implementable. Biometrics approach uses the individual personal characteristics to verify one person's identity. It is applied together with the automated methods to measure and analyze an individual identity using either physiological or behavioral characteristics. Implementation of biometric technologies included face recognition, fingerprints identification, iris identification, voice recognition, signature recognition, retina identification, and DNA sequence matching. Biometrics technologies are good in authentication system as it is by nature and is not easy to be stolen by others (Huang, 1998).

Face recognition has more advantages over other biometric methods as both ease of use and convert use especially for police surveillance. Face images enable more easy audits and verification performed by human operators when using the biometrics records. While other biometric methods such as fingerprints and hand geometry identificatin may have problem when skin cut on finger and damage of epidermis tissue to hand. Face recognition system is more hygienic and no maintenance acquired as the face is analyzed in a distance (Huang, 1998).

The face recognition can be divided into two parts which is verification and identification process. Face recognition algorithms are developed to solve the problem of both processes. When verification process is taking part, the recognition system is given a face image and a claimed identity. The system will either approve or disapprove the identification. On the other hand, in the authentication process, the system must be able to decide which individual the test image identity is based on the trained images in the database (Rizvi et al., 1998).

Video based face recognition is actually expansion of the still image face recognition. The static images used for recognition are obtained from each of the video frame image. Searching an identity manually throughout thousands of video frames by few personnel becomes a heavy task which is leading to the motivation of this project.

1.2 Problem Statement

Video based face recognition has become one of the most challenging tasks in numerous fields and disciplines especially the cases where there is the security reason to identify certain person in Closed Circuit TV (CCTV) video surveillance system. It is very tiring, tedious and troublesome for everyone to search for an identity from CCTV video surveillance footage. Currently, all the searching jobs are being done manually, where a few personnel will search the Identity of Interest (IoI) within all CCTV footages by looking one by one.

1.3 Objectives of the Project

In the proposed project, activity of IoI searching through CCTV devices is able to be done automatically. Contribution of this research is the identity searching system using face recognition technology for video surveillance system. The objectives in this project are as follow:

i. To develop an Identity of Interest (IoI) system from video sequence that able to recognize selected person and count the number of appearance of that identity in the loading video

ii. To evaluate the performance of well-known face recognition algorithms, which are Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), as the feature extractor

iii. To determine the best classifier among Euclidean Distance, Manhattan Distance, and Neural Network (Learning Vector Quantization Neural Network (LVQNET)) for face matching

1.4 Scope of the Project

Face detection for the video is only limited for one face per video frame. If multiple face image are detected, only the largest face size image will be extracted out and used for testing purpose.

PCA and LDA as the feature extractor in this video based face recognition system are highly dependency on the similarity of the training database to the testing images. Dataset used for training require minimum variations in illumination, brightness, and poses compared to testing images.

1.5 Thesis Outline

In chapter 2, review of journals and papers on face recognition methods for video and static is given. The approach of video based face recognition is mainly about face detection and segmentation of image from video frame and extraction of the features and classification of static images. Hence, techniques of feature extractors and classifiers used in video and static images of face recognition are discussed.

Chapter 3 is based on the details of the proposed video based face recognition method and the system developed in this research using MATLAB GUI. Combination of PCA and LDA techniques and three features classifiers ,i.e., ED, MD, and NN together with MATLAB coding implementation will be covered.

In chapter 4, the experimental results using the database will be given. Results obtained are discussed based on different factors related to recognition rate.

Finally in chapter 5, conclusion and possible directions for future work are given.

CHAPTER 2

LITERATURE REVIEW

Nowadays, multimedia applications are becoming universal due to the increasing technology of microelectronics for the processor performance speed and the capacity of storage devices. Video cameras are broadly used in everywhere for security purpose instead of normal digital cameras. Every mobile phone now is adapted with camera capability for taking photo and video. Video processing turns to be very important as the application of video becomes common in our life. The related background and work are given in this chapter.

2.1 Static Image Face Detection and Segmentation Techniques

Face recognition systems with static still image are reviewed in this research. Face detection in static image by using skin color is proposed by Kumar and Bindu (2006). Different body posture, facial expression, changing in lighting brightness and some uncertain parameters can increase the difficulties in face recognition. Therefore, image preprocessing and segmentation play an important role to eliminate and enhance the image quality before training and testing stages. Qualities of image can be improved through image preprocessing such as filtering and segmentation. The face detection algorithm must able to cope with different conditions and complicated background. Lighting compensation techniques are used to solve the different skin tones and complex background. Signature of various skin regions are classified to special identity for the face subjects individually (Kumar and Bindu, 2006).

