CHAPTER 1

INTRODUCTION

1.1 Overview

Visual surveillance is one of the active research topic in the field of computer vision, which attempts to detect, recognize and track certain objects from image sequences, and more generally to understand and describe object behaviors, whether the image sequences are obtained from Closed-Circuit Television (CCTV) cameras or other digital video cameras (Hu et al., 2004). Hu et al. (2004) also stated that the aim of visual surveillance is to develop intelligent visual surveillance to replace the traditional passive video surveillance that is proving ineffective as the number of cameras exceeds the capability of human operators to monitor them. In short, an automated surveillance system is desired to accomplish the entire surveillance task as automatically as possible without the need of involving human operators.

Generally, basic automated object recognition system consists of two main components: automated detection, and recognition. In automated detection, the main objective is to find the region of interest from the entire digital image captured by using still camera or video camera correctly without affecting much of the frame rate or in other word, fast detection rate with low error rate. In the case of face detection, this is usually done by identifying the skin color, shape, or facial features based on a model that had been trained offline. After the region of interest is identified from the digital image, the information or features whether in the form of pixel values or other descriptors, are passed to classifier to be assigned with a label which associate to a particular object class. Although there is no limitation on what object to be detected and identified, the main interest of this work is focused on face detection and recognition in

surveillance system.

Face detection (Viola and Jones, 2004; Zakaria and Suandi, 2011; Ma et al., 2013; Artan et al., 2014; Verma et al., 2014) and recognition (Ayad et al., 2014; Wang et al., 2014; Kang et al., 2014; Yi and Su, 2014; Zhu, Hu, Sun and Hu, 2014; Huang and Lin, 2014; Zhu, Zuo, Zhang, Shiu and Zhang, 2014) are current famous biometric techniques which are also the challenging tasks in the computer vision field. This system can be categorized into two types of system: cooperative and non-cooperative system. A cooperative face detection and recognition system refers to the system that is designed to request the cooperation of the subject to provide a near perfect face for the system to operate. This can be done by specifying the distance of the subject from the camera, ample lighting to illuminate the facial features, or even requesting the subject to remove any accessories or hair that might occlude the face. Non-cooperative system, on the other hand, refers to the system that tries its best to detect and recognize the identity of the subjects captured by the camera without needing them to look exactly at the camera. Non-cooperative system is a much more challenging system whereby the chances of obtaining a suitable face for detection and recognition are relatively low compared to cooperative system, and the administrators are unable to request the subjects to face at the camera to obtain another face image. Surveillance camera based face detection and recognition systems belong to the non-cooperative system.

Due to the fact that non-cooperative system is more challenging than cooperative system, a few methods should be introduced to the system to reduce the chance of misclassification: image pre-processing, and watch list. Image pre-processing is a method to normalize the image intensity quality where the image intensity quality might be affected by uncontrolled illumination or even hardware differences. The watch list method listed out a few of the most possible candidates as the recognition output.

Visual surveillance in dynamic scenes has a wide range of potential applications, such as a security guard for communities and important buildings, traffic surveillance in cities and expressways, detection of military targets, etc. (Hu et al., 2004). With the never ending of crime over the nations, law enforcement agencies are starting to look into the possibilities of automated detection and recognition of criminals or suspects in public space, where the face of the suspects might be caught by the surveillance cameras, in order to reduce the crime rate. Figure 1.1 shows the surveillance cameras that are installed at a traffic light junction in Malaysia.

(a) (b)

Figure 1.1: Surveillance cameras installed at traffic light junction in Malaysia: (a) surveillance cameras installed at the middle section of the pole; (b) close up view of the surveillance cameras.

Face detection and recognition in video based surveillance system is more challenging than normal face detection and recognition system that utilizes photo-realistic images (Lei et al., 2009; Wheeler et al., 2010; Ho et al., 2003; Lee et al., 2005). Figure 1.2 shows a sample of frontal pose photo-realistic image from the Surveillance Camera Face Database (SCface) database. The challenges of using commercialized surveillance cameras are the images taken using these cameras are low in frame rate, and low in resolution, but high in noise. The detection and recognition are even more difficult when added with known problems in computer vision such as changes in facial appearance, pose variations, occlusions, uncontrolled lighting conditions, and changes in facial expression. Figure 1.3 shows the facial images captured by a Pan-Tilt-Zoom (PTZ) surveillance camera over a few minutes at $10-20$m range outdoors. The effect of pose variation, illumination variation, and blurring that can occur when imaging non-cooperative subjects can be observed clearly in this figure.

