

國立交通大學

電控工程研究所

碩士論文

基於幾何人臉特徵之智慧型頭部姿態估測

Intelligent Head Attitude Estimation based on Geometric

Facial Features

研究生：王宣竣

指導教授：陳永平 教授

中華民國一百零二年六月

基於幾何人臉特徵之智慧型頭部姿態估測

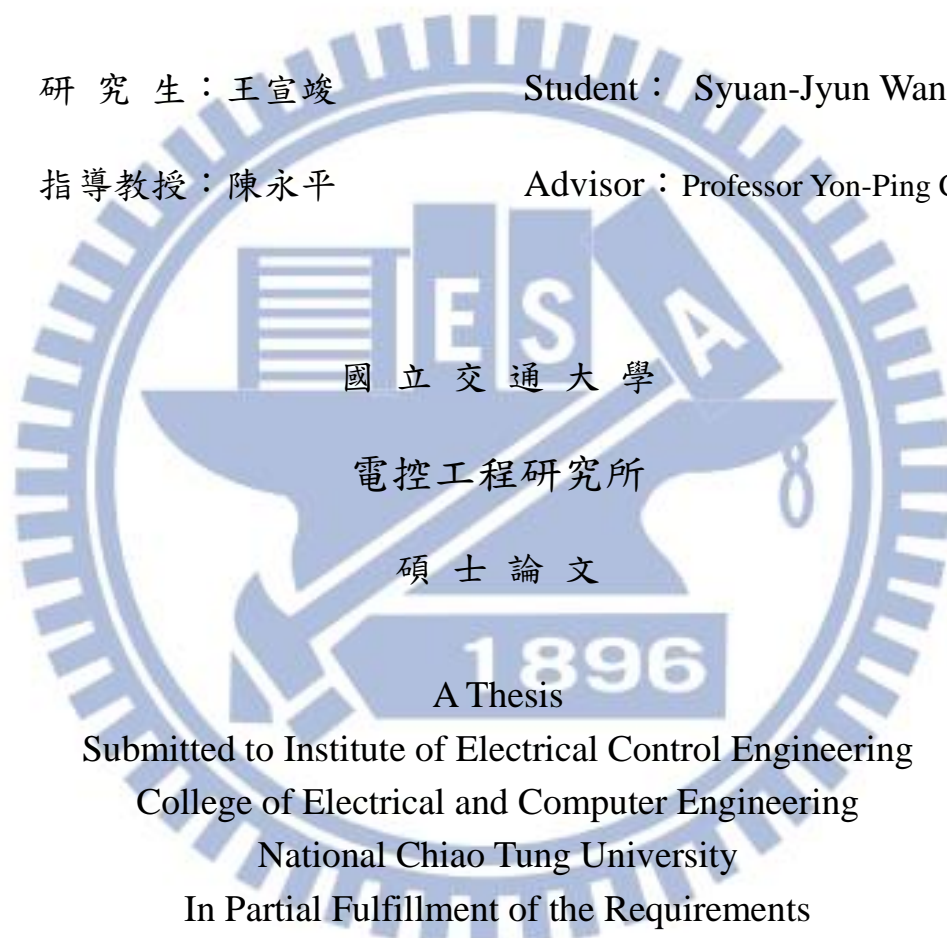
Intelligent Head Attitude Estimation based on Geometric
Facial Features

研究生：王宣竣

Student： Syuan-Jyun Wang

指導教授：陳永平

Advisor： Professor Yon-Ping Chen



A Thesis

Submitted to Institute of Electrical Control Engineering
College of Electrical and Computer Engineering
National Chiao Tung University

In Partial Fulfillment of the Requirements
For the degree of Master

In

Electrical Control Engineering

June 2013

Hsinchu, Taiwan, Republic of China

中華民國 一 百 零 二 年 六 月

基於幾何人臉特徵之智慧型頭部姿態估測

學生：王宣竣

指導教授：陳永平博士

國立交通大學電控工程研究所

摘要

近年來臉部特徵偵測和人臉辨識已被廣泛地研究，許多利用臉部特徵偵測的應用也隨之發展。本篇論文即利用幾何人臉特徵，針對頭部姿態的方向和角度，提出智慧型頭部姿態估測系統之設計。此系統能夠自動偵測影像中的人臉特徵，包括眼睛及嘴巴，進而判斷頭部姿態的方向和角度。本系統分成三個步驟完成智慧型頭部姿態估測系統設計，首先，利用膚色找出臉部位置後，使用幾何人臉特徵達到眼睛及嘴巴的高偵測率，第二，製作頭部姿態的人臉模擬立體模型，可調整臉部轉向角度 β 及頭部傾斜角度 γ ，在轉盤上標有七個偵測點，根據不同的 β 及 γ ，製作臉部模擬影像，並記錄每張影像的七個偵測點，作為類神經網路學習之用，第三，經由監督式學習的類神經網路設計達到頭部姿態的估測。本論文所提出的智慧型頭部姿態估測在於特定範圍內正確率可達 97.3%。

Intelligent Head Attitude Estimation based on Geometric Facial Features

Student: Syuan-Jyun Wang Advisor: Dr. Yon-Ping Chen

Institute of Electrical Control Engineering
National Chiao Tung University

ABSTRACT

Recently, facial feature detection and face recognition have been studied extensively and many applications using facial feature detection have been developed. This thesis is aimed at the development of head attitude estimation system (HAES) based on geometric facial features to detect the face orientation and angle. The HAES automatically detects eyes and mouth in the image as the facial features, and then determine the head attitude. There are three steps to complete the intelligent head attitude estimation. First, detect the human face based on skin color and use the geometric facial features to the detection of eyes and mouth in high accuracy rate. Second, build up a stereo facial model to simulate the head attitude which is able to adjust the face orientation and angle by seven detecting points marked on the face model. Record the seven detecting points on each image referring to a specific face orientation and angle, which will be used in neural network learning. Third, the HAES is completed by intelligent neural networks under supervised learning. The proposed HAES achieves a high accuracy rate up to 97.3%.

Acknowledgement

誠摯感謝指導教授 陳永平老師在這兩年中的悉心指導與教誨，老師嚴謹的治學態度，理論與實務並重的訓練，使得本論文得以順利完成。除了學術上的指導，在待人處事方面的啟發更讓我獲益良多，這份師恩會令我永生難忘。同時也感謝口試委員 林昇甫教授 張浚林教授對本論文所提出的珍貴意見與指證，讓本論文能更加的完整。

除此之外，感謝可變結構控制實驗室的世宏學長平日在攻讀博士學位之餘，不吝傳授知識與經驗及給予指導。此外，學長澤翰、文俊、文榜、榮哲、崇賢、孫齊與振方，同學兆村、仕政、咨璋與谷穎，學弟妹仁傑、惠琪、御旻與冠銘在課業與研究上一起學習與勉勵，以及實驗室上的協助，感謝妳們在生活中帶給我許多歡笑，使我兩年的研究所生活更加多采多姿。

最後，更要感謝我的父親、母親、大姊、小妹，你們的關心與鼓勵，給了我許多的溫暖，由於你們的支持，使我能專心在學習領域上衝刺，最後再次由衷的謝謝所有支持、關心與幫助過我的人。

謹以此篇論文獻給所有關心我、照顧我的人，你們的恩惠我銘感於心，由衷感謝你們。

王宣竣 2013 .6

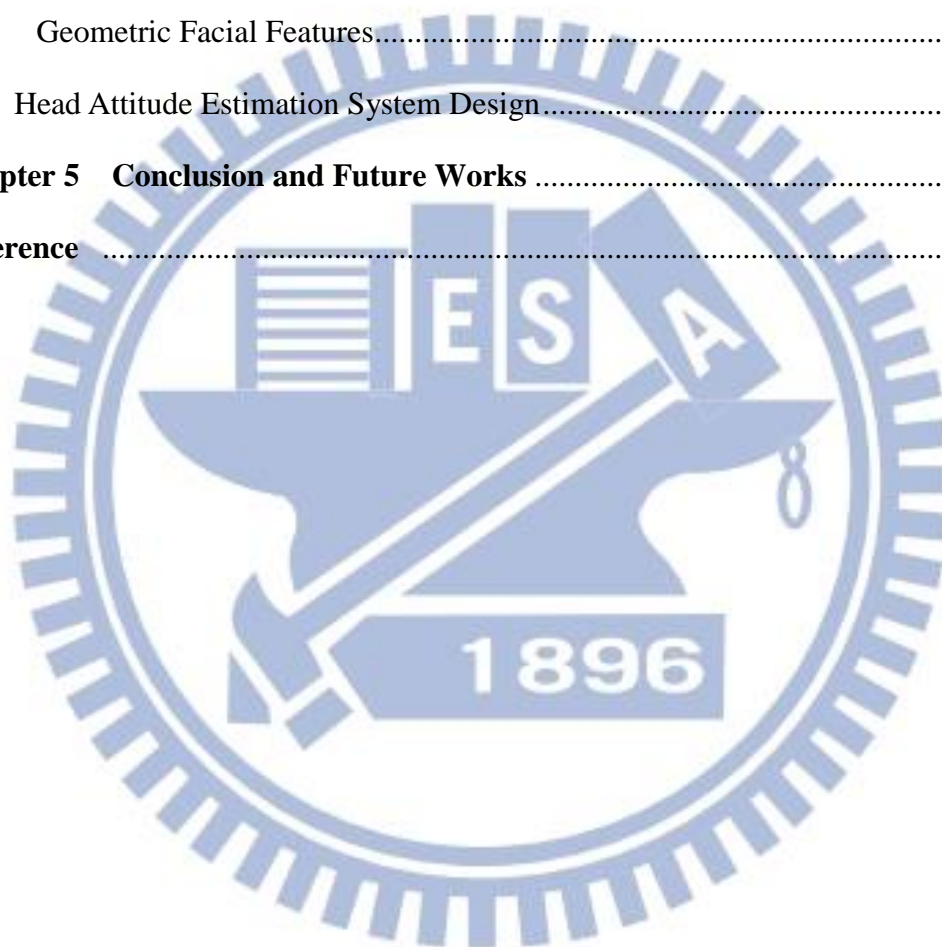
Contents

Chinese Abstract	i
English Abstract	ii
Contents	iii
List of Figures	vi
List of Tables	viii

Chapter 1 Introduction	1
1.1 Preliminary.....	1
1.2 System Overview	2
Chapter 2 Related Works	4
2.1 Introduction to ANNs.....	4
2.2 Back-Propagation Network.....	7
2.3 Skin Color Detection.....	10
2.4 Edge Detection.....	11
2.5 Morphology Operation.....	12

Chapter 3 Head Attitude Estimation System	14
3.1 Facial Features Detection.....	14
3.1.1 Human Face Detection.....	14
3.1.2 Morphology Operation.....	18
3.1.3 Connected Components Labeling.....	20
3.1.4 Face Classification System	21

3.1.5 Facial Feature Detection	25
3.2 Geometric Facial Features	30
3.3 Head Attitude Estimation System Design	31
Chapter 4 Experiment Results.....	38
4.1 Facial Features Detection.....	38
4.2 Geometric Facial Features.....	40
4.3 Head Attitude Estimation System Design.....	42
Chapter 5 Conclusion and Future Works	54
Reference	56



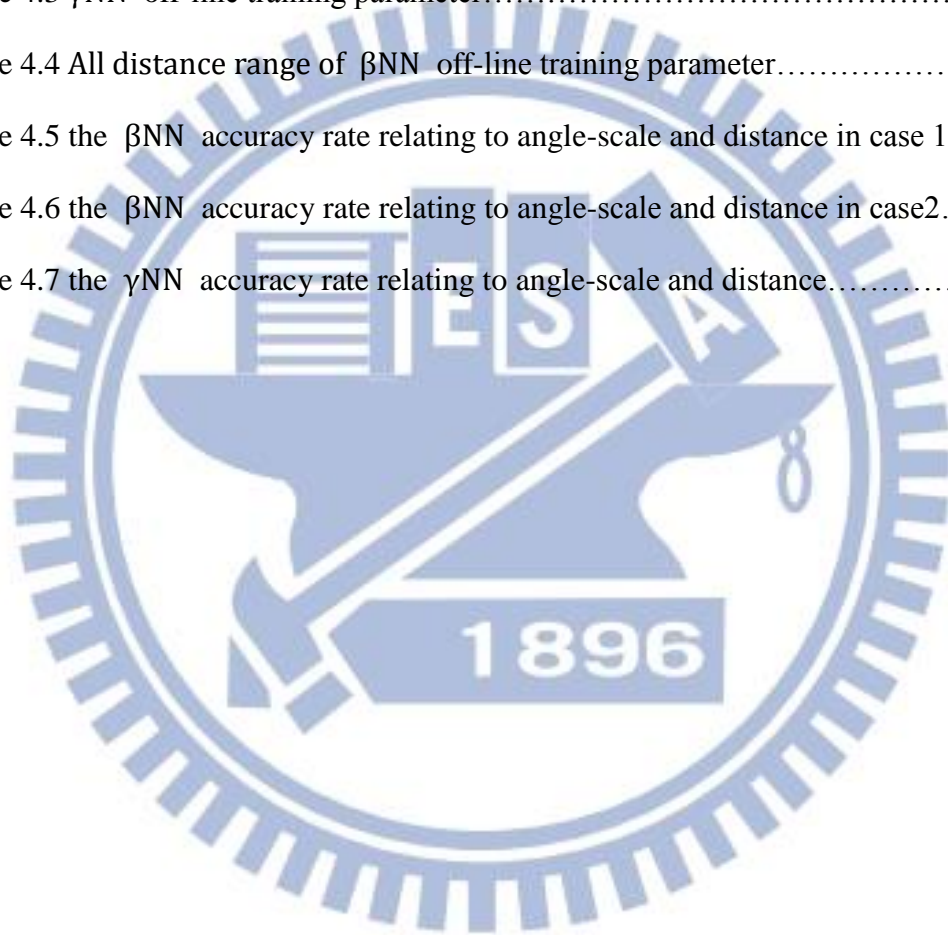
List of Figures

Figure 1.2.1 Architecture of head attitude estimation	3
Figure. 2.1 Basic structure of a neuron	4
Figure. 2.2 Multilayer feed-forward network	6
Figure. 2.3 Neural network with one hidden layer	8
Figure. 2.4 Example of dilation	12
Figure. 2.5 Example of erosion	13
Figure 3.1.1.1 statistic range of skin	16
Figure 3.1.1.2 (a) Color image, (b) thresholded image using Eqs.(3.1.1)-(3.1.6)	17
Figure. 3.1.2.1 (a) Structuring element B (b) Structuring element C	18
Figure. 3.1.2.2 Steps for morphology operations (a) Initial image (b) Result of erosion using structuring element B (c) Result of dilation using structuring element C	19
Figure. 3.1.3.1 Scanning the image	20
Figure. 3.1.3.1 Example of 4-pixel connected CCL	21
Figure 3.1.4.1 The proportion of the human face's width and height	22
Figure 3.1.4.2 The flow chart of the human face detection	23
Figure 3.1.4.3 To show real human faces detection of per step from the flow chart .	24
Figure 3.1.4.1 Extraction the ROI of human face get the edge detection (a) shows original image (b) get ROI of human face (c) edge detection of human face.	26
Figure 3.1.4.2 Head geometry	29
Figure 3.1.4.3 Head geometry divided to three part	29
Figure 3.2.1 Geometric facial features	30