The proposed algorithm by Kumar and Bindu (2006) is able to detect the skin regions and crop the face portion out from the detected skin regions. Input color image which is originally in RGB color space component is changed to YCbCr color space component. Y and Cb components from the image are removed resulting the Lighting Compensate Red Component (LCRC). When the LCRC image reverted back to RGB color space from YCbCr color space, red component is the only color remain in the final image. The human skin is red component due to the existence of the blood irrespective of the skin color. The brightness information is excluded from the computation where valid skin color extraction skill is only applied to fixed rules based on the pixel and region. By finding the portion where is skin colored, face region can be detected as skin tone as it is faster with fixed head pose orientation and size. Overview of face detection of this skin illumination compensation model is illustrated in Figure 2.1 Kumar and Bindu (2006).

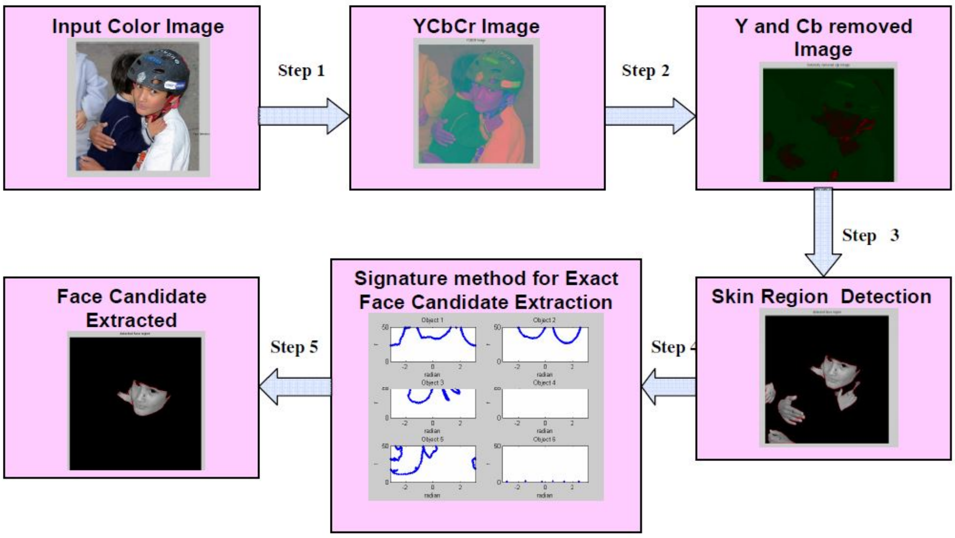

Figure 2.1: Overview of skin illumination compensation model (Kumar and Bindu, 2006)

2.2 Features Extraction Techniques

Features of face recognition can be categorized into geometrical feature based approach and holistic template matching based approach. In geometric feature-based matching system, the recognition can be performed even the information of the individual features are not resolved. Purely geometrical of the remaining details are at a very rough resolution. The geometrical feature based approach extracts and measures the relative location of features such as eyes, nose, mouth and chin. One of the examples are distance between the left and right eye are estimated. The overall configuration can be represented by a vector of numerical data which describing the location and size of the main facial features such as eyes and eyebrows, nose, and mouth. Shape of the face outline had complemented these details (Brunelli and Poggio, 2002).

Holistic template matching considers the face image as a global representation and the face features are extracted from the whole face region. In template matching, the face image is treated as a bi-dimensional array of intensity values. Face features in holistic system is usually extracted by using statistical methods such as PCA and LDA. These techniques are projection based methods which are similiar to template matching that suffers difficulties when the face image has different face orientation, expression, structural components and brightness intensity. This is due to the characteristic of template matching as a direct correlation-based matching technique (Brunelli and Poggio, 2002).

Brunelli and Poggio (2002) had developed and implemented two new algorithms by combination of geometrical features and template matching. The geometrical features are computed based on nose width and length, mouse position and chin shape while the template matching is applied with grayscale images. The implementation of this method is successful under limited conditions which is sensitive to incident illumination and face pose. The recognition system requires to tolerate to certain deviation between the template and the testing image. Direct template matching is computed with high dimensionality which dimension reduction technique has been developed by extracting only the specific features of the face (Brunelli and Poggio, 2002).