Figure 1.2: Frontal pose image from SCface database (Grgic et al., 2010).

Figure 1.3: Facial images captured by a PTZ surveillance camera over a few minutes at a $10-20$m range outdoors (Wheeler et al., 2010).

Vapnik (1998) and his co-workers proposed Support Vector Machine (SVM) as a very effective method for general purpose pattern recognition. Given a set of points belonging to two classes, SVM finds the hyperplane known as the Optimal Separating Hyperplane (OSH) that separates the largest possible fraction of points of the same class on the same side while maximizing the distance from either class to hyperplane, which in turn minimizes the risk of misclassifying not only to the examples in the training set but also the unseen examples of the test set (Guo et al., 2000).

Zhao et al. (2009) proposed an approach based on multi-class SVM with Radial Basis Function (RBF) as the kernel function to recognize a face in the images by adjusting the penalty factor and kernel parameter to obtain a relatively satisfactory result. Liu et al. (2007) proposed learning based approach to improve the perceptual image quality during video conferencing.

They use a set of professional-taken face images as training examples, and then adjust the color of the input image so that the color statistics in the image is similar to the training examples. This procedure automates the enhancement process and is extremely efficient to compute (Liu et al., 2007).

Face detection in visual surveillance systems can be realized by implementing various techniques, such as template matching (Iso et al., 1996), neural network-based face detection (Rowley et al., 1998), skin color regions (Garcia and Tziritas, 1999), Adaboost face detection (Viola and Jones, 2004), or even the combination of neural network and Adaboost (Zakaria and Suandi, 2011). In recent years, many researches on face detection and recognition had been carried out by researchers around the world for various purposes, e.g. surveillance applications (Bhaskar, 2012), human identification at a distance (Park and Lee, 2004; Maeng et al., 2011; Tan and Kumar, 2012), matching facial composites to mugshots (Klum et al., 2014), gender classifications (Rai and Khanna, 2014) etc. Various techniques had also been studied to improve the performance of face detection and recognition system (Park et al., 2007; Chitaliya and Trivedi, 2010; Huang et al., 2011; Narang et al., 2013; Falcao et al., 2014).

Face detection and recognition in surveillance system is possible, but efforts to improve the recognition rate are still needed as this will reduce the constraints being put to the system. Therefore, this research project is initiated to improve the recognition rate of such system.

1.2 Application Domain

Many face recognition systems have been envisaged over various application domains, and this section explains the possible application domain of face recognition in surveillance system, and some of the challenges that the system might face while in operation.

1.2.1 Public Surveillance

The most interesting application domain of face recognition in surveillance system is face recognition in public surveillance. Video surveillance system contains rich information that can be used to aid police investigation, e.g. vehicle type, vehicle number plate, the process of a certain incident, face of the suspects, etc. For application that needs identity recognition, face is the most important feature that can be used from video surveillance system. As mentioned in Section 1.1, face recognition in surveillance system belongs to non-cooperative system, in which the system can operate without the subject's active participation and conscience (as shown in Figure 1.4). The automated face recognition system can then search for the identity or watch list of the suspect from a database that comprises of millions of subjects.

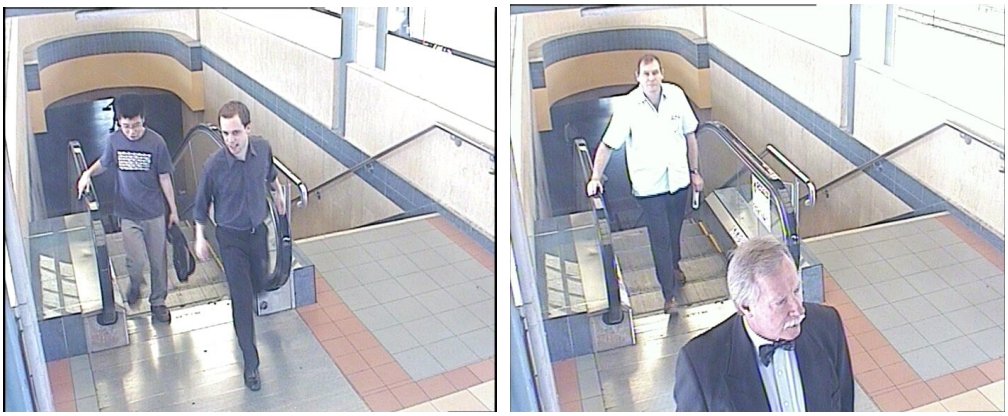


Figure 1.4: Examples of typical face pose under surveillance conditions (Lovell et al., 2008).