Figure3.3.2 Train vectors on human face	31
Figure3.3.2 Train vectors on human face	33
Figure3.3.3 Neural network structure for β angle of the human head geometric	33
Figure 3.3.4 Neural network structure for γ angle of the human head geometric	34
Figure.3.3.5 linear transfer function	36
Figure.3.3.6 β NN	36
Figure.3.3.7 γ NN	37
Figure 4.1.1 The result of different people facial detection	38
Figure 4.1.1 The result of different people facial detection	39
Figure 4.2.1 The turntable appearance	40
Figure 4.2.2 RGB color label seven points	41
Figure 4.2.3 Detection RGB colors	41
Figure 4.3.1 The different distance between webcam and stereo facial model.	47
Figure 4.3.2 A histogram relates to accuracy rate and angle-scale of β NN in case1	48
Figure 4.3.3 In different angle-scale curves, β NN relates to accuracy rate and distance in case 1.....	49
Figure 4.3.4 A histogram relates to accuracy rate and angle-scale of β NN in case2	50
Figure 4.3.5 In different angle-scale curves, β NN relates to accuracy rate and distance in case2.....	51
Figure 4.3.6 A histogram relates to accuracy rate and angle-scales of γ NN	52
Figure 4.3.7 In different angle-scale curves, γ NN relates to accuracy rate and distance	53
Figure 4.3.8 the HAES final results	54

List of Tables

Table 3.1 Two angle of human head geometric.....	31
Table 4.1 Accuracy rate of three facial features detection.....	38
Table 4.2 30CM distance of β NN off-line training parameter.....	43
Table 4.3 γ NN off-line training parameter.....	44
Table 4.4 All distance range of β NN off-line training parameter.....	45
Table 4.5 the β NN accuracy rate relating to angle-scale and distance in case 1....	47
Table 4.6 the β NN accuracy rate relating to angle-scale and distance in case2.....	49
Table 4.7 the γ NN accuracy rate relating to angle-scale and distance.....	51



Chapter 1

Introduction

1.1 Preliminary

Facial feature detection and face recognition have been studied extensively in last decade and many applications used to detect facial features have been developed.

Several approaches have been proposed in the literature for facial feature detection in front-view head images. Yuille, et al. [1] use deformable templates to search for the facial features around the peaks and valleys of the intensity image. A similar approach is used by Hallinan [2] to detect the eyes in an image. Ahmed Fadzil and Abu Baker [3] adopt a multi-layered neural network to search the head area and locate the eyes. Chang and Huang [4] and Ohya, et.al.[5] employ skin color to locate facial features in colored images.

This thesis uses the methods proposed from Huang and Ohya, which are based on skin color to detect facial features such as eyes and mouth and can achieve high accuracy rate. According to the locations of eyes and mouth, which form geometric facial features, we can develop the head attitude estimation system (HAES) which determines the face orientation and angle as the head attitude.

Recently, in many researches the neural network has been carried out to deal with problems of system modeling concerning nonlinearities and uncertainties. It is well known that the neural network possesses excellent learning and mapping ability for nonlinear system modeling and the most prevalent neural network architecture is based on the back-propagation technique. Since the geometric facial features are nonlinear and complex, in order to determine the face orientation and angle, the proposed HAES in this thesis chooses the neural network with back-propagation algorithm to learn the face orientation and angle from the geometric facial features.

In the thesis, a stereo facial model is built up to simulate the head attitude which includes seven points to represent the on geometric facial features. Three points are the mouth and two eyes, and the other four points M1, M2, M3 and M4 construct a rectangle to mask the face. The stereo facial model includes a human face image and two angle-scaled turntables to indicate the facial angle γ and orientation β . To achieve the head attitude features, the left-eye is assigned as the original point. Then, choose the difference of the right-eye to the original point as the vector V1 and the difference of the mouth to the original point as the vector V2. Both V1 and V2 are adopted as the neural network inputs. Similarly, the differences of the M1, M2, M3 and M4 to the original point are set to be the vectors V3, V4, V5 and V6, respectively. Based on the neural network back-propagation algorithm with inputs V1 and outputs V2, V3, V4, V5, V6, β , and γ , the HAES is implemented. Most importantly, the HAES achieves a high accuracy rate up to 97.3% in the detection of face orientation and angle.

1.2 System Overview

In this thesis, a system is proposed for head attitude estimation (HAES). For software architecture, the image shown in Figure 1.2.1 is the flow chart of the proposed system.

There are three steps to complete the intelligent head attitude estimation, First, detect the human face based on skin color and use the geometric facial features to the detection of eyes and mouth in high accuracy rate. Second, build up a stereo facial model to simulate the head attitude which is able to adjust the face orientation and angle by seven detecting points marked on the face model. Record the seven detecting points on each image referring to a specific face orientation and angle, which will be used in neural network learning. Third, the HAES is completed by intelligent neural networks

under supervised learning. The remainder of this thesis is organized as follows. Chapter 2 describes the related works of the system. Chapter 3 describes intelligent head attitude estimation based on geometric facial features system. Chapter 4 shows the experiment results. Chapter 5 is the conclusions of the thesis and the future works.

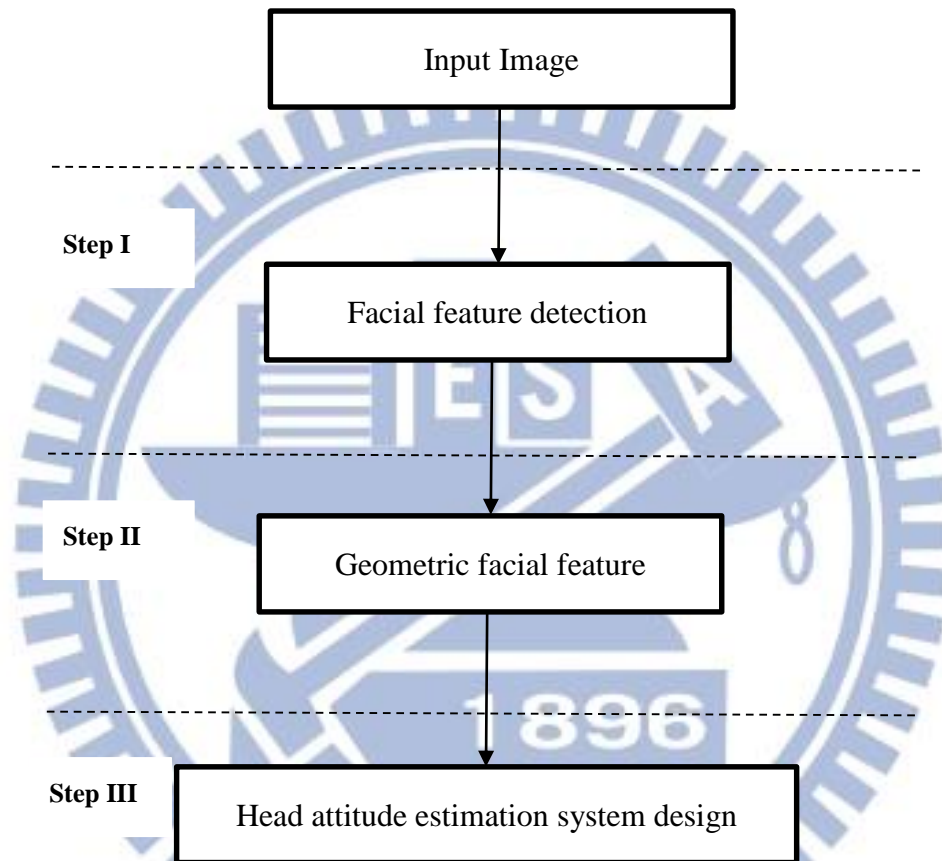


Figure 1.2.1 Architecture of head attitude estimation.

Chapter 2

Related Work

2.1 Introduction to ANNs

The human nervous system consists of a large amount of neurons, including somas, axons, dendrites and synapses. Each neuron is capable of receiving, processing, and passing signals from one to another. To mimic the characteristics of the human nervous system, recently investigators have developed an intelligent algorithm, called artificial neural networks (ANNs). In the artificial intelligence field, ANNs have been applied successfully to speech recognition, image analysis and adaptive control. This thesis will apply ANNs to the face detection in an eyeball system through learning.

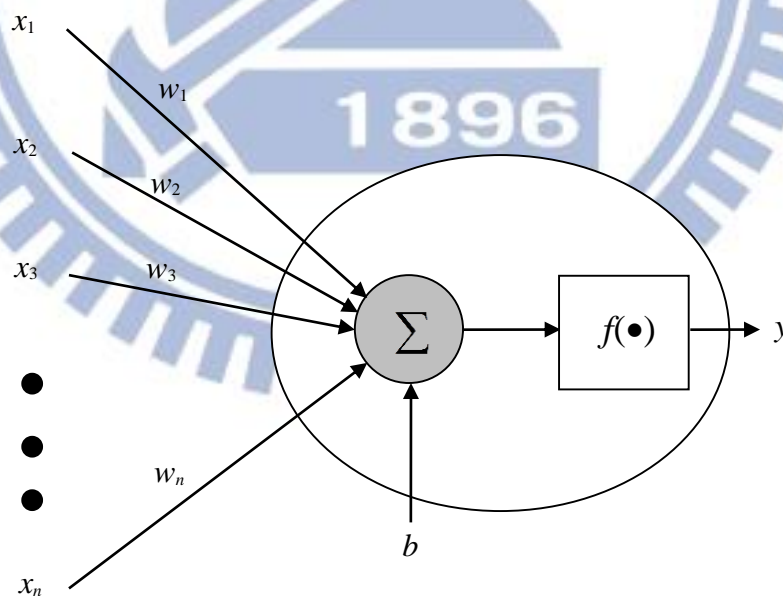


Figure. 2.1 Basic structure of a neuron.

Figure. 2.1 shows the basic structure of a neuron, whose input-output relationship is described as

$$y = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (2.1)$$

where w_i is the weight at the input x_i and b represents the bias. The activation function $f(\bullet)$ can be linear or nonlinear, such as linear function, log-sigmoid function and tan-sigmoid function, respectively expressed as below:

(1) linear function

$$f(x) = x \quad (2.2)$$

(2) log-sigmoid function

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.3)$$

(3) tan-sigmoid function

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.4)$$

Here, each input x_i is multiplied by a corresponding weight w_i , analogous to synaptic strengths. The weighted inputs are summed to determine the activation level of the neuron.

A general multilayer feed-forward network is composed of one input layer, one output layer, and some hidden layers. For example, Figure. 2.2 shows a neural network with one input layer, one output layer, and two hidden layers. Each layer is formed by neurons with basic structure depicted in Figure. 2.1. The input layer receives signals and response from the outside world, and then through the hidden layer to the output layer, the response of the net can be read. Note that in some cases only the input layer and output layer are required and the hidden layer can be omitted, i.e., the hidden layer is not necessary to be used.

Compared with networks using single hidden layer, networks with multi-hidden layer can solve more complicated problems. However, the training process of multi-hidden layer networks may be more difficult.

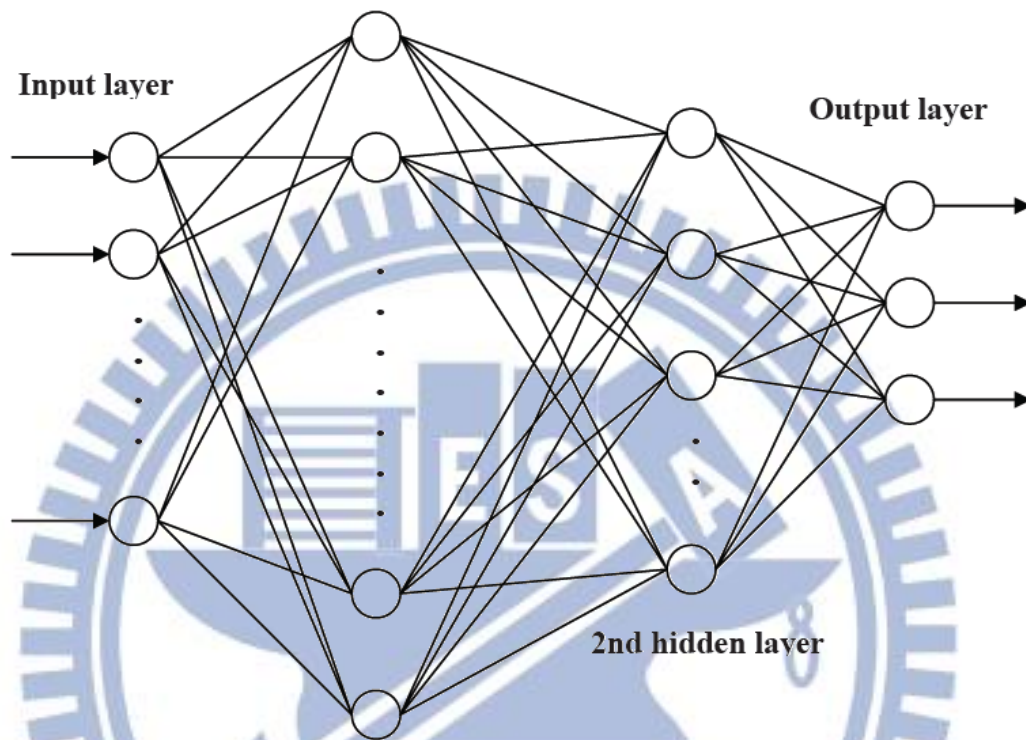


Figure. 2.2 Multilayer feed-forward network.