2.2.1 Principal Component Analysis (PCA)

Turk and Pentland (1991b) had proposed Eigenfaces as the method for the detection and identification of faces in real-time face recognition system. Eigenfaces is computed by using PCA technique. The method here is considering for two-dimensional and full frontal upright images as testing image. Features in the face images are extracted using PCA and projected onto a face space which captures all variation between the face images as training data. Then, the principal components of the distribution of the faces are identified.

Eigenvectors and Eigenvalues are two important components in PCA. Eigenvectors describe all the difference between the face images while Eigenvalue evaluate the differences of variance between the training images. The greater Eigenvalue for Eigenvectors, the better Eigenvector representing for the variance within the face images. Eigenfaces are referring to Eigenvectors which appear like the ghostly faces and having the same amount as the number of the face images in the training database (Turk and Pentland, 1991b).

Figure 2.2: Average face and some of the Eigenfaces image (Zhao, Chellappa and Krishnaswamy, 1998)

Average face and some of the Eigenfaces images are shown in Figure 2.2 (Zhao, Chellappa and Krishnaswamy, 1998). The face images are restored by weights sum which associated to each of the Eigenfaces. Face recognition system is created by extracting the specific features from a set of face image and use these features to recognize the correct person by analyzing the feature weights required for restoration with the weights contributed to each known persons (Turk and Pentland, 1991b).

Faruqe and Hasan (2009) have implemented PCA as feature extractor and Support Vector Machine (SVM) as feature classifier in the face recognition system. Face image $I(x,y)$ is defined as a vector of two-dimensional N^2 array of intensity values. Assuming an image of size 256 by 256, is constructing a vector of dimension 65,536 or a point in 65,536-dimensional space. All of these images are accumulated to project to a collection of points in large spaces. PCA finds the vectors that best represent the distribution of face images among the entire image space. Descriptions on the procedures to find the vectors of PCA are shown below (Faruqe and Hasan, 2009).

1. Given a training set, S where M is the total number of face images.

$$S = \{\Gamma_1, \Gamma_2, \Gamma_3, \ldots, \Gamma_M\} \tag{2.1}$$

   In Equation 2.1, Γ_1 is formed using the image I (N_x x N_y pixels) which are converted into a vector of dimension size as P x 1 where P =N_x x N_y. Thus, S has a size of P x M.

2. Mean image of the training set, Ψ is calculated using the formula. Ψ has a size of ($Px1$) and n is the number of the face images in Equation 2.2.

$$\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_i \tag{2.2}$$

3. Mean Subtracted Image, Φ_i, the difference between the input image and the mean image is calculated by using Equation 2.3.

$$\Phi_i = \Gamma_i - \Psi \tag{2.3}$$

4. Difference Matrix, A consists large vectors which is subjected to PCA in Equation 2.4.

$$A = [\Phi_1, \Phi_2, \Phi_3, \ldots, \Phi_M] \tag{2.4}$$

5. A set of M orthonormal vectors, u_k which represents the nearest to the distribution of the training data is computed. Orthonormal k^{th} vector, u_k is described in Equation 2.5.

$$\lambda_k = \frac{1}{M} \sum_{n=1}^{M} (u_k^T \Phi_M)^2 \tag{2.5}$$

Maximum of u_k is subject to Equation 2.6.

$$u_l^T u_k = \delta_{lk} = \begin{cases} l & \text{if l = k} \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

where l is the numbering of the training face images, u_k and λ_k are referring to the Eigenvectors and Eigenvalues of the covariance matrix, C. u_l is the Eigenvector of each number of training face image.

6. Covariance matrix, C contains the vectors of size (PxP), which is expected to have (PxP) Eigenfaces.

$$C = A \cdot A^T \tag{2.7}$$

$$C = \frac{1}{M} \sum_{n=1}^{M} \Phi_n \Phi_n^T \tag{2.8}$$

where A is difference matrix in Equation 2.4 and A^T is the transpose of the difference matrix, A.

Eigenvectors, v and Eigenvalues, λ of C are computed by Equation 2.9.

$$Cv = \lambda v \tag{2.9}$$

7. To avoid computing a large vectors, Eigenvector of covariance matrix by size M by M is computed instead of calculating for Eigenvectors for N^2 by N^2. Let $L = A^T \cdot A$, where L is covariance matrix with size of M by M. M Eigenvector, v_l of L is calculated.