1.2.2 Closed Environment Surveillance

Another interesting application of face recognition in surveillance system is in the closed environment surveillance system. Closed environment can be related to the corporate environment, where all the subjects in that environment had already enrolled in the system with frequent presence. The usage can be either like verification system that only allows certain people to enter

that particular area or attendance based system. Unlike the public surveillance type, where the environment might be affected by weather condition and amount of sunlight throughout the day, this system can be setup to have optimum condition for face detection and recognition, e.g. proper lighting, suitable location to install the camera, etc.

1.3 Problem Statement

Face recognition in surveillance system has unlimited possibilities, but it is still very limited to the quality of the input image. This is due to several factors listed below:

1. If a surveillance camera can only provide poor quality images, the difficulties of recognizing the subject will be very high. Sometimes, visual surveillance systems have surveillance cameras of different specifications installed at different locations due to different surveillance purposes. Therefore, an automated face recognition system that can perform under different image qualities is required.

2. Surveillance camera manufacturers manufacture different model of surveillance camera, or even model hardware revision due to component availability which cause differences in input image quality. If different model of camera is installed over a period of time, the input image quality might not be the same, in which some might have different hue, saturation, or even brightness, even though they can be integrated into the same system. Thus, a face recognition system that can even perform under such intensity quality differences caused by hardware differences is a vital requirement.

3. Facial features and lighting condition difference under different distances captured by a fixed surveillance camera. Moses et al. (1994) mentioned that the variations between the images of the same face due to illuminations almost always being larger than image

variations due to changes in face identity. Face region size also plays important role in automated face recognition system as it determines how much information is captured by the camera for recognition purpose. Here, the challenge is to find the most appropriate face size that is suitable for face recognition under different distances between the subject and camera.

4. Most commercialized surveillance cameras capture low resolution images or video to reduce the storage consumption. The resolution in this case is not simply referring to the pixel resolution of the captured image, but it also refers to the spatial resolution. An image with high pixel resolution, but low in spatial resolution will produce a blurred image when compared to an image with higher spatial resolution. This challenge requires facial features be detectable even under low spatial resolution images.

5. Subject recognition or identification based on the face captured from visual surveillance system is very challenging. The uncontrollable elements such as the lighting condition or even the angle of the camera to the face of the subject due to the installation location of the surveillance camera may severely affect the performance of the face recognition system. Therefore, a suitable algorithm should be implemented to ensure that the system can perform well under challenging environment.

1.4 Objectives

The main objectives of this research undertaken here are listed below:

1. To determine the most appropriate image pre-processing technique and face recognition classifiers that are robust against the effect of image intensity quality difference due to camera hardware difference in visual surveillance system.

2. To observe the effect of face size difference based on detected facial features to the performance of the visual surveillance face recognition system.

3. To design and develop a robust automated face recognition algorithm in surveillance system that performs well even if the subject moves towards the surveillance camera by integrating score fusion to form a multimodal algorithm and watch list monitoring principle to list out possible subjects.

1.5 Scope of Thesis

This thesis covers the following scope:

1. This research only considers face images that are easily detected using common face detection algorithm. Database images that have color error are not considered as part of the subjects. This is to evaluate the performance of the face recognition system based on the effectiveness of different image pre-processing techniques.

2. The main database used for this research is SCface database (Grgic et al., 2010), which is a publicly available surveillance camera face database. The images in this database are taken from color surveillance camera and infrared camera with only single face per image. Infrared camera is not included in the performance evaluation. Only color surveillance camera images are considered in this research.