In addition to the architecture, the method of setting the weights is an important matter of different neural networks. For convenience, the training for a neural network is mainly classified into supervised learning and unsupervised learning. Training via supervised learning is mapping a given set of inputs to a specified set of target outputs. The weights are then adjusted according to a pre-assigned learning algorithm. For the unsupervised learning, it can self-organize a neural network without training data, i.e., only input vectors are provided, but no target vectors are specified. Through the unsupervised learning, the network modifies its weights so that the most similar input vectors can be assigned as the same group. In this thesis, the neural network is designed

for image feature extraction and recognition which requires two images, input image and target image, as the training input-output pairs. Hence, the neural network will be trained via supervised learning.

2.2 Back-Propagation Network

In supervise learning, the back propagation learning algorithm, is widely used in most application. The back propagation, BP in brief, algorithm was proposed in 1986 by Rumelhart, Hinton and Williams, which is based on the gradient steepest descent method for updating the weights to minimize the total square error of the output. To clearly explain the BP algorithm, an example is given in Fig. 2.3 which is a neural network with one hidden layer. Let the inputs be $x_i, i=1,2,\dots,I$, where I is the total number of input nodes and let the outputs be $y_j, j=1,2,\dots,J$, where J is the total number of output nodes. For the hidden layer, the k -th hidden node, $k=1,2,\dots,K$, with K being the total number of hidden nodes, receives information from input layer and sends out h_k to the output layer. These three layers are connected by two sets of weights, v_{ik} and w_{kj} . The weigh v_{ik} connects the i -th input node and the k -th hidden node, while the weigh w_{kj} connects the k -th hidden node and the j -th output node.

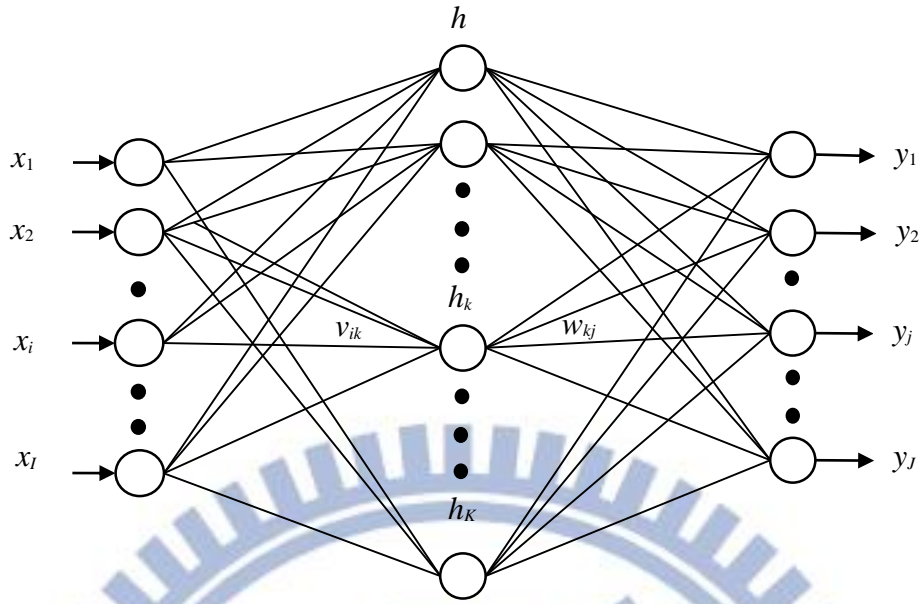


Figure. 2.3 Neural network with one hidden layer.

Based on the neural network in Figure. 2.3, the BP algorithm for supervised learning is generally processed by eight steps as below:

Step 1: Set the maximum tolerable error E_{\max} and then the learning rate η between 0.1 and 1.0 to reduce the computing time or increase the precision.

Step 2: Set the initial weight and bias value of the network randomly.

Step 3: Input the training data, $x = [x_1 \ x_2 \ \dots \ x_I]^T$ and the desired output

$$d = [d_1 \ d_2 \ \dots \ d_J]^T.$$

Step 4: Calculate each output of the K neurons in hidden layer

$$h_k = f_h \left(\sum_{i=1}^I v_{ik} x_i \right), \quad k = 1, 2, \dots, K \quad (2.5)$$

where $f_h(\bullet)$ is the activation function, and then each output of the J neurons in output layer

$$y_j = f_y \left(\sum_{k=1}^K w_{kj} h_k \right), \quad j = 1, 2, \dots, J \quad (2.6)$$

where $f_y(\bullet)$ is the activation function.

Step 5: Calculate the following error function

$$E(w) = \frac{1}{2} \sum_{j=1}^J (d_j - y_j)^2 = \frac{1}{2} \sum_{j=1}^J \left[d_j - f_y \left(\sum_{k=1}^K w_{kj} h_k \right) \right]^2 \quad (2.7)$$

Step 6: According to gradient descent method, determine the correction of weights as below:

$$\Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}} = -\eta \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial w_{kj}} = \eta \delta_{kj} h_k \quad (2.8)$$

$$\Delta v_{ik} = -\eta \frac{\partial E}{\partial v_{ik}} = -\eta \sum_{j=1}^J \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial h_k} \frac{\partial h_k}{\partial v_{ik}} = \eta \delta_{ikj} x_i \quad (2.9)$$

where

$$\delta_{kj} = (d_j - y_j) \left[f'_y \left(\sum_{k=1}^K w_{kj} h_k \right) \right]$$

$$\delta_{ikj} = \sum_{j=1}^J \left[(d_j - y_j) f'_y \left(\sum_{k=1}^K w_{kj} h_k \right) w_{kj} \right] f'_h \left(\sum_{i=1}^I v_{ik} x_i \right)$$

Step 7: Propagate the correction backward to update the weights as below:

$$\begin{cases} w(n+1) = w(n) + \Delta w \\ v(n+1) = v(n) + \Delta v \end{cases} \quad (2.10)$$

Step 8: Check the next training data. If it exists, then go to Step 3, otherwise, go to Step 9.

Step 9: Check whether the network converges or not. If $E < E_{\max}$, terminate the training process, otherwise, begin another learning circle by going to Step 1.

BP learning algorithm can be used to model various complicated nonlinear functions. Recently years The BP learning algorithm is successfully applied to many domain applications, such as: pattern recognition, adaptive control, clustering problem, etc. In the thesis, the BP algorithm was used to learn the input-output relationship for clustering problem.

2.3 Skin Color Detection

Color is an important source of information during the human visual perception activities. Skin color in a color image is relatively concentrated and stable. In recent years, skin color detection has become a popular research topic, and reached a great number of achievements. Nowadays, skin color detection has applied to a variety of tasks, for examples detecting and tracking human faces and gestures, filtering web image contents, and diagnosing disease [6,7,8,9].

As the first task in face detection technique, skin color detection can highly reduce the computational cost [10], and then extracts the potential face regions. To obtain the face locations in the image, these potential face regions are analyzed based on a face model including face shape and physical geometric information [11]. Furthermore, color image segmentation is computationally fast while being relatively robust to changes in scale, viewpoint, and complex background.

According to the characteristics of skin color in color space distribution, skin color pixels can be detected quickly by a skin color model. However the use of different color spaces for different races and different illuminations often results in different detection accuracy [12]. In this thesis, the experimental environment is our laboratory and the lighting condition is fixed.

Skin color characteristics are mainly described by skin color model. Usually, the skin color detection should be considered two aspects: color space selection and how to use the color distribution to establish a good skin color model. Nowadays main color spaces include RGB, HSV, HSI, YCrCb, some of their variant, etc, while RGB is the foundational method to represent color.

2.4 Edge Detection

Edge detection is a fundamental tool in image processing and computer vision, particularly suitable for feature detection and feature extraction, which aim at identifying points with brightness changing sharply or discontinuously in a digital image.

In the ideal case, the result of applying an edge detector to an image may lead to a set of connected curves that indicate the boundaries of objects and surface markings. Based on the boundaries that preserve the important structural properties of an image, the amount of data to be processed may be reduced since some irrelevant information is neglected. Following the edge detection, it seems that the task of abstracting information from the original image will be much simpler.

The common edge detection methods are based on differential operators, such as Laplacian [13], Roberts [14], Sobel [15], LOG [16], Prewitt [17], and Canny [18] operator, etc. In these classic methods, firstly masks are moved around the image. The pixels which are the dimension of masks are processed. Then, new pixels values on the new image provide us necessary information about the edge. These differential operators are all sensitive to abrupt change of pixel gray level so that they are sensitive to noises. Most of the existing edge detection techniques are effective in a certain cases

but often subject to a large amount of computation time and threshold setting. With neural networks, not only the existing approaches can be improved, but also develop new ones.

2.5 Morphology Operation

Morphology has two simple function dilation and erosion.

Dilation is defined as:

$$A \oplus B = \{x : (\hat{B})_x \cap A \neq \phi\} \quad (2.11)$$

where A and B are sets in Z . This equation simply means that B is moved over A and the intersection of B reflected and translated with A is found. Usually A will be the signal or image being operated on and B will be the structuring element. Figure. 2.4 Shows how dilation works.

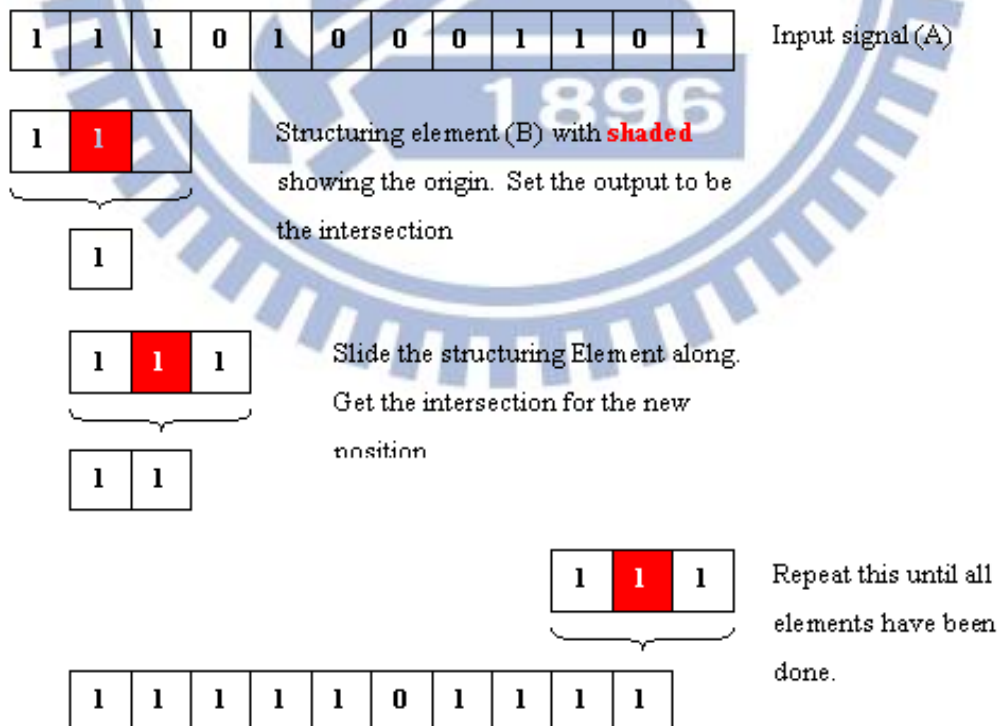


Figure. 2.4 Example of dilation.

The opposite of dilation is known as erosion. This is defined as:

$$A \ominus B = \{x : (B)_x \subseteq A\} \quad (2.12)$$

which simply says erosion of A by B is the set of points x such that B , translated by x , is contained in A . Figure. 2.5 shows how erosion works. This works in exactly the same way as dilation. However equation (2.12) essentially says that for the output to be a one, all of the inputs must be the same as the structuring element. Thus, erosion will remove runs of ones that are shorter than the structuring element. This thesis will applied two kind of this operation to process the image.

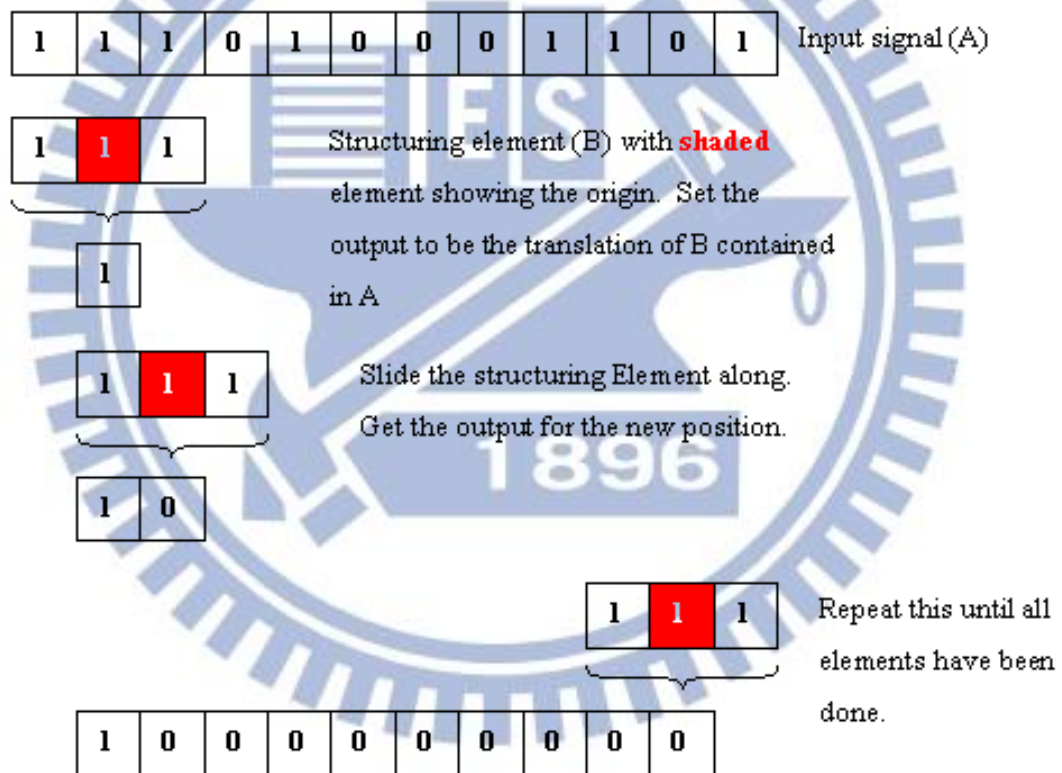


Figure. 2.5 Example of erosion.

Chapter 3 Head Attitude Estimation System

The image process of detecting human facial features such as the eyes, nose and mouth is crucial to applications like automatic face recognition [19] and head attitude estimation [20]. This thesis further achieves geometric facial features based on the detected human facial features and proposes a head attitude detection system using artificial neural network (ANN) to detect the orientation and angle of a human head.