   Consider the Eigenvectors of v_l of $A^T \cdot A$ is as shown in Equation 2.10.

$$A^T A v_l = \lambda_l v_l \tag{2.10}$$

   where λ_l and v_l are the Eigenvalue and Eigenvectors of the L. Pre-multiplying both sides by A will result in Equation 2.11.

$$A A^T A v_l = \lambda_l A v_l \tag{2.11}$$

   From which Av_l is the Eigenvectors of C in Equations 2.9 and 2.12.

$$C A v_l = \lambda_l A v_l \tag{2.12}$$

$$C u_l = \lambda_l u_l, \qquad where \quad u_l = A v_l \tag{2.13}$$

   Linear combinations of M training set face images are constructed by these vectors, v_l to form Eigenfaces, u_l in Equation 2.14 where u_l is equal to Eigenvectors of C by Equation 2.13.

$$u_l = \sum_{k=1}^{M} v_{lk} \Phi_k, \qquad l = 1, 2, 3, \ldots, M \tag{2.14}$$

8. To classify the faces, each of the original images is projected into Eigenspace. Weights contributed to each of the Eigenfaces in the reconstruction of input images are computed as in Equation 2.15.

$$w_k = u_k^T (\Gamma - \Psi) \tag{2.15}$$

$$\Omega = [\omega_1, \omega_2, \omega_3, \ldots, \omega_M]^T \tag{2.16}$$

where u_k is the k^{th} Eigenvector and ω_k is the k^{th} weight in the vectors in Equations 2.15 and 2.16.

However, there are few characteristics that affect the face recognition rate using PCA. The number of training images, level of noise and blurriness, image size of face images and different databases are investigated for testing criteria of PCA. The rising of the number of training images increase the recognition rate until a saturation level. The amount of noise and blurriness impact the recognition performance while the different size of image does not bring a significant effect as long as the number of training data is more than total amount of the testing image. Recognition accuracy is improved if the face images are preprocessed with geometric alignment and intensity normalization. Number of the training images which required for recognition are reduced (Poon et al., 2009).

2.2.2 Linear Discriminant Analysis (LDA)

Etemad and Chellappa (1996) had developed a new scheme using LDA technique as feature extraction in spatial and wavelet domains. LDA is carried out via scatter matrix analysis. The important features of different parts of face for each person's identity are identified in LDA through Eigenvector analysis of scatter matrices. LDA aims to maximize the variations between-class scatter matrix and minimize the difference within-class scatter matrix Etemad and Chellappa (1996).

Another techniques which combine both PCA and LDA are proposed by Belhumeur et al. (1997) and Zhao, Chellappa and Nandhakumar (1998). The idea of combining these two methods is to enhance the generalization capability of LDA when there are limited samples per class (Belhumeur et al., 1997). PCA is first used to project face images into the Eigenfaces space for dimensionality reduction and the Eigenfaces vectors are projected into LDA classification space (Zhao, Chellappa and Nandhakumar, 1998). PCA maximizes all the scatters while LDA maximizes the differences between class scatter of the face data and minimizes differences within class scatter of the face data. This hybrid feature extractor using PCA and LDA has improved performance over pure PCA and pure LDA individually (Zhao, Chellappa and Krishnaswamy, 1998).

Dimensionality reduction for maximum discrimination of classes is an important part for LDA. LDA is working on top of PCA technique for the dimension reduction. The outputs of dimension reduction from PCA are applied to derive the within-class scatter matrix and covariance matrix. Separability of classes and dimension reduction are defined by significant Eigenvectors of $S_w^{-1}\ S_B$. $S_w^{-1}\ S_B$ is derived and Eigenvectors with larger Eigenvalues are chosen to be used. Figure 2.3 displays the nine Fisherfaces in the proposed system of (Sahoolizadeh and Ghassabeh, 2008). The first Fisherface has the most discriminatory ability and the last one has the least discriminatory power. Computation of the discriminatory power are described in Equation 2.19 (Sahoolizadeh and Ghassabeh, 2008).

Projection matrix in PCA while its columns are the Eigenvectors of $S_w^{-1}S_B$ is applied to optimize the ratio of determinant of between-class scatter matrix to the determinant of the within-class scatter matrix of the projected samples. The original space is projected to PCA space first before mapped to LDA space. This is to avoid the problem when S_W becomes singular. PCA space becomes the middle person for LDA before transformed to LDA space (Sharkas and Elenien, 2008). To avoid the complicated computation of singular within-class class matrix,S_W, image set is projected to a lower dimensional space first for dimension reduction. This technique is called as Fisherfaces method (Belhumeur et al., 1997).