1.6 Outline of Thesis

This thesis is organized into five chapters. In Chapter 1, the general overview and application domain of automated face recognition in surveillance system is presented. It also shows the problem statements, along with the research objectives and scope. Chapter 2 explains and

reviews various image pre-processing techniques, and classifiers. Chapter 3 details the experiments setup, and illustrated the experiment flows using process flow. Chapter 4 exhibits the experiment results, analysis, and discussions. Chapter 5 concludes the research project and suggestions for future studies.

CHAPTER 2

LITERATURE REVIEW

2.1 Background

Digital images are the basis of all digital based applications, such as digital image processing, medical image processing, satellite image processing, visual based object detection and recognition, motion detection etc. Digital images can be simply categorized into two different types: color images, and grayscale images. Color images are usually represented in three intensity components corresponding to the red, green, and blue components of the Red, Green, Blue (RGB) model, whereas grayscale images are represented with only one intensity component. The intensity of each component is represented by an 8 bits value, and when the red, green, and blue components are combined into a single digital color image, it is known as the 24 bits true color images which able to represent $16,777,216$ different colors simultaneously. The information that is stored within the digital images can be used directly or after digital image processing techniques are applied to enhance the image for further application.

In the case of object matching, there are two different types of matching available: verification, and recognition. Verification refers to the one-to-one matching of the object of interest to find out whether the object is in fact itself or not, while recognition (also known as identification) refers to the matching of one-to-many to find out what is the object of interest. The decision of the matching is done by trained classifier(s), either supervised or unsupervised classifiers can be used for that purpose.

Several digital image processing techniques and object classifiers will be discussed in detail in this chapter.

2.2 Image Pre-Processing Techniques

Image pre-processing and feature extraction techniques are mandatory for any image based applications. The accuracy and convergence rate of such techniques must be significantly high in order to ensure the success of the subsequent steps (Jude Hemanth and Anitha, 2012).

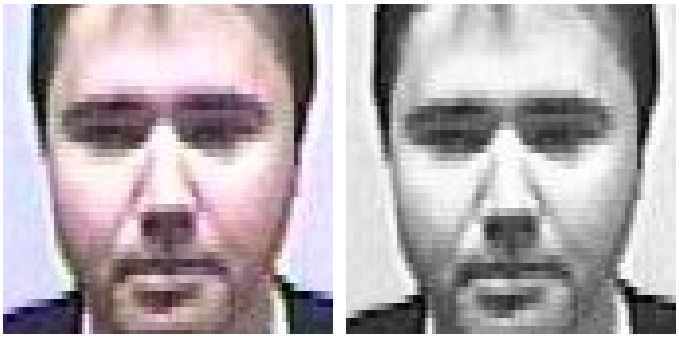
2.2.1 RGB to Grayscale Conversion

The conversion from digital color images to grayscale images is not unique as the use of different photographic filters will have difference in color channels weighting that result in an output with different level of contrast, sharpness, shadows, or even the structure of the image when color images is converted to grayscale images. Saravanan (2010) proposed a new algorithm for color image to grayscale image conversion, and discussed on several linear and nonlinear color image to grayscale image conversion techniques.

The main computer vision library utilized in this research is Open Source Computer Vision (OpenCV) library which utilizes the equation below for the conversion of color image to grayscale image. RGB to grayscale conversion image is shown in Figure 2.1.

$$Y = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B \tag{2.1}$$

where Y is the luminance, R, G, and B stands for the intensity value of red, green, and blue channel, respectively.



(a) (b)

Figure 2.1: RGB to grayscale conversion: (a) RGB face image; (b) Grayscale face image.

Alternative color space conversions include Hue, Saturation, Lightness (HSL) and Hue, Saturation, Value (HSV) color space conversion, whereby the grayscale image is obtained by taking only the lightness of HSL color space or value of HSV color space.

For HSL color space, let $M = sup(R,G,B)$ and $m = inf(R,G,B)$, where $sup(.)$ is the supremum of R, G, and B, $inf(.)$ is the infimum of R, G, and B, R, G, and B is the red, green, and blue pixel value, respectively in the range of $[0,1]$. Hence, the equation for each channel is given as below:

$$L = \frac{M+m}{2} \tag{2.2}$$

$$S = \begin{cases} \frac{M-m}{M+m} & L \leq \frac{1}{2} \\ \frac{M-m}{2-M-m} & L > \frac{1}{2} \end{cases} \tag{2.3}$$

$$H' = \begin{cases} undefined & M-m=0 \\ \frac{G-B}{M-m} & R=M \\ \frac{B-R}{M-m}+2 & G=M \\ \frac{R-G}{M-m}+4 & B=M \end{cases} \tag{2.4}$$

$$H = H' \times 60^\circ \tag{2.5}$$

For HSV color space, the equation for H is the same as HSL color space, which is given in Equation (2.4) and (2.5). The equation for the other channels are given below:

$$V = M \tag{2.6}$$

$$S = \begin{cases} \frac{M-m}{M} & M \neq 0 \\ 0 & otherwise \end{cases} \tag{2.7}$$