3.1 Facial Features Detection

Automatic human face analysis and recognition has received significant attention during the past decades, due to the emergence of many potential applications such as person identification, video surveillance and human computer interface. An automatic face recognition usually begins with the detection of face pattern, and then proceeds to normalize the face images using information about the location and appearance of facial feature such as eyes and mouth [21] ,[22]. Therefore, detecting faces and facial features are a crucial step. 3.1.1-3.1.3 will introduce how to detect the human face in a successive frames. After the human face is confirmed, 3.1.4 will introduce a method to detection human eyes and mouth.

3.1.1 Human Face Detection

Human face detection and recognition have long been a popular research topic. In the last decades, researchers have devoted much effort to these two problems and have obtained some satisfactory results. Some of these previous efforts were focused on face recognition. However, an accurate and efficient method for human face detection is still lacking.

Popular algorithms for face detection include template matching, geometry features and skin color detection. Skin color detection has been gaining popularity and important in pattern recognition. Generally, it is the first step of computer vision tasks, such as

detection, tracking, and recognition of face. Many researches have indicated that skin color can be captured easily under suitable color space. Because of the human's skin color can be limited in a range of some specific color spaces even if the human races are different. Hence, several color spaces have been used for displaying the skin color distribution introducing normalized RGB, HSV, YCbCr, CIE-Lab color space, etc.

In the many methods, this thesis uses the normalized RGB method, which is effective used for skin color segmentation. Because this method consider the white balance effective and the illumination variable, which are both reasons to perform whether skin color detection well or not. In the practice, the normalized RGB method can show the good performance of detection skin color, which is the most importance reason to use this method of the thesis.

Through the computerized statistics difference illumination condition and human skin color range know the RGB space sensitive to external environment, therefore, thesis convert RGB space to NCC (Normalized Color Coordinate), formats are showing as:

$$r = \frac{R}{R+G+B} \quad (3.1)$$

$$g = \frac{G}{R+G+B} \quad (3.2)$$

The format (3.1) and (3.2) are normalized red color and green color, respectively, which target is reduce original color dependence the brightness. Figure 3.1.1.1 shows the skin locus which the X coordinate represents r and the Y coordinate represents g , therefore, we can observe the figure 3.1.1.1 which the skin range are very centralize. The values range from 0.2 to 0.6 of the X coordinate, on the other hand, the values range from 0.2 to 0.4 of the Y coordinate, furthermore, the statistic result can define boundary function, which are defined as following:

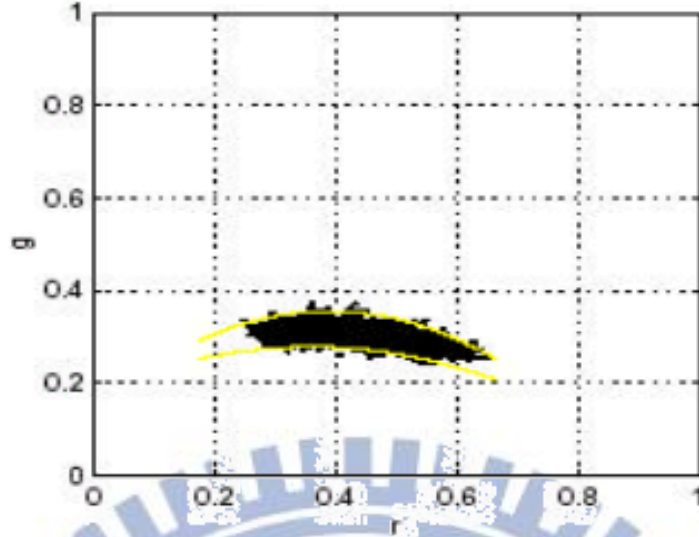


Figure 3.1.1.1 statistic range of skin.

First, a simple membership function to the skin locus is a pair of quadratic functions defining the upper and lower bound of the cluster. For each r , the maximum and minimum g was used to estimate the upper and lower quadratic function. Using least square estimation, the upper bound quadratic coefficients are found to be $A_u = -1.3767$, $b_u = 1.0743$, $c_u = 0.1452$; the lower bound coefficients are $A_d = -0.776$, $b_d = 0.5601$, $c_d = 0.1766$. Therefore, the Q_+ and Q_- are define as (3.3) and (3.4):

$$Q_+ = A_u r^2 + b_u r + c_u \quad (3.3)$$

$$Q_- = A_d r^2 + b_d r + c_d \quad (3.4)$$

Because the white points are included the Q_+ and Q_- so we have to eliminate the white points, therefore the quadratic are showing in (3.5):

$$W = (r - 0.33)^2 + (g - 0.33)^2 \quad (3.5)$$

Pixels with chromaticity (r, g) are then given skin locus membership values $S(r, g)$

where

$$S = \begin{cases} 1 & \text{if } (g < Q_+) \& (g > Q_-) \& (W > 0.0004) \\ 0 & \text{otherwise} \end{cases} \quad (3.6)$$

If the S be assign to 1 then represent to skin regin, otherwise, 0 to represent to non-skin region. Figure 3.1.1.2 shows the thresholded image obtained by above equations. Using Eqs.(3.1)-(3.6), the binary image is obtained as the shown in Fig.5.1.1.2(b), where the black color represents the non-skin region, in this case white objects, while white color represents skin region.

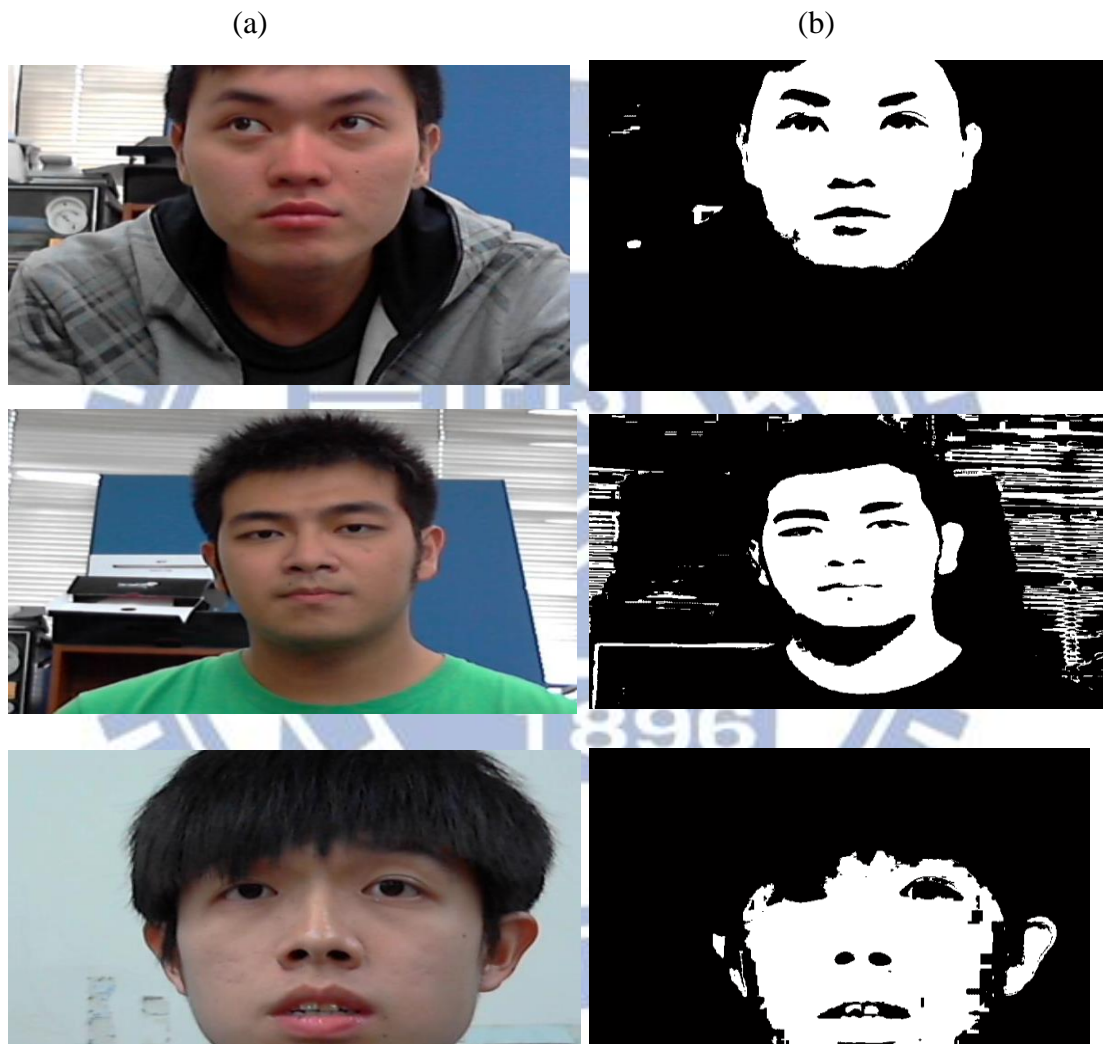


Figure 3.1.1.2 (a) Color image, (b) thresholded image using Eqs.(3.1)-(3.6)

3.1.2 Morphology Operation

After applying color extraction, color regions are extracted from the original image, but some noise still exists therein. One of the conventional ways to eliminate noise regions is using the morphology operations. In the thesis, the noises are eliminated by the morphology erosion operation expressed as

$$A \ominus B = \{x : (B)_x \subseteq A\} \quad (3.7)$$

where B is a disk-shaped structuring element with radius 4 as shown in Fig. 3.5.2.1 and the noises in image A with region smaller than B are erased after operation. However, some gaps may be also generated in isolated regions after erosion. In order to repair these gaps, further employ the morphology dilation operation expressed as

$$A \oplus C = \{x : (\hat{C})_x \cap A \neq \phi\} \quad (3.8)$$

where C is a disk-shaped structuring element with radius 10 as shown in Figure. 3.1.2.1 (b) and the gaps in image A are repaired after operation. Fig. 3.1.2.2 shows an example of erosion and dilation using the structuring elements B and C .

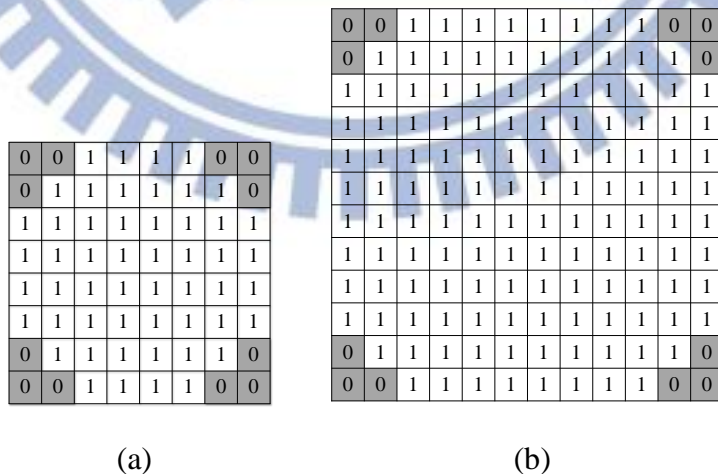


Figure. 3.1.2.1 (a) Structuring element B (b) Structuring element C



Figure. 3.1.2.2 Steps for morphology operations (a) Initial image (b) Result of erosion using structuring element B (c) Result of dilation using structuring element C

3.1.3 Connected Components Labeling

After morphology operation different components are identified by using Connected Components Labeling (CCL), which is often used in computer vision to detect connected regions containing 4 or 8 pixels in the binary digital image [24]. In this thesis, the 4-pixel connected component will be used to label potential face regions.

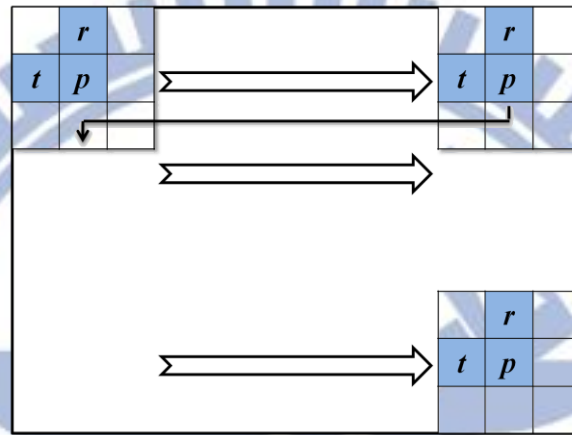


Figure. 3.1.3.1 Scanning the image.

The 4-pixel connected CCL algorithm can be partitioned into two processes, labeling and componentizing. During the labeling, the image is scanned pixel by pixel, from left to right and top to bottom as shown in Figure. 3.1.3.1, where p is the pixel being processed, and r and t are respectively the upper and left pixels to p . Defined $v(\cdot)$ and $l(\cdot)$ as the binary value and the label of a pixel. If $v(p)=0$, then move on to next pixel, otherwise, i.e., $v(p)=1$, the label $l(p)$ is determined by following rules:

- R1. For $v(r)=0$ and $v(t)=0$, assign a new label to $l(p)$.
- R2. For $v(r)=1$ and $v(t)=0$, assign $l(r)$ to $l(p)$, i.e., $l(p)=l(r)$.
- R3. For $v(r)=0$ and $v(t)=1$, assign $l(t)$ to $l(p)$, i.e., $l(p)=l(t)$.
- R4. For $v(r)=1$, $v(t)=1$ and $l(t)=l(r)$, then assign $l(r)$ to $l(p)$, i.e., $l(p)=l(r)$.
- R5. For $v(r)=1$, $v(t)=1$ and $l(t) \neq l(r)$, then assign $l(r)$ to both $l(p)$ and $l(t)$,

i.e., $l(p)=l(r)$ and $l(t)=l(r)$.

For example, after the labeling process, Figure. 3.1.3.1(a) is changed into Figure. 3.1.3.1(b). It is clear that some connected components contain pixels with different labels. Hence, it is required to further execute the process of componentizing, which sorts all the pixels connected in one component and assign them by the same label, the smallest number among the labels in that component. Figure. 3.1.3.1(c) is the result of Figure. 3.1.3.1(b) after componentizing.