LDA is dealing to characterize instead of showing the information. For LDA feature vec-

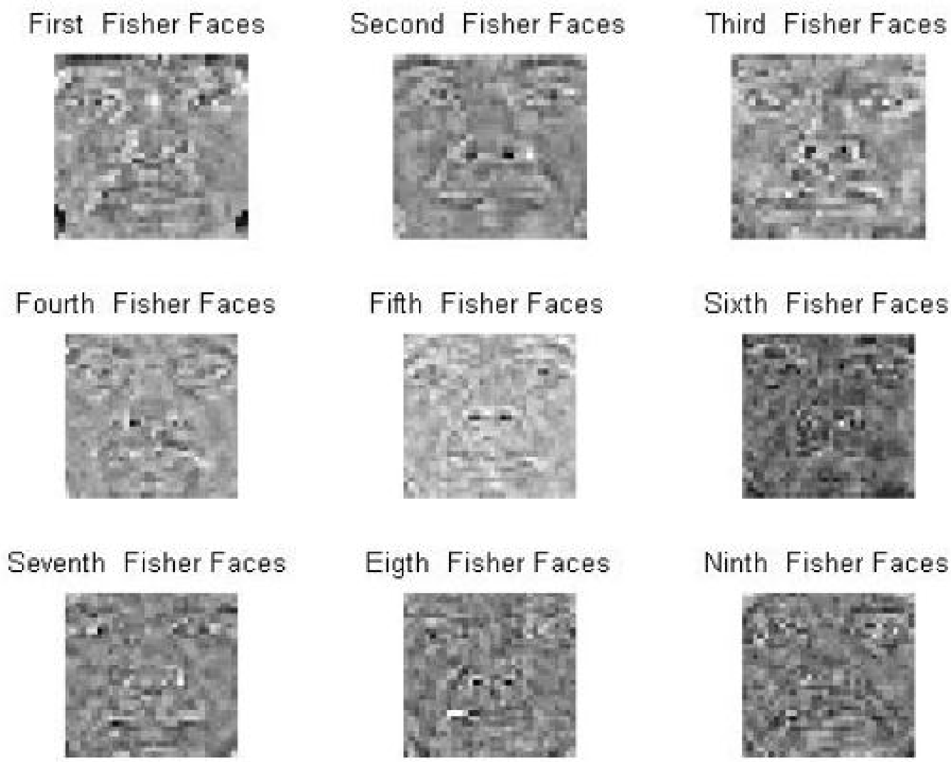

Figure 2.3: Nine Fisherfaces images (Sahoolizadeh and Ghassabeh, 2008)

tors, there are 2 scatter matrices, which are between-class and within-class scatter matrices (Sharkas and Elenien, 2008). Between-class scatter matrix, S_B is shown in Equation 2.17.

$$S_B = \sum_{j=1}^{R} (\mu_j - \mu)(\mu_j - \mu)^T \tag{2.17}$$

where μ_j is the mean of class j while μ is the mean of the all classes and R is the total number of classes.

Within-class scatter matrix, S_W is show in Equation 2.18.

$$S_W = \sum_{j=1}^{R} \sum_{i=1}^{M_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T \tag{2.18}$$

where j and i define the class and the image number and M_j is referring to the total amount of images in class j and the total number of classes, R.

The discriminatory power of the technique is raised by extracting all LDA features. When S_W is not singular, the optimal projection with orthonormal columns, W_{opt} in Equation 2.19 is used to optimize the ratio of det|S_B|/det|S_W| (Sahoolizadeh and Ghassabeh, 2008).

$$W_{opt} = arg \max_{w} \frac{|W^T S_B W|}{|W^T S_W W|} = [w_1, w_2, \ldots, w_m] \tag{2.19}$$

where $w_i | i = 1, 2, \ldots, m$ is the set of generalized Eigenvectors of S_B and S_W corresponding to the m largest generalized Eigenvalues $\lambda_i | i = 1, 2, \ldots, m$ shown in Equation 2.20.

$$S_B w_i = \lambda_i S_W w_i, \quad i = 1, 2, \ldots, m \tag{2.20}$$

The maximum number of nonzero generalized Eigenvalues is $R - 1$ as there is only $R - 1$ of the upper bound for between-class scatter matrix where R is the total number of classes (Belhumeur et al., 1997).