2.2.2 Histogram Equalization

Histogram Equalization (HE) (Nixon and Aguado, 2008; Gonzalez and Woods, 2008) is a non-linear contrast adjustment technique intended to highlight image brightness based on image histogram for visual analysis. HE aims to change a picture by remapping the histogram of the image to produce a flatter histogram image. The intensity is redistributed to form a histogram that has a near uniform probability density function, meaning that the peaks and valleys of the original histogram will be shifted after the equalization (Zakaria et al., 2010).

Cameras and image sensors are assembled with different hardware sensors, in which the exposure to the resulting light of the scene and the contrast of the scene will be affected. The shutter and lens aperture setting are being juggled between exposing the sensors to too much or too little light in standard camera, where the contrast is often too much for the sensors to deal with; hence trade-off between capturing dark and bright areas are taken into account. There is nothing can be done about the data recorded by the sensors after the picture has been taken, however the dynamic range of the image can still be extended using HE (Bradski and Kaehler, 2008).

Let f be an image represented as M by N matrix of integer pixel intensities ranging from 0 to $L-1$ where L is the number of possible intensity values, which is normally 256 for 8-bits digital images, and p denote the normalized histogram of f with a bin for each possible intensity which can be represented by the equation shown below (Gonzalez and Woods, 2008):

$$p_n = \frac{\text{number of pixels with intensity } n}{\text{total number of pixels}} \qquad n = 0, 1, \ldots, L-1. \tag{2.8}$$

The histogram equalized image g will be defined by

$$g_{i,j} = \left\lfloor (L-1) \sum_{n=0}^{f_{i,j}} p_n \right\rfloor, \tag{2.9}$$

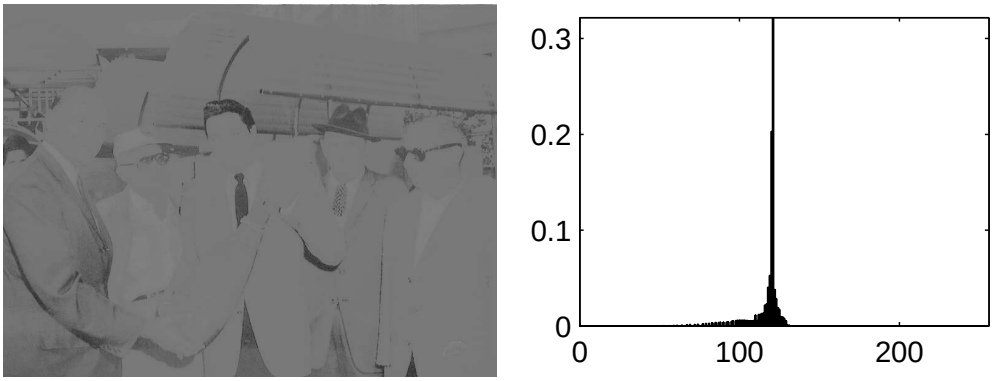
where $\lfloor \cdot \rfloor$ rounds down to the nearest integer. This is equivalent to transforming the pixel intensities, k, of f by the function