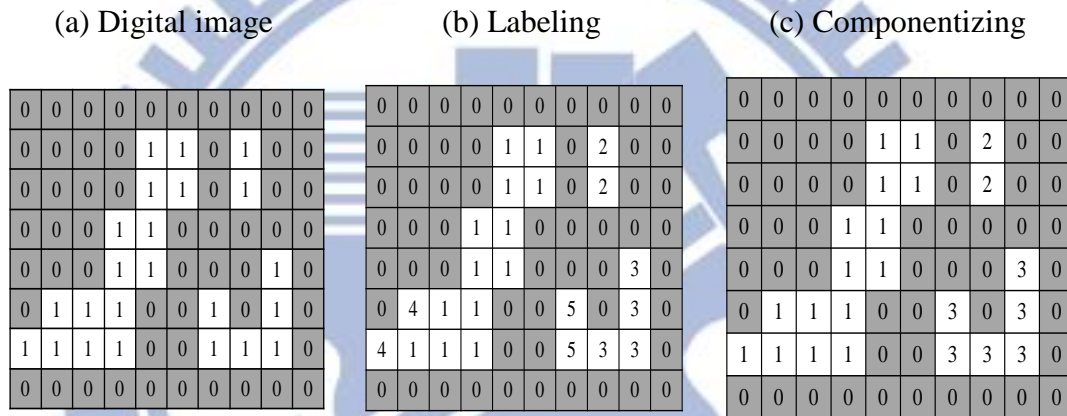


Figure. 3.1.3.1 Example of 4-pixel connected CCL.

3.1.4 Face Classification System

After the skin color extraction and connected components labeling (CCL), which get the face candidates, therefore, this thesis define two conditions to confirm the face location of per frame.

- (I) Areas are judged which one is face region. After the CCL, the potential face regions are located. The face area have to fit (3.9), too small areas will difficult search eyes and mouth, on the other hand, if the area is too huge not real face

size so both condition have to eliminate.

$$1000 \text{ Pixels} \leq \text{AREA} \leq 8000 \text{ Pixels} \quad (3.9)$$

- (II) The proportions of the human face's width and height have to fit the (3.10). In the general condition, human face's height bigger than the width, the figure 3.1.4.1 show the proportion of the human face's width and height, besides, the human face candidates include some non-face skin regions(the neck and arms), hence, we have define the proportion of the human face's width and height to confirm the human face region.

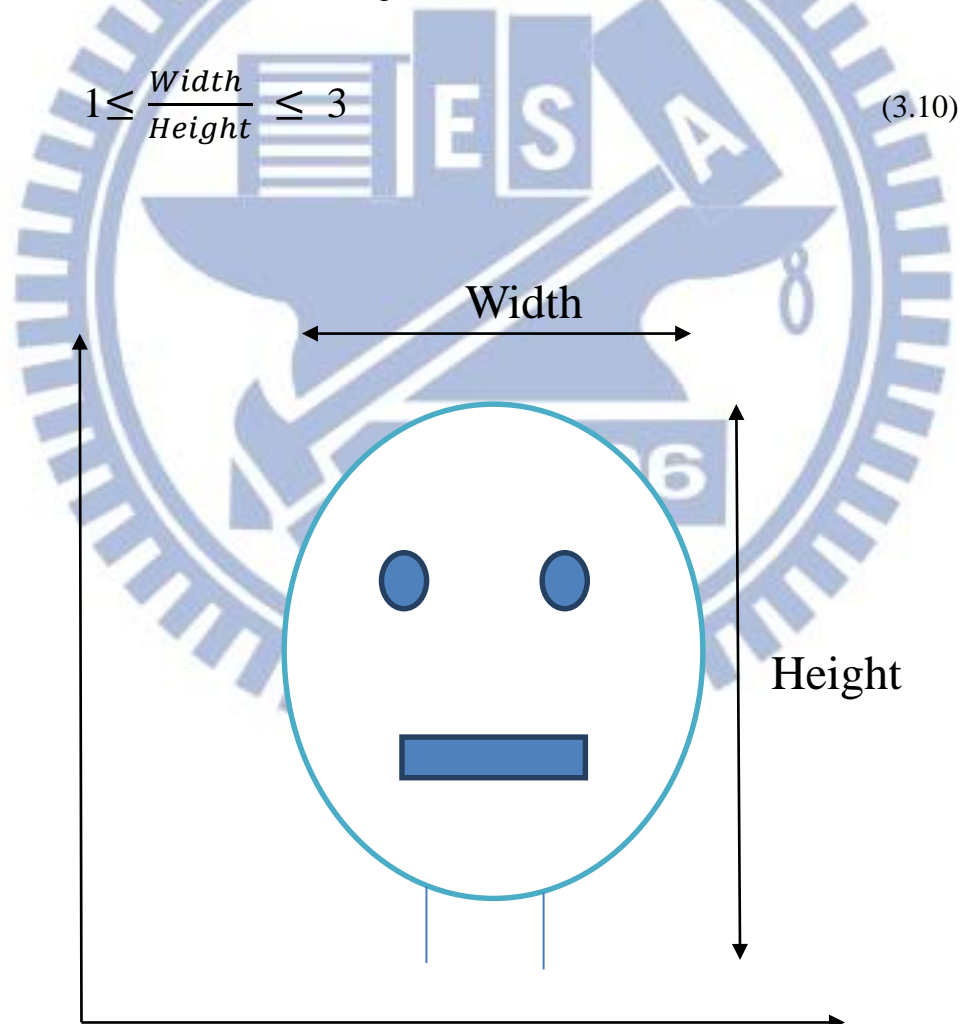


Figure 3.1.4.1 The proportion of the human face's width and height.

Before the detect eyes and mouth we have to human face detection, hence, I will introduce the human face detection conclusion steps, which are as following figure 3.1.4.2 and the figure 3.1.4.3 shows Figure 3.1.4.2 flow chart per steps real human face plots.

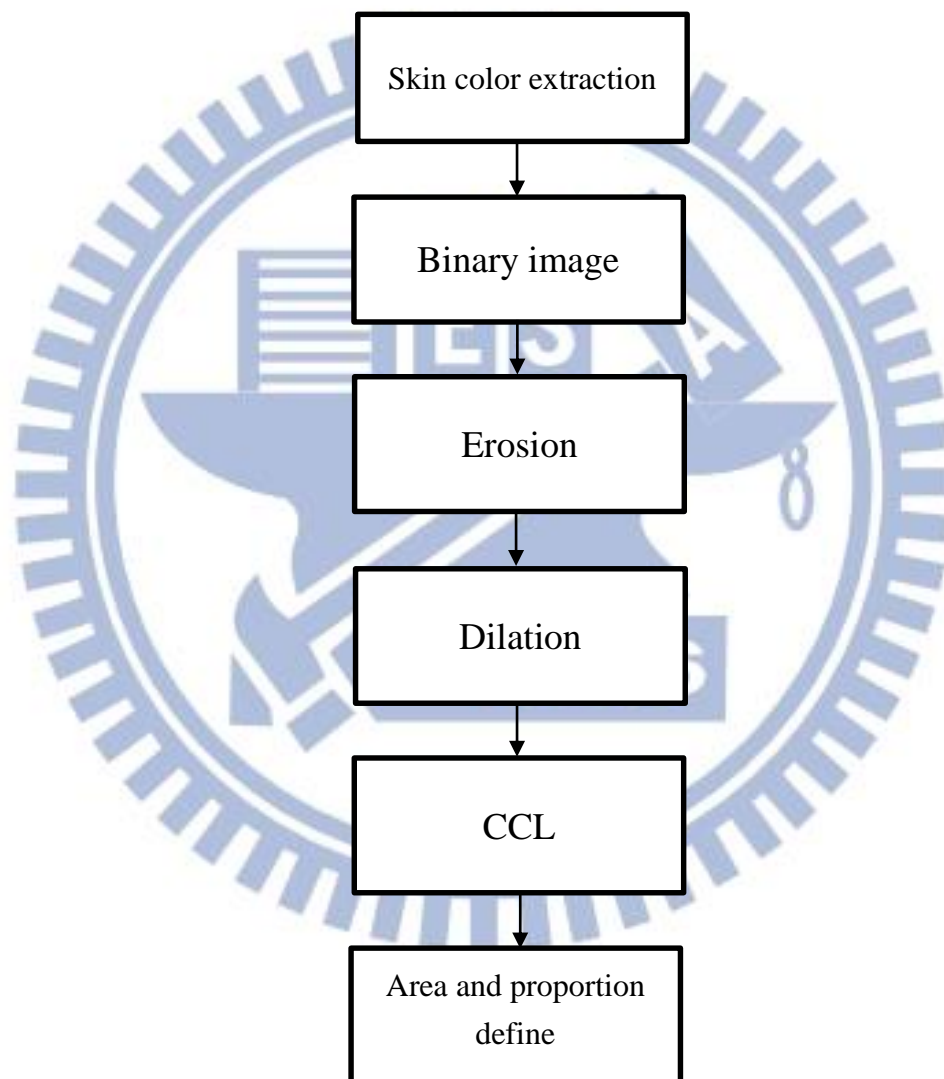


Figure 3.1.4.2 The flow chart of the human face detection.

Skin color extraction and Binary image

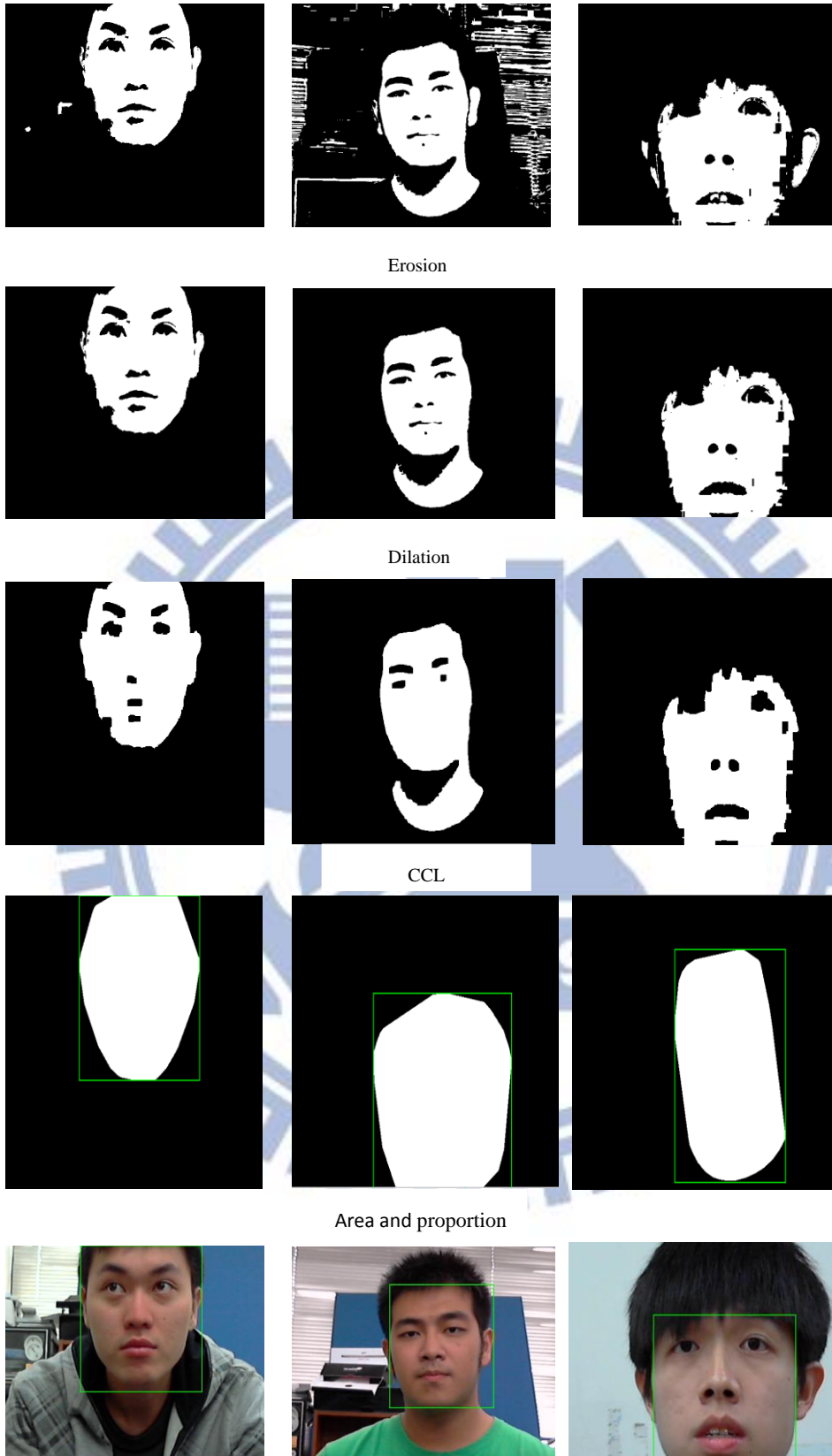


Figure 3.1.4.3 To show real human faces detection of per step from the flow chart.

3.1.4 Facial Feature Detection

In this thesis goal is to know human face faces where then get two information which are orientation and angle, therefore, this thesis has to face detection and facial feature detection. In the last sections already get human face position, which are known the human face position where in per successive images. In this section wants to use last section result then reach goal which are detection human facial feature, just like eyes, mouth, nose and so on, but this thesis just to search two features on the human face which are mouth and eyes. In this section will show more details how to detect human eyes and mouth in this thesis.

3.1.4.1 Eye and mouth detection

The eye is the most significant and important feature in the human face, as extraction of the eye are often easier as compared to other facial features. Eye detection is also used in person identification by iris matching. Only those image region that contain possible eye pairs will be fed into a subsequent face verification system. Localization of eyes is also a necessary step for many face classification methods. Eyes can be used for crucial face expression analysis for human computer interactions as they often reflects a person's emotions.

After last section of face detection, which are get human face ROI. In this step in eye detection involves edge detection. Morphological techniques are used for boundary detection. Dilation followed by erosion and the calculation of differences between the two produces an image with boundaries. For the purposed at hand, this technique is found to be more efficient that the laplacian edge detection. This is followed by suitable threshold of the image. The figure 3.1.4.1 shows original image to get ROI of human

face and edge detection of human face.