2.3 Features Classification Techniques

PCA is implemented as feature extractor and ED and MD are proposed as classification methods in Malkauthekar (2013). MD shows the better performance for the image with the changed angle while ED is good in prediction for the frontal faces than MD through experimental results. Others classifiers such as K-Nearest Neighbor (KNN) (Parveen and Thuraisingham, 2006; Zhang and Chen, 2012; Wang et al., 2008), SVM(Parveen and Thuraisingham, 2006; Guo et al., 2000), and NN (Yesu et al., 2012; Lawrence et al., 1997; Sahoolizadeh and Ghassabeh, 2008) are also discussed in this section.

2.3.1 Euclidean Distance (ED)

ED for images, Image Euclean Distance (IMED) has considered the spatial relationships of pixels. Advantages of this method is that it is applicable with mostly all image classification. ED is used in the matching and face recognition for texture features which have high rejection to noise (Wang et al., 2008).

Euclidean distance between two point, $a_1 = (x_1, y_1)$ and $a_2 = (x_2, y_2)$ are shown in Equation 2.21 (Malkauthekar, 2013).

$$ED(a_1, a_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{2.21}$$

If there are n-dimensions points, $b_1 = (p_1, p_2, \ldots, p_n)$ and $b_2 = (q_1, q_2, \ldots, q_n)$, then ED between ponts b_1 and b_2 is given in Equation 2.22.

$$ED(b_1, b_2) = \sqrt{(p_1 - q_2)^2 + (p_1 - q_2)^2 + \ldots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \tag{2.22}$$

2.3.2 Manhattan Distance (MD)

MD between two points, $a_1 = (x_1, y_1)$ and $a_2 = (x_2, y_2)$ are shown in Equation 2.23 (Malkauthekar, 2013).

$$MD(a_1, a_2) = |x_1 - x_2| + |y_1 - y_2| \tag{2.23}$$

If there are n-dimensions points, $b_1 = (p_1, p_2, \ldots, p_n)$ and $b_2 = (q_1, q_2, \ldots, q_n)$ in Equation 2.24.

$$MD(b_1, b_2) = |p_1 - q_2| + |p_1 - q_2| + \ldots + |p_n - q_n| = \sum_{i=1}^{n} |p_i - q_i| \tag{2.24}$$

2.3.3 K-Nearest Neighbor (KNN)

Parveen and Thuraisingham (2006) had implemented PCA together with KNN in real time face recognition system to reduce the computation time for KNN. Multiple classification approaches SVM, KNN, KNN with PCA, LDA, and LDA with PCA had been testing on benchmark dataset. The better performance of KNN with PCA over SVM and LDA are obtained (Parveen and Thuraisingham, 2006).

KNN classifier is a basic algorithm that captures all data and differentiate testing image by using a similarity estimation. During the experiment, k-nearest neighbors of testing instance to training data are captured. Their class labels are identified and the major label is concluded as the prediction of the testing image. The optimal value of k is based on the data. The noise on classification can be removed with large values of k but at the same time causing the boundaries of between classes to be less obvious. This is difference with Euclidean distance measure which is calculating the distance between each training class with the testing image (Zhang and Chen, 2012).

The minimum distance of the class is classified as the classification of the testing image. However, processing of KNN takes time as all training data matrix require to be compared to KNN testing image and due to the high dimensionality of data. Thus, PCA is implemented as feature extractor together with KNN to speed up the computation process. Detail description of implementation of PCA and nearest neighbor and KNN are described (Zhang and Chen, 2012).

New KNN techniques are proposed based on the evidence theory. Global frequency estimation of prior probability (GE) relies on the overall training data space while local frequency estimation of prior probability (LE) is estimated in neighborhood space. Difference between these two terms is used to come out with the solution to the inequality data without recollecting the sample data for face recognition (Wang et al., 2008).

2.3.4 Support Vector Machine (SVM)

SVM can be applied for both linear and non-linear classification. Linear classifier works with a hyperplane that split the training data into two classes (positive and negative trained examples) with the largest margin. Largest margin hyperplane is identified where the feature space is projected to a higher dimensional space. The support vectors and margin of a linear SVM are described in Figure 2.4. Red circles and black squares are defined as negative and positive instances respectively. The plane in bold line split the positive and negative instances (Parveen and Thuraisingham, 2006).