$$T(k) = \left\lfloor (L-1) \sum_{n=0}^{k} p_n \right\rfloor. \tag{2.10}$$

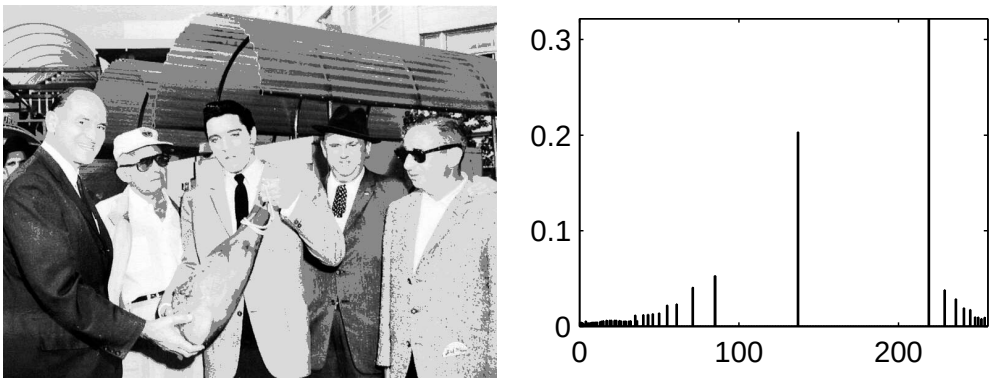
The motivation for this transformation comes from thinking of the intensities of f and g as continuous random variables X, Y on $[0, L-1]$ with Y defined by

$$Y = T(X) = (L-1) \int_X^0 p_X(x) dx, \tag{2.11}$$

where p_X is the probability density function of f. T is the cumulative distributive function of X multiplied by $(L-1)$. Figure 2.2 is the example of image and its histogram after applying HE.



(a) Original image (b) Original histogram



(c) Transformed image (d) Transformed histogram

Figure 2.2: Example of image after applying histogram equalization (Gonzalez and Woods, 2008).

2.2.3 Contrast Limited Adaptive Histogram Equalization

Contrast Limited Adaptive Histogram Equalization (CLAHE) (Pizer et al., 1987; Zuiderveld, 1994; Pizer et al., 1990) has produced good results on medical images (Reza, 2004). This method is proposed by Pizer et al. (1987) and summarized by Zuiderveld (1994), in which the original image is divided into several non-overlapping contextual regions also known as tiles, of almost equal sizes before HE was made on each of these sub regions. Any artificially induced boundaries are eliminated by applying bilinear interpolation to combine the neighboring tiles. Thus, better contrast and more accurate results can be obtained. Noise removal step is done beforehand to avoid enhancing the noise (Zhao et al., 2010).

The equation of three different histogram distributions, namely the uniform distribution, exponential distribution, and Rayleigh (Gaussian) distribution as the basis for creating the contrast transform function are given below respectively:

$$g_{uniform} = \Big[g_{max} - g_{min}\Big] P(f) + g_{min} \tag{2.12}$$

$$g_{exponential} = g_{min} - \frac{1}{\alpha} ln\Big[1 - P(f)\Big] \tag{2.13}$$

$$g_{Rayleigh} = g_{min} + \sqrt{2\alpha^2 ln\left(\frac{1}{1 - P(f)}\right)} \tag{2.14}$$

where g is the computed pixel value, g_{min} is the minimum pixel value, g_{max} is the maximum pixel value, α is the distribution parameter, and $P(f)$ is the cumulative probability distribution. Figure 2.3 is the example of image and its histogram after applying CLAHE.

2.2.4 Learning-Based Color Tone Mapping

Learning-based Color Tone Mapping (CTM) is proposed by Liu et al. (2007) to improve the perceptual image quality during video conferencing. The example of off-the-shelf webcam images and the CTM enhanced images are shown in Figure 2.4, where the pale and unpleasant look of the left images are less appealing to most users compared to the right images. The basic idea is to select a set of tone mapping training images which look good perceptually, and build a Gaussian Mixture model for the color distribution in the face region. CTM is applied to any input image so that its color statistics in the face region matches the tone mapping training examples.

Consider n as the number of training images. Automatic face detection (Viola and Jones, 2004) is carried out to identify the face region for each training image I_i. For each color

(a) Original image

(b) Transformed image

(c) Original color image

(d) Transformed color image

Figure 2.3: Example of image after applying CLAHE (*Contrast-limited adaptive histogram equalization (CLAHE) - MATLAB adapthiseq*, 2014).

channel, the mean and standard deviation are computed for all pixels in the face region. Let $v_i = \left(m_1^i, m_2^i, m_3^i, \sigma_1^i, \sigma_2^i, \sigma_3^i\right)^T$ denote the vector that consists of the means, m^i and standard deviations, σ^i of the three color channels in the face region, denoted by the subscripts of 1, 2, and 3, respectively.