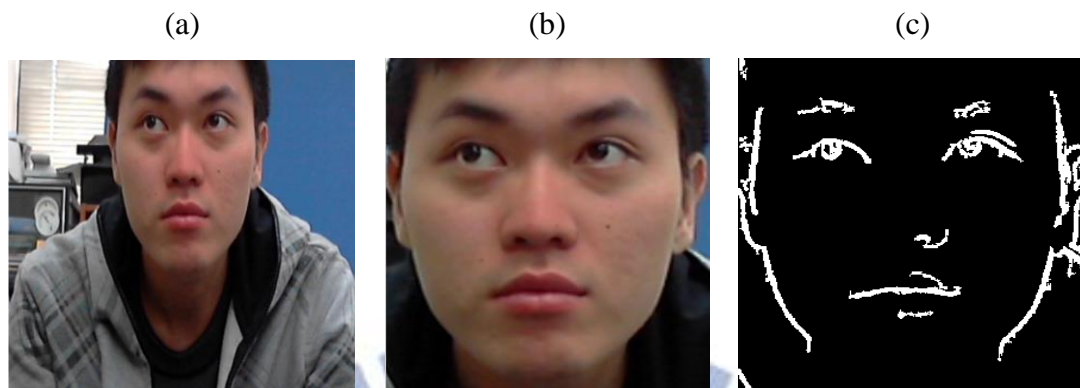


Figure 3.1.4.1 extraction the ROI of human face get the edge detection (a) shows original image (b) get ROI of human face (c) edge detection of human face.

The figure 3.1.4.2 is well known from the art of drawing the human head. The average human head is approximately five eye-lengths wide [25]. Both eyes lie on the line midway between the top of the head and the bottom of the chin, and the distance between them is approximately equal to one eye-length. The distance between the central line of the mouth and the center of the face is approximately one third of the head length. Therefore, after the edge detection this thesis accords to the figure 3.1.4.2 where the head geometry is divided into three parts (see Figure 3.1.4.3) which are left-part, right-side part and down-part showing the more detail (i), (ii) and (iii), respectively.

- (i) The red rectangle in Figure 3.1.4.3 is left-part which is included right-eye. In the right-part we try to find the right-eye, therefore, in this part I label some points and set some threshold to detect the right-eye position. P2 is human face centroid (RFC), P1 is the half human face (HF) of the coordinate y , P11 is half human face of the coordinate x , P9 is half face of the coordinate x multiplied by 0.2, P10 is half face of the coordinate x multiplied by 0.8, P4 is right eyebrow centroid(REC),P6 is right-eye centroid (REC). After define all points at right-

side part then set some threshold to search right-eye, the red rectangle is connected to P9, P10, P1 and P2. there are two binary objects which are right-eyebrow and right-eye in the red rectangle region, therefore, I have to determine which one is the right-eye then mark it. Because the right-eyebrow centroid of coordinate y is less than the right-eye centroid of the coordinate y , therefore, I let P6 of the coordinate y bigger than P4 of the coordinate y . finally, I successful detect right-eye.

(ii) The yellow rectangle in Figure 3.1.4.3 is left-part which is included left-eye. In the left-part we try to find the left-eye, therefore, in this part I label some points and set some threshold to detect the left-eye position. P2 is human face centroid (HFC), P3 is the half human face (HF) of the coordinate y , P11 is half human face of the coordinate x , P12 is half face of the coordinate x multiplied by 0.2 and add half human face of the coordinate x , P13 is half face of the coordinate x multiplied by 0.8 and add the half human face of the coordinate x , P5 is left eyebrow centroid (LEC), P7 is left-eye centroid (LEC). After define all points at left-side part then set some threshold to search left-eye, the yellow rectangle is connected to P3, P7, P12 and P13. there are two binary objects which are left-eyebrow and left-eye in the yellow rectangle region, therefore, I have to determine which one is left-eye then mark it. Because the left-eyebrow centroid of coordinate y is less than the left-eye centroid of the coordinate y , therefore, I let P7 of the coordinate y bigger than P5 of the coordinate y . finally, I successful detect left-eye.

(iii) The blue rectangle in Figure 3.1.4.3 is down-part which is included mouth. In the down-part we try to find the mouth, therefore, in this part I label some points and use the facial relation between eyes and mouth to check mouth position. P2 is human face centroid (HFC), P1 is the half human face (HF) of the coordinate

y, P3 is half human face of the coordinate y, P14 is half face of the coordinate x multiplied by 0.2 and add half human face of the coordinate x, P15 is half face of the coordinate x multiplied by 0.8 and add the half human face of the coordinate x, P8 is mouth centroid (MOUC). After define all points at down part then set some check conditions to search mouth, the yellow rectangle is connected to P14, P15, P16 and P17. In the down-part have four relation between eyes and mouth to check whether mouth is successful found position or not. There are some mathematics function and mouth check which are defined as below: The distance between right-eye and left-eye is assume Math1, the distance between left-eye and mouth is assume Math 2, the distance between left-eye and mouth is assume Math3, the distance between middle of eyes and mouth is assume Math4.

$$\text{Math 1: } \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3.11)$$

$$\text{Math 2: } \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2} \quad (3.12)$$

$$\text{Math 3: } \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2} \quad (3.13)$$

$$\text{Math 4: } \sqrt{(x_4 - x_3)^2 + (y_4 - y_3)^2} \quad (3.14)$$

Check1: Check whether the mouth is in down-part or not.

$$\text{Check2: } 0.9 \leq \frac{\text{math1}}{\text{math4}} \leq 1.5 \quad (3.15)$$

$$\text{Check 3: Check whether the } |\text{math2}-\text{math3}| < 0.25*\text{math1} \text{ or not.} \quad (3.16)$$

$$\text{Check4: Check whether the } |\text{math1}-\text{math4}| < 0.25*\text{math1} \text{ or not.} \quad (3.17)$$

After confirm Check1 to Check4 then the human facial of mouth is detected

position.

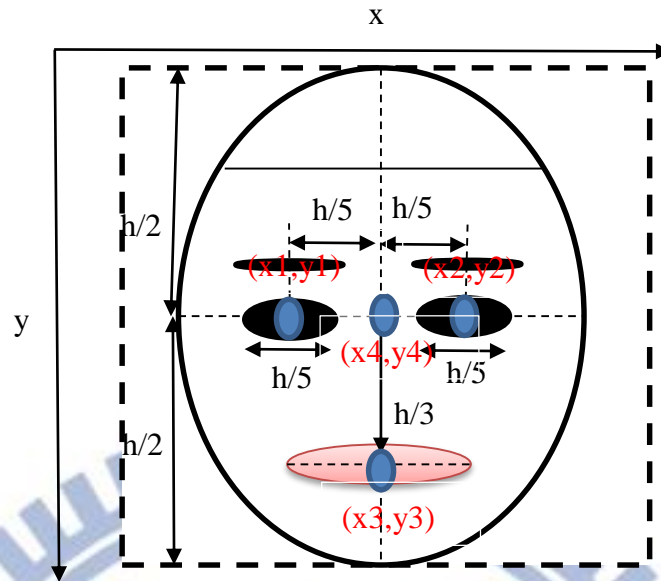


Figure 3.1.4.2 head geometry.

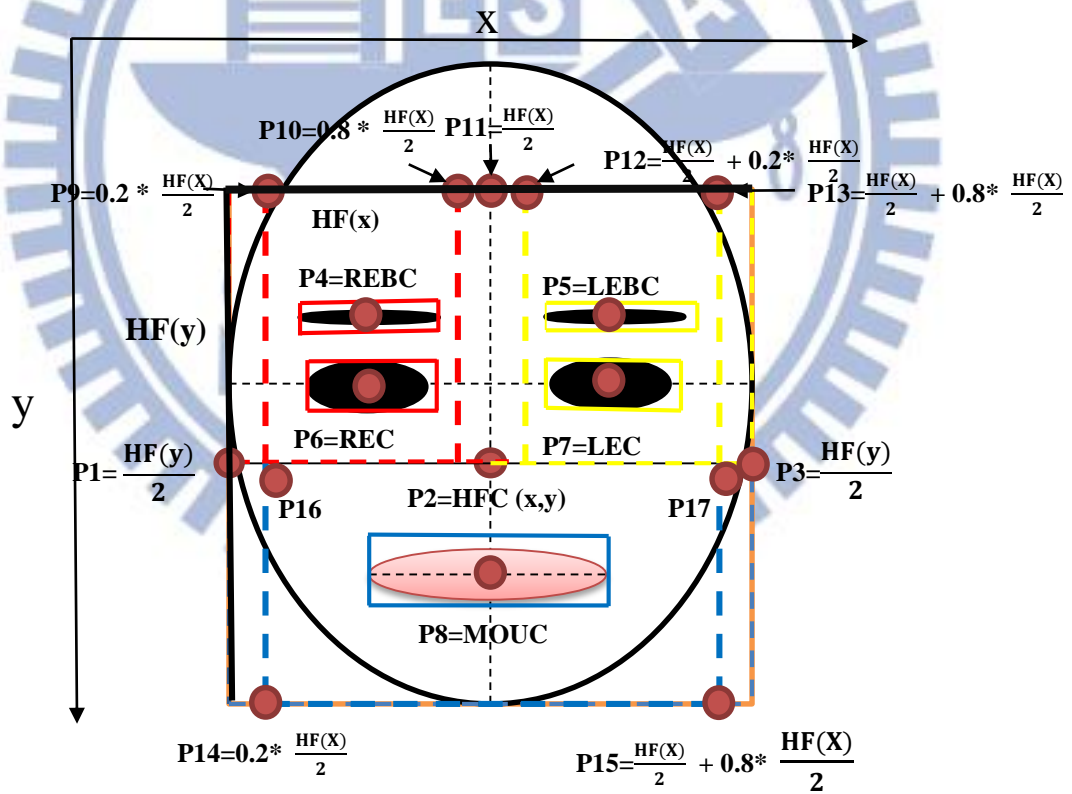


Figure 3.1.4.3 Head geometry divided to three part.

3.2 Geometric Facial Features

This thesis methods uses human's eyes and mouth as feature, human's eyes and mouth have fixed proportion of face. Using this characteristic we can know the human's face orientation and rotation and angles. First, let the right-eye be the original point and by the statistic average human's right-eye to left-eye and right-eye to mouth, we can get vectors $V1$ and $V2$, respectively. After averaging a statistic of 100 real human faces facial features of right-eye to left-eye and right-eye to mouth the results are, $V1$ divided by $V2$ equals 0.81 ± 0.1 which shows the figure 3.2.1.

The head of the a human being can be thought of as an ellipsoid that sits on the top of the torso and is attached to it at the center by the neck. The head movement can then be viewed as rotations β , γ and α around the X, Y and Z -axes, respectively. This thesis focuses on β and γ of the human head rotation angles and orientation which shows table 3.1.

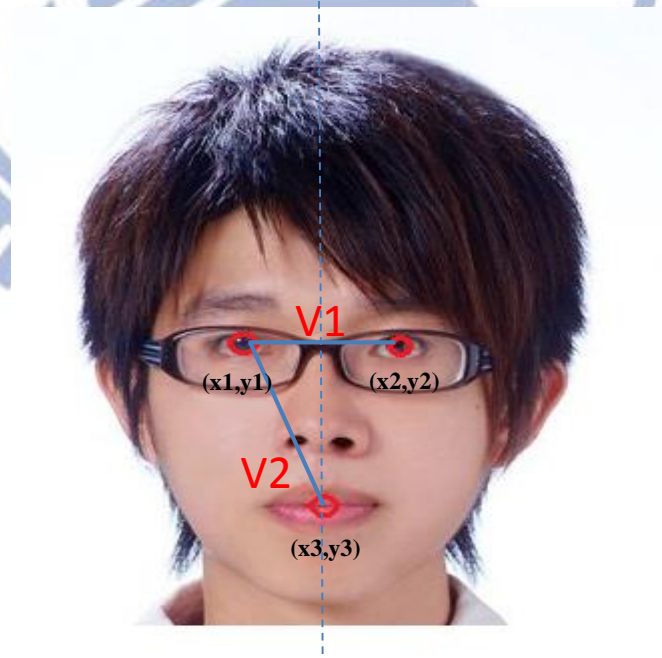


Figure 3.2.1 Geometric facial features.

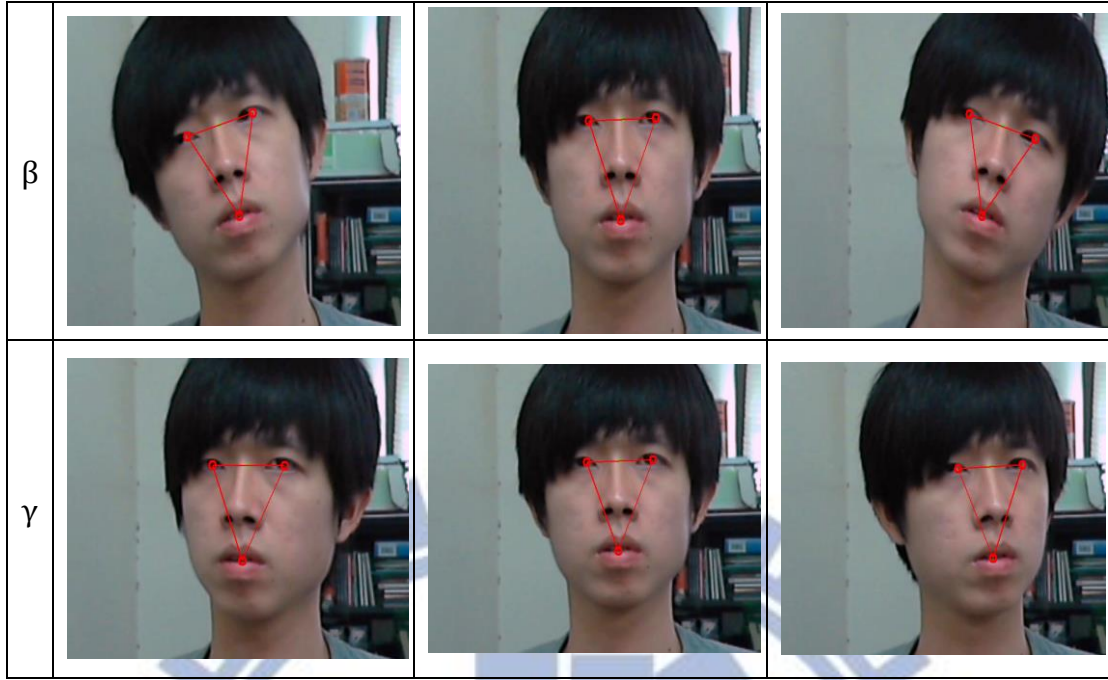


Table 3.1 Two angle of human head geometric.

3.3 Head Attitude Estimation System Design

Figure 3.3.2 shows the human face including right-eye coordinates, left-eye coordinates, mouth coordinates and four mask face coordinates. The three facial points which are left-eye, right-eye and mouth coordinate are (x_1, y_1) , (x_2, y_2) and (x_3, y_3) , respectively. The facial vector had been introduced in 3.2 section and the mask of the human face points are (x_4, y_4) , (x_5, y_5) , (x_6, y_6) and (x_7, y_7) , after subtracting them with the origin point we get four vector (V_3, V_4, V_5, V_6) .