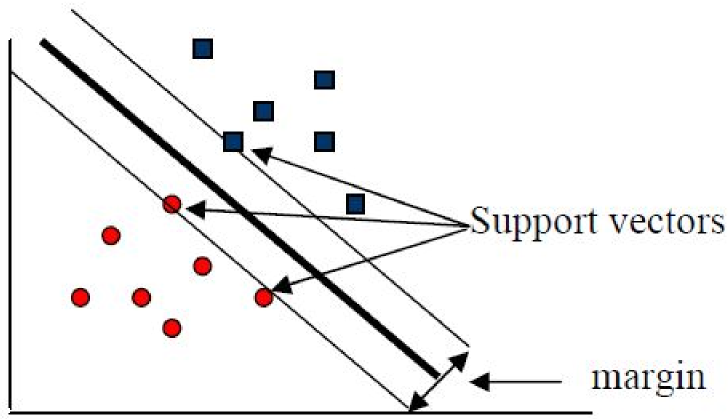

Figure 2.4: Illustration of support vectors and margin of a linear support vector machine (Parveen and Thuraisingham, 2006)

All circles and squares must not be located farer than the minimum distance (margin) from the hyperplane. Support vectors are the points that are located exactly at the distance of margin of the hyperplane. The hyperplane with largest number of the points of same class on the same side is identified by SVM and the distance from all class to the hyperplane is enlarged. For non-linear classification, Kernel trick (kernel function) is used for SVM to largest margin hyperplanes which virtually mapping the testing image into high dimensional feature spaces (Parveen and Thuraisingham, 2006).

New method for pattern recognition implemented with SVM is a bottom-up binary tree classification strategy in face recognition is proposed. This new technique can easily solve the face recognition problem. Decision tree of eight classes in data set are described in Figure 2.5. The numbers 1 to 8 encode the classes without any definition on the ordering. One class number is chosen to be "winner" of the current two classes. The selected classes from the lowest level of the binary tree are compared one by one level for two round of the tests. Unique class is the one appeared on the top of the binary tree. (Guo et al., 2000).

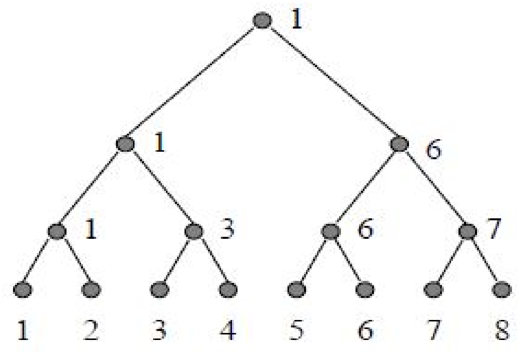

Figure 2.5: Binary tree structure (Guo et al., 2000)

SVM is also implemented together with a wavelet statistics subsystem features that giving the position of eyes and mouth. Wavelet statistics subsystem fasten the processing period for recognition significantly. The method had been tested with different people in multiple of face sizes with changing poses which showed good results with a high speed less than 0.1 per image (Xi et al., 2002).

2.3.5 Neural Network (NN)

Hybrid features using Artificial Neural Network (ANN) are proposed by combining local and global features based approaches (Yesu et al., 2012). The global features are applied with PCA while local features are dealing with standard deviation of eyes, nose and mouth segments of face. ANN are non-linear mapping structures where the idea similar as the networks of biological neurons in human brain. Neurons are the basic computational units that greatly interlinked together. ANN classify and allocate the patterns between training sets and expected results although there are some unknown relationships between each other. The neural network is able to estimate the output of a new testing data.

ANN have advantage in their flexible through learning from the past experience in problem solving strategy. This benefits when the testing data is well prepare but there is little or insufficient information of issue to be settle. Multilayer Feed-Forward Network (MFFN) with multiple layers and more hidden layers is applied. Hidden neurons are the components inside the hidden layers. These hidden layers did the midway assessment prior the transferring information of inputs to output layers. The hidden layer neurons are connected to input layer neurons where the weights on the links are defined as input-hidden layer weights. Output-hidden layer weights are denoted as the hidden layer neurons and the relating weights (Yesu et al., 2012).

Lawrence et al. (1997) had presented a hybrid neural-network solution. The local image sampling, Self-Organizing Map (SOM) neural-network and a conventional neural-network are applied in this system. Quantization of the training samples into a topological face space are provided by SOM. The data that are closer in the original space are also closer in the result space which resulting dimensionality minimization and insensitive to small changes of training sample. Karhunen-Loève (KL) applied with SOM and Multilayer Perceptron (MLP) with the convolutional neural network results. The technique is fast in classification with simple normalization and preprocessing and perform well compared to Eigenfaces method that having training database that varies from one to maximum up to five images per person.