Then the distribution of the vectors $\{v_i\}_{1 \leq i \leq n}$ is modeled as a mixture of Gaussians, and m denote the number of mixture components. Let (μ_j, Σ_j) indicate the mean vector and covariance matrix of the j'th Gaussian mixture component, $j = 1, ..., m$, given any input image, the means and standard deviations of the three color channels in the face region is indicated by

(a) Original image

(b) Transformed image

(c) Original image

(d) Transformed image

Figure 2.4: Images on the left are input frames captured by two different webcams. Images on the right are the enhanced frames by CTM (Liu et al., 2007).

$v = \left(m_1, m_2, m_3, \sigma_1, \sigma_2, \sigma_3\right)^T$. The Mahalanobis distance $D_j(v)$ from v to j'th component is,

$$D_j(v) = \sqrt{(v-\mu_j)^T \sum_j^{-1} (v-\mu_j)}\,. \tag{2.15}$$

From the work of Liu et al. (2007), the target mean and standard deviation vector for v is defined as a weighted sum of the Gaussian mixture component centers μ_j, $j = 1, \ldots, m$, where the weights are inversely proportional to the Mahalanobis distances. More specifically, denoting

$\bar{v} = (\bar{m}_1, \bar{m}_2, \bar{m}_3, \bar{\sigma}_1, \bar{\sigma}_2, \bar{\sigma}_3)^T$ as the target mean and standard deviation vector, resulting

$$\bar{v} = \sum_{j=1}^{m} (w_j \times \mu_j), \qquad \text{where } w_j = \frac{1/D_j(v)}{\sum_{l=1}^{m} 1/D_l(v)}. \tag{2.16}$$

CTM is performed for each color channel to match the target distribution using the target mean and standard deviation vector. For color channel c, $c = 1,2,3$, consider $y = f_c(x)$ as the desired tone mapping function. To map the average intensity from m to $\bar{m}_c$, $f_c(x)$ needs to fulfill

$$f_c(m_c) = \bar{m}_c. \tag{2.17}$$

The derivative at m_c needs to be equal to $\frac{\bar{\sigma}_c}{\sigma_c}$ in order to modify the standard deviation σ_c to match the target $\bar{\sigma}_c$, such that,

$$f_c'(m_c) = \frac{\bar{\sigma}_c}{\sigma_c}. \tag{2.18}$$

In addition, $f_c(x)$ needs to be in the range of $[0,255]$, therefore,

$$\begin{matrix} f_c(0) = 0 \\ f_c(255) = 255 \end{matrix}. \tag{2.19}$$

A simple function that satisfies these constraints is

$$f_c(x) = \begin{cases} 0 & x < 0 \\ \frac{\bar{\sigma}_c}{\sigma_c}(x - m_c) + \bar{m}_c & 0 \le x \le 255 \\ 255 & x > 255 \end{cases}. \tag{2.20}$$

The drawback of this function is that it saturates quickly at the low and high intensities. A piecewise cubic spline that satisfies these constraint is fitted to overcome this problems (Liu et al., 2007). In order to prevent quick saturation at the two ends, constraints are added on the derivatives $f_c'(0)$ and $f_c'(255)$ as follows:

$$\begin{aligned} f_c'(0) &= 0.5 \times \frac{\bar{m}_c}{m_c} \\ f_c'(255) &= 0.5 \times \frac{255-\bar{m}_c}{255-m_c} \end{aligned} \quad . \tag{2.21}$$

Given the constraints of Equations (2.17), (2.18), (2.19), and (2.21), a piecewise cubic spline that satisfies these constraints can be fitted (Farin, 2002).

Note that the color tone mapping function is created based on the pixels in the face region, but it is applied to all the pixels in a given input image.

2.3 Face Recognition Classifiers

Face recognition has several advantages, i.e. natural, non-intrusive, and easy to use. The main advantage of the face technology is that it can be captured invisibly at a distance. Besides, facial features achieve the highest compatibility with a Machine Readable Travel Documents (MRTD) system based on several evaluation factors such as enrollment, renewal, machine requirements, and public perception. Due to these reasons along with its potential for continuously increasing law enforcement and commercial applications, it has been a research topic for decades (Al-Arashi, 2014).

However, face as a three-dimensional object is subjected to different variations such as, camera noise, illumination, pose and facial expression. These variations degrade face recognition system efficiency. Therefore, suitable classifier is needed to overcome the variations after

image pre-processing techniques.