Before using the proposed method, first we need to normalize $V_1, V_2, V_3, V_4, V_5, V_6$ and two angles of the human head geometry which are β and γ , the equations are from (3.18) to (3.30) :

$$Dis = \sqrt{(vx1)^2 + (vy1)^2} \quad (3.18)$$

$$f1 = Vx1/ Dis \quad (3.19)$$

$$f2 = Vy1/ Dis \quad (3.20)$$

$$f3 = Vx2/ Dis \quad (3.21)$$

$$f4 = Vy2/ Dis \quad (3.22)$$

$$f5 = Vx3/ Dis \quad (3.23)$$

$$f6 = Vy3/ Dis \quad (3.24)$$

$$f7 = Vx4/ Dis \quad (3.25)$$

$$f8 = Vy4/ Dis \quad (3.26)$$

$$f9 = Vx5/ Dis \quad (3.27)$$

$$f10 = Vy5/ Dis \quad (3.28)$$

$$f11 = Vx6/ Dis \quad (3.29)$$

$$f12 = Vy6/ Dis \quad (3.30)$$

where Dis is magnitude of V1.

Two angles of the human head geometry are limited between -30 and 30 degrees because of the range of a normal human head geometry. Furthermore, before training the angles have to be normalized, the angle degrees are shown in (3.31) and (3.32):

$$\beta_n = \beta/30 \quad (3.31)$$

$$\gamma_n = \gamma/30 \quad (3.32)$$

where β_n and γ_n are normalize range between 1 and -1.

The artificial neural network (ANN) is used to train the six normalize vectors and normalize β angle, where two normalized vectors are the inputs which are right-eye to left-eye and right-eye to mouth, other normalized vectors are the outputs which are four normalize vectors belong to extended normalize vectors of every four extend points to subtract origin points . The normalized β angle and four extend normalized vectors are learned by the neural network structure in Figure3.3.3 based on the back-propagation. The other normalized angle γ of human head geometric uses the same ANN structure to train four normalize vectors and normalize γ angle which shows in the Figure 3.3.4.

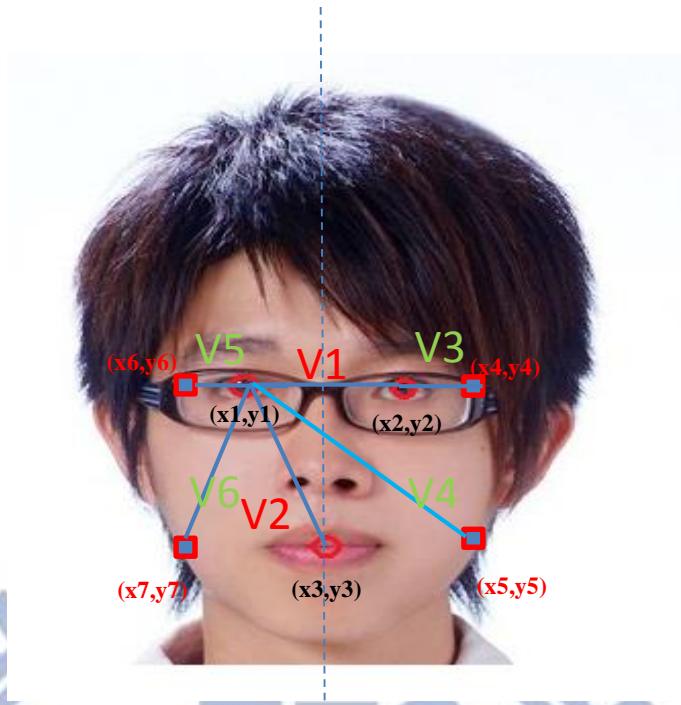


Figure.3.3.2 Train vectors on human face.

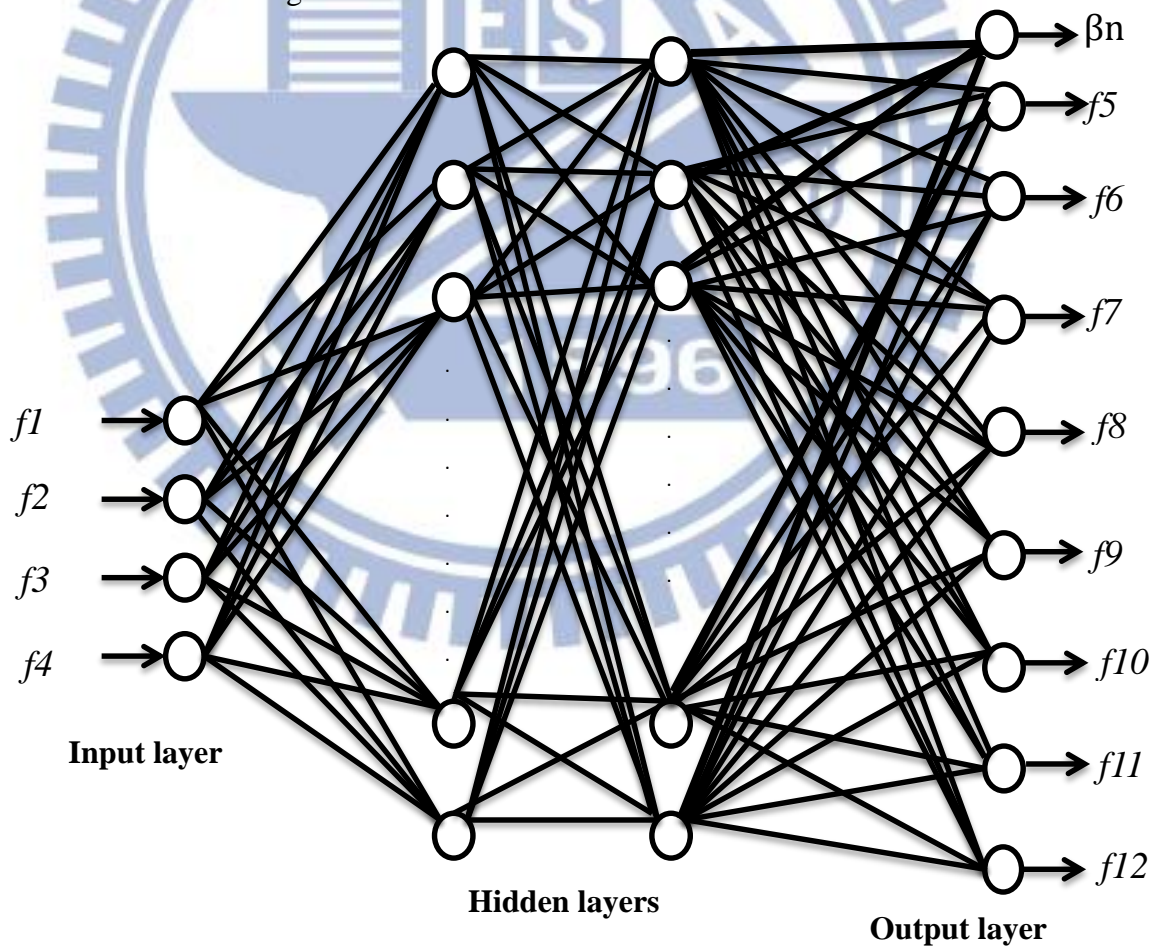


Figure 3.3.3 Neural network structure for β angle of the human head geometric.

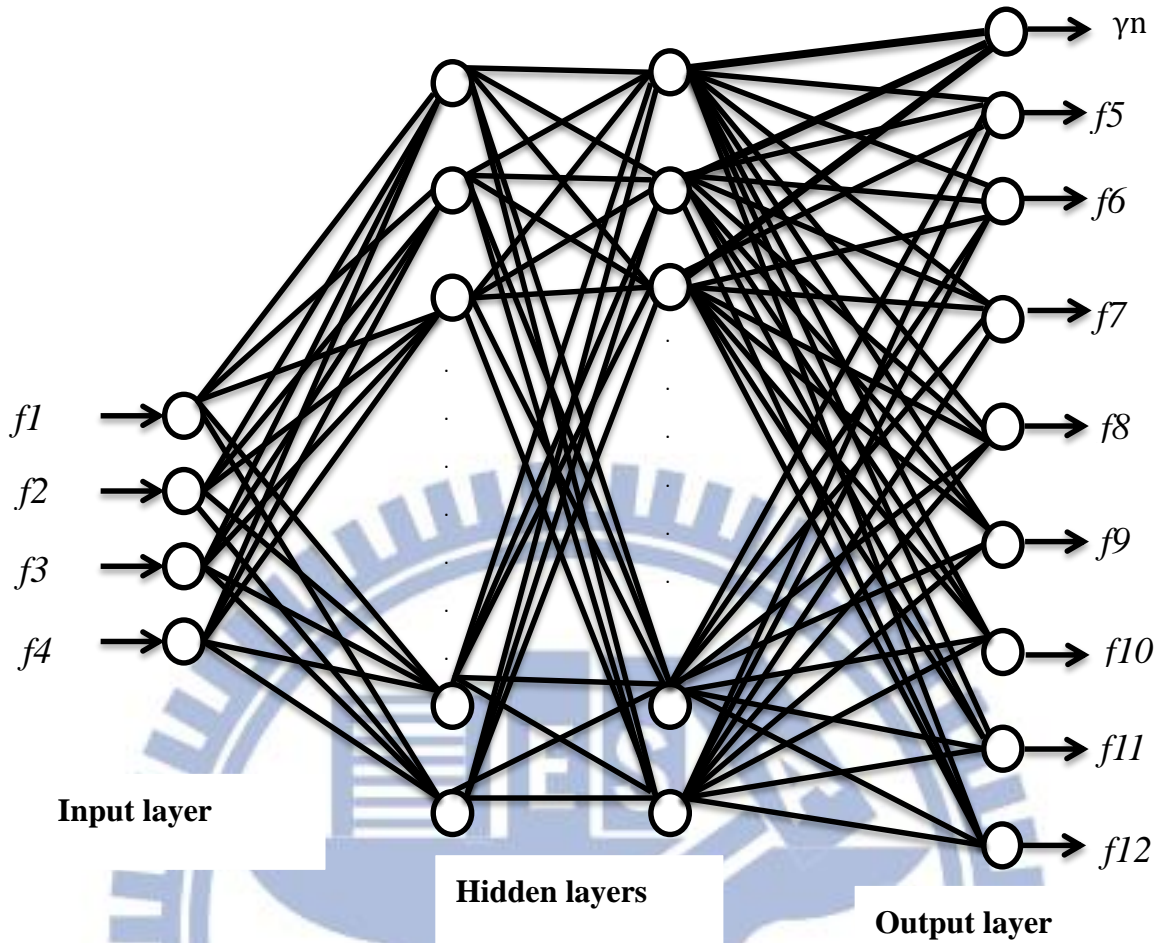


Figure 3.3.4 Neural network structure for γ angle of the human head geometric.

For β angle of human head geometric neural network (β NN) based on the structure in Fig3.3.3, there include one input layer with four neurons, two hidden layer each layer with 30 neurons, and one output layer with 9 neurons. The facial features vectors are sent into 4 neurons of the input layer, represented by $C(p)$, $p=1, 2, 3, 4$, correspondingly. The p -th input neuron is connected to the q -th neuron, $q=1, 2, 3, 4, \dots, 30$, of the first hidden layer with weighting $W_{c1}(p, q)$. Besides, the q -th neuron of the first hidden layer is also with an extra bias $b_{c1}(q)$. Hence, there exists a weighting array $W_{c1}(p, q)$ of dimension 4×30 . The q -th of first hidden neuron is connected to the r -th neuron, $r=1, 2, 3, 4, \dots, 30$, of the second hidden layer with weighting $W_{c2}(q, r)$. There exists a weighting array $W_{c2}(q, r)$ of dimension 30×30 . Besides, the r -th neuron of the second hidden layer is also with an extra bias $b_{c2}(r)$. Finally, the r -th neuron of the second

hidden layer is connected to the s -th neuron, $s=1, 2, 3, 4, \dots, 9$ with weighting array $W_{c3}(r,s)$ of dimension 30×9 , and a bias $bc_3(s)$ is added to the output neuron.

Let the activation function of the hidden layer be the linear transfer function which shows Fig.3.3.5 and the q -th output neuron $O_{c1}(q)$ is expressed as:

$$O_{c1}(q) = \text{purelin}(n_1(q)), q=1, 2, 3, 4, \dots, 30. \quad (3.33)$$

where

$$n_1(q) = \sum_{p=1}^4 W_{c1}(p,q)C(p) + bc_1(q) \quad (3.34)$$

Let the activation function of the second hidden layer be the linear transfer function which shows Fig.3.3.3 and the second hidden neuron $O_{c2}(r)$ is expressed as:

$$O_{c2}(r) = \text{purelin}(n_2(r)), r=1, 2, 3, 4, \dots, 30. \quad (3.35)$$

where

$$n_2(r) = \sum_{q=1}^{20} W_{c2}(q,r)O_{c1}(q) + bc_2(r) \quad (3.36)$$

Let the activation function of the output layer be the linear transfer function which shows Fig.3.3.3 and output neuron $O_{c3}(s)$ is expressed as:

$$O_{c3}(s) = \text{purelin}(n_3(s)), s=1, 2, 3, \dots, 9. \quad (3.37)$$

where

$$n_3 = \sum_{r=1}^{20} W_{c3}(r,s)O_{c2}(r) + bc_3(s) \quad (3.38)$$

The above operations are shown in Fig.3.3.6.

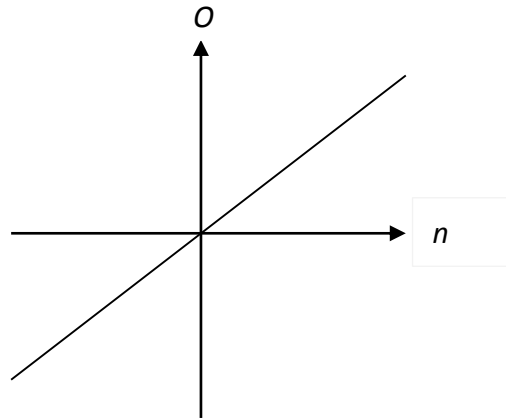


Figure.3.3.5 linear transfer function.