Another face recognition techniques using PCA, LDA and neural network had been proposed by Sahoolizadeh and Ghassabeh (2008). PCA method is first applied to decrease the data dimensionality and then LDA technique is used to extract the important features. NN is used as the feature classifier is to minimize the number of wrong classification due to non-linearly separable classes. This technique has been investigated on Yale face database and the results proven the accuracy of the technique with minimum wrong assessment comparing to other face recognition approaches (Sahoolizadeh and Ghassabeh, 2008).

2.4 Face Detection and Segmentation from Video

Face detection and segmentation are one of the important factor for video recognition as the static images used for recognition are extracted out from the streams of video. Few papers (Srikantaswamy and Samuel, 2006; Chowdhury et al., 2014; Ping, 2008; Xuan and Nitsuwat, 2007) are studied for face detection and segmentation methods in video based face recognition.

Srikantaswamy and Samuel (2006) had proposed an effective method to eliminate all the complicated background first from each frame and a raw face portion is detected using skin color method. Dynamic template matching technique is applied to extract out the face persuasively. This method is not effected by the different in pose and huge scale. Face recognition is handling by PCA and LDA (Srikantaswamy and Samuel, 2006).

Another technique local image gradients are used to flexibly enlarging the window size and it is able to locate all the possible face regions. Directional change in the intensity or color is considered as the image gradients that use to extract important information from images. Local image gradients locate the whole face boundary that describe the left and right sides of the face profile (Chowdhury et al., 2014).

Proposed two elastic window for left and right expand the window upward and downward in the directions guided by the local pixel gradients. These window combine all lines describing the left and right side of face contour to detect the face region within the frame. This technique is less sensitive to the different poses, illumination and sizes of faces as the location of the face is detected by the two elastic windows (Chowdhury et al., 2014).

Before the localization of a face in video frame is carried out, there are no any predefined conditions been made on the pose and face expressions. Prior the face detection process, there are some of preprocessing tasks for each video frame to eliminate out the unwanted noise, background and others. The results came out with the slim binary image which contains only the face regions and some cluttered noises. Discriminant facial features are segmented out from these face images. Multiclass SVM is use as feature classifier in this recognition system (Chowdhury et al., 2014).

Some preprocessing methods on face images had been done to normalize the face image with fixed size. The face image are cropped based on location of the center of two eyes from original face image and rescaled to fix the distance between two eyes. Histogram equalization are also applied on the cropped image. These technique had helped to improve the recognition rate in video based recognition (Gong et al., 2013).

Ping (2008) had proposed a method with three fast cascade face verification models with an ensemble classifier. Face skin, face symmetry, and eye template validation models are implemented to test the detected face from video. These three models are able to remove the oblique faces, the behind part of the head, and all other moving subjects that contain no faces in the video. The face recognition system are only deal with frontal face images and the accuracy are improved by applying the verification thresholds in each models. The ensemble classifier applied together with three different ANN classifiers which well trained by three hybrid feature sets. The recognition is still high even though with low quality video images and overall recognition accuracy and performance had improved with the ensemble classifier

(Ping, 2008).

Another popular technique for face detection is adaptive skin-color model which implemented together with the Eigenface method. Skin-color model is applied to split the skin portion from the background to segment out the face regions for recognition. Skin-color model is using the human face template to decide the location of the face regions for extraction (Xuan and Nitsuwat, 2007)

2.5 Video Based Face Recognition System

Video based face recognition are been investigated for multiple field of algorithms. The complexity of video recognition are increased by some factors that keep changing throughout the stream of the video. Quality of video is also one of the important item in video recognition system. The effect of these factors in video recognition performance had been investigated. Nine factors are investigated by testing the Point-and Shoot Challenge (PaSC) dataset based on three face recognition techniques. The most important factors for accuracy of methods in video are identified (Lee et al., 2014).

Evaluation is also done on four factor metrics that characterize an individual video and two competitive metrics for videos pair. From the analysis, distribution-based metrics is perform well for quantifying parameter values compared to algorithm-dependent metrics. Face detection confidence and face size are important in quality measures. Analysis also showed that males are easier to classify compared to females and Asians are easier compared to Caucasians. Environment and sensor parameters are driving for face recognition technique accuracy for PaSC video dataset (Lee et al., 2014).

The difficulties of video based face recognition are handling multiple frames where the processing in high dimensional data size are more complex than static images. Besides, the face