2.3.1 Support Vector Machines

The SVM is a relatively new type of a classifier originally invented by Vapnik (1998). Significance of SVM is the transformation of data into a feature space, in which the classification can hopefully be achieved with a linear hypersurface. The new space is also of a higher dimension than the original (Cyganek, 2013).

2.3.1(a) Binary Classification

The goal for binary SVM classification problem is to separate the classes by a function which is induced from the available training examples by finding a decision surface that has the maximum distance to the closest points in the training examples which are termed the support vectors (Faruqe and Hasan, 2009; Guo et al., 2000; Vapnik, 1998). Figure 2.5 shows that there are many possible linear classifiers (hyperplanes) that can separate the data, but there is only one that maximizes the margin (the distance between the hyperplane and the support vectors). This linear classifier is known as the Optimal Separating Hyperplane (OSH).

Consider each point in the data set is referred as $\mathbf{x}_i \in \mathbb{R}^n$, $i = 1, 2, \ldots, l$ and belongs to class $y_i \in \{-1, +1\}$, presented in the form of $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_i, y_i)$, with a hyperplane $\langle \mathbf{w} \cdot \mathbf{x} \rangle + b = 0$. For linear classification, the two classes and OSH can be identified by

$$\mathbf{w} \cdot \mathbf{x}_i + b \geq 1, \qquad y_i = 1 \tag{2.22}$$

$$\mathbf{w} \cdot \mathbf{x}_i + b \leq -1, \qquad y_i = -1. \tag{2.23}$$

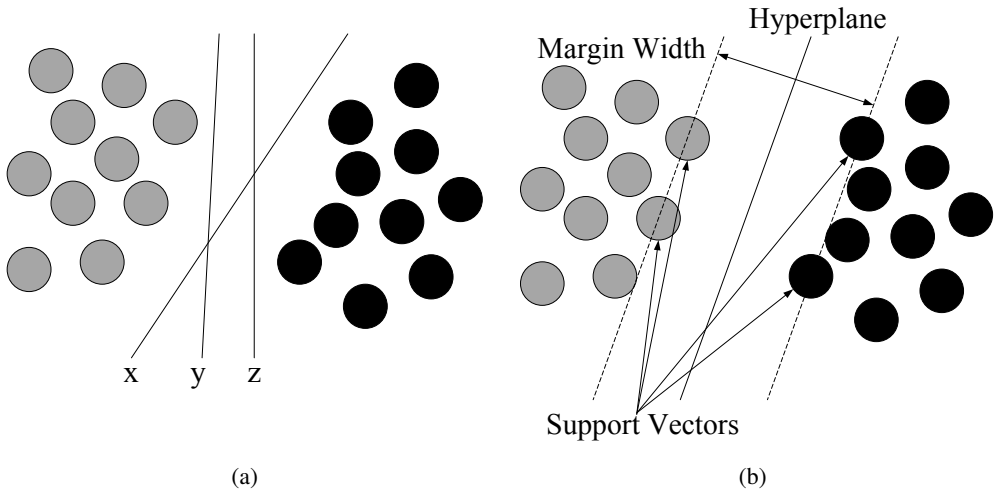


Figure 2.5: Binary classification: (a) arbitrary hyperplanes x, y, and z; (b) OSH with the largest margin passing the support vectors.

Generalizing (2.22) and (2.23) to

$$y_i \cdot \Big[(\mathbf{w} \cdot \mathbf{x}_i) + b\Big] \geq 1, \qquad i = 1, \ldots, l. \tag{2.24}$$

The distance between a point $\mathbf{x}$ and the hyperplane is,

$$d(\mathbf{w}, b; \mathbf{x}) = \frac{\mid \mathbf{w} \cdot \mathbf{x} + b \mid}{\parallel \mathbf{w} \parallel}. \tag{2.25}$$

The margin width is,

$$d = \frac{2}{\parallel \mathbf{w} \parallel}. \tag{2.26}$$

A better separation between two classes can be achieved with bigger d. Hence the hyperplane that optimally separates the data is the one that minimizes

$$\Phi(\mathbf{w}) = \frac{\parallel \mathbf{w} \parallel^2}{2}. \tag{2.27}$$