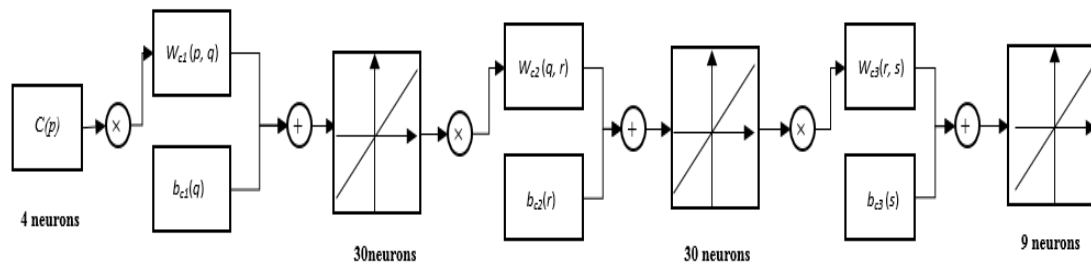


Figure.3.3.6 BNN.

For γ angle of human head geometric neural network (γ NN) based on the structure in Figure 3.3.4, there include one input layer with four neurons, two hidden layer each layer with 25 neurons, and one output layer with 9 neurons. The facial features vectors are sent into 4 neurons of the input layer, represented by $G(p)$, $p = 1, 2, 3, 4$, correspondingly. The p -th input neuron is connected to the q -th neuron, $q = 1, 2, 3, 4, \dots, 25$, of the first hidden layer with weighting $W_{G1}(p, q)$. Besides, the q -th neuron of the first hidden layer is also with an extra bias $b_{G1}(q)$. Hence, there exists a weighting array $W_{G1}(p, q)$ of dimension 4×25 . The q -th of first hidden neuron is connected to the t -th neuron, $t = 1, 2, 3, 4, \dots, 25$, of the second hidden layer with weighting $W_{G2}(q, t)$. There exists a weighting array $W_{G2}(q, t)$ of dimension 25×25 . Besides, the t -th neuron of the second hidden layer is also with an extra bias $b_{G2}(t)$. Finally, the t -th neuron of the second hidden layer is connected to the u -th neuron, $u = 1, 2, 3, 4, \dots, 9$ with

weighting array $W_{G3}(t, u)$ of dimension 25×9 , and a bias $b_{G3}(u)$ is added to the output neuron.

Let the activation function of the hidden layer be the linear transfer function which shows Fig.3.3.5 and the q -th output neuron $O_{G1}(q)$ is expressed as:

$$O_{G1}(q) = \text{purelin}(n_1(q)), q = 1, 2, 3, 4, \dots, 25. \quad (3.39)$$

where

$$n_1(q) = \sum_{p=1}^4 W_{G1}(p, q)G(p) + b_{G1}(q) \quad (3.40)$$

Let the activation function of the second hidden layer be the linear transfer function which shows Fig.3.3.5 and the second hidden neuron $O_{G2}(t)$ is expressed as:

$$O_{G2}(t) = \text{purelin}(n_2(t)), t = 1, 2, 3, 4, \dots, 25. \quad (3.41)$$

where

$$n_2(t) = \sum_{q=1}^{20} W_{G2}(q, t)O_{G1}(q) + b_{G2}(t) \quad (3.42)$$

Let the activation function of the output layer be the linear transfer function which shows Fig.3.3.5 and output neuron $O_{G3}(u)$ is expressed as:

$$O_{G3}(u) = \text{purelin}(n_3(u)), u = 1, 2, 3, \dots, 9. \quad (3.43)$$

where

$$n_3 = \sum_{t=1}^{20} W_{G3}(t, u)O_{G2}(t) + b_{G3}(u) \quad (3.44)$$

The above operations are shown in Fig.3.3.7.

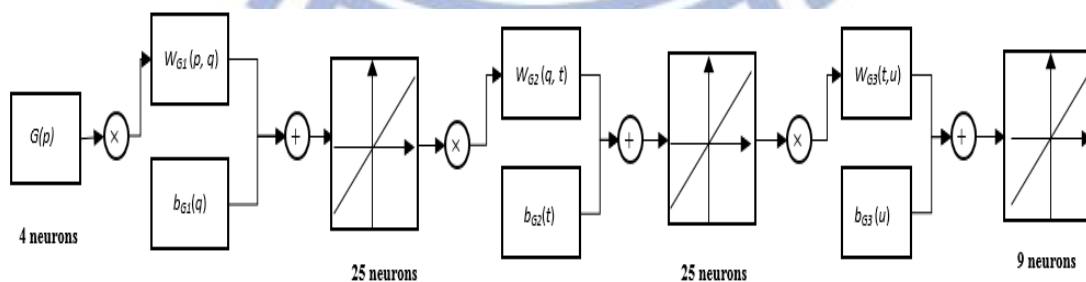


Figure.3.3.7 γ NN.

Chapter 4 Experiment Result

In the previous chapter, the three main steps of the proposed head attitude estimation system are introduced. In this chapter, the experiment results of each step will be expressed and the result of the proposed algorithm will be obtained by MATLAB R2011b.

4.1 Facial Features Detection

For facial feature detection, there two features, eyes and mouth are detected in thesis. A set of experimental results will be used to show the effectiveness and efficiency of the proposed system. In this dissertation, a webcam is applied to the experiment in order to catch appropriate images. In the previous chapter had introduced how to detection human eyes and human mouth, here 300 images under different people human, with size of (640× 480) for test. The table 4.1 shows eyes detection and mouth accuracy rate. After experiment the results achieve good performance which get high accuracy rate to detect human eyes and mouth. The figure 4.1 shows different people facial features success detection result.

Facial features detection	Accuracy rate %
Left-eye correct	96.66 % (290/300)
Right-eye correct	95.66% (287/300)
Mouth	91.6 % (275/300)
Three feature correct	91.6 % (275/300)

Table 4.1 Accuracy rate of three facial features detection.



Figure 4.1.1 the result of different people facial detection.

4.2 Geometric Facial Features

This thesis has to statistic more information of the geometric facial feature which are β and γ . We don't know human head rotation information, therefore, this thesis manufactured a stereo facial model showing in the figure 4.2.1 which are showed the appearance including a protractor and indicator which achieve more precise in the statistic human attitude estimation system information. On the table, there are seven points shows figure 4.2.2. Thesis uses red, green, blue colors represent right-eye, left-eye, mouth and four points of mask human face.



Figure 4.2.1 the turntable appearance.

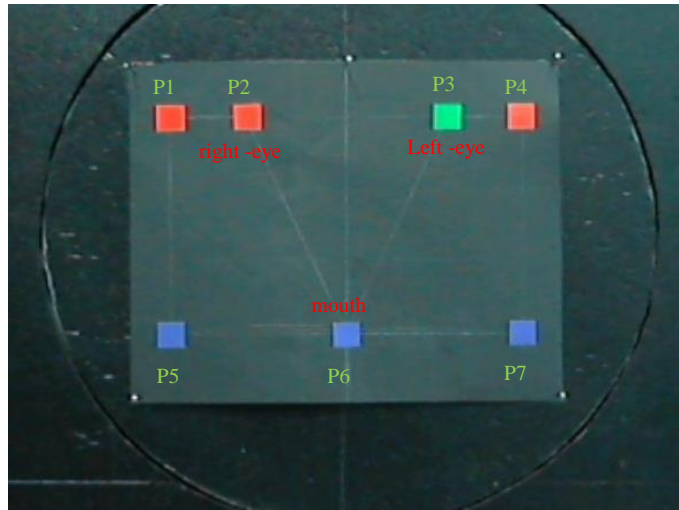


Figure 4.2.2 RGB color label seven points.

After labeling red, green, blue colors automatic detected labeling, therefore, thesis uses two steps to detect red, green and blue colors which shows figure 4.2.3, first, gray level three colors and second step is binary three colors, finally, we can successful automatic detect red, green, blue colors.

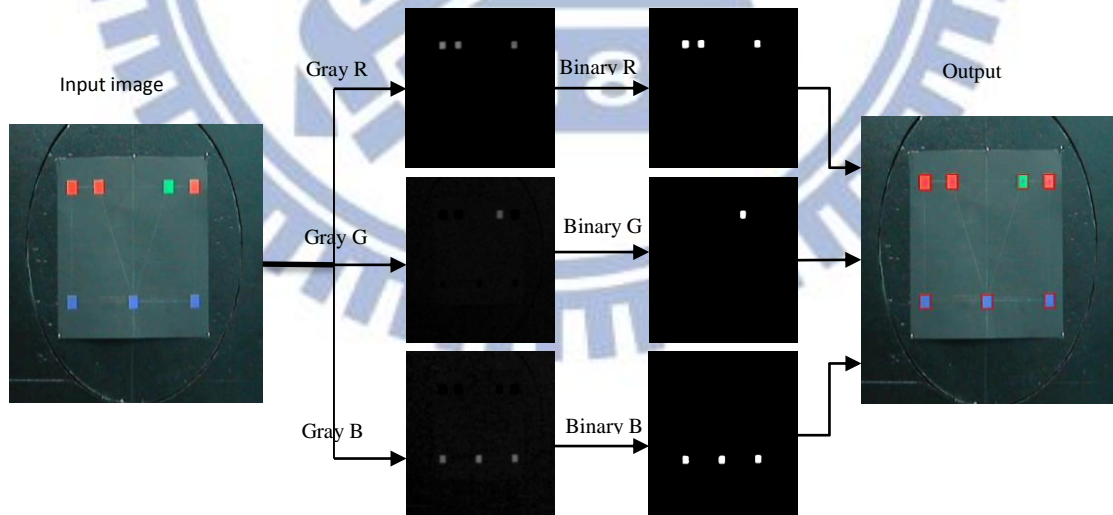


Figure 4.2.3 detection RGB colors.

4.3 Head Attitude Estimation System Design

This section shows the head attitude estimation system design experiment result, including the neural network off-line training, test the neural network off-line training performance and head attitude estimation final result.

4.3.1 Neural Network Off-line Training

This section focuses on the off-line training of the two neural networks, β NN and γ NN, used in the head attitude estimation. It is known that different types of the head attitude estimation system requires different types of neural networks. Besides, all the neural networks are designed to have three layers, the input layer, output layer and hidden layer. The number of neurons of the input layer is chosen to the same as the number of input data, so is the number of neurons of the output layer, corresponding to the output data. However, how many neurons are needed for the hidden layer should be determined by experiments, via neural network off-line training in this thesis. According to the experiment, β NN performance are influenced to the distance between webcam and stereo facial model. The thesis divides two cases to determine off-line training parameter.

Case1:

The 30CM distance from webcam to stereo facial model for case1. First, let's find the suitable number of neurons of β NN and γ NN which will be applied to head attitude estimation system. The off-line training of the β NN is executed in difference cases, named as β NN- k, h where k is the number of neurons of the first hidden layer and is chosen from 10 to 30, where h is the number of the second hidden layer and is chosen from 10 to 30. Based on the off-line training, it can be found that the performance is

best while the pair of (k, h) are changed from $(20, 20)$ to $(30, 30)$, as show in Table 4.2. The performance from $(20, 20)$ to $(30, 30)$ are the same, thus, thesis choses $(20, 20)$ pair less neurons than $(30, 30)$. Similarly, Table 4.3 shows the results of the off-line training of the γ NN. Obviously, the γ NN-25, 25 is the best structure with minimal learning mean square error (MSE). Hence, the γ NN-25, 25 will be used in the HAES.

Experiments	β NN-10,0	β NN-20,0	β NN-30,0	β NN-10,10	β NN-20,20	β NN-25,25	β NN-30,30
Learning Time(sec)	7.1	4.8	6.4	13.8	18.31	29.364	31.4
Epochs	321	163	259	273	383	600	658
First hidden neurons	10	20	30	10	20	25	30
Second hidden neurons	0	0	0	10	20	25	30
Tolerance	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}
MSE (performance)	3.6×10^{-4}	3.83×10^{-4}	2.7×10^{-4}	5.6×10^{-5}	2.59×10^{-5}	2.67×10^{-5}	2.66×10^{-5}
Input neuron	4	4	4	4	4	4	4
Output neuron	9	9	9	9	9	9	9

Table 4.2 30CM distance of β NN off-line training parameter.

Experiments	γ NN-10, 0	γ NN-20, 0	γ NN-30, 0	γ NN-10, 10	γ NN-20, 20	γ NN-25, 25	γ NN-30, 30	γ NN-35, 35
Learning Time	17.24	12.3	19.75	20.48	28.14	31.9	33.8	36.7
Epochs	640	319	427	540	713	720	729	778
First hidden neurons	10	20	30	10	20	25	30	35
Second hidden neurons	0	0	0	10	20	25	30	35
Tolerance	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}
MSE (performance)	3.7×10^{-4}	3.6×10^{-4}	3.6×10^{-4}	3.4×10^{-4}	4.61×10^{-5}	4.6×10^{-5}	4.6×10^{-5}	4.61×10^{-5}
Input neuron	4	4	4	4	4	4	4	4
Output neuron	9	9	9	9	9	9	9	9

Table 4.3 γ NN off-line training parameter

Case2:

It is all range of distance from webcam to stereo facial model for case1. First, let's find the suitable number of neurons of β NN which will be applied to head attitude detection system. The off-line training of the β NN is executed in difference cases, named as β NN- k, h where k is the number of neurons of the first hidden layer and is chosen from 10 to 40, where h is the number of the second hidden layer and is chosen from 10 to 40. Based on the off-line training, it can be found that the performance is best while the pair of (k, h) are changed from $(30, 30)$ to $(40, 40)$, as show in Table 4.4. The performance from $(30, 30)$ to $(40, 40)$ are the same, thus, thesis choses $(30, 30)$ pair less neurons than $(40, 40)$.

Experiments	γ NN-10, 0	γ NN-20, 0	γ NN-30, 0	γ NN-10, 10	γ NN-20, 20	γ NN-25, 25	γ NN-30, 30	γ NN-35, 35	γ NN-40, 40
Learning Time	17.24	12.3	19.75	20.48	28.14	31.9	33.8	36.7	36.7
Epochs	640	319	427	540	713	720	729	778	778
First hidden neurons	10	20	30	10	20	25	30	35	35
Second hidden neurons	0	0	0	10	20	25	30	35	35
Tolerance	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}	10^{-10}
MSE (performance)	3.7×10^{-4}	3.6×10^{-4}	3.6×10^{-4}	3.4×10^{-4}	4.61×10^{-5}	4.6×10^{-5}	2.57×10^{-5}	2.61×10^{-5}	2.58×10^{-5}
Input neuron	4	4	4	4	4	4	4	4	4
Output neuron	9	9	9	9	9	9	9	9	9

Table 4.4 All distance range of β NN off-line training parameter.

4.3.2 Test neural Network Performance

In this section, this experiment will test the performance both of β NN and γ NN. According to the experiment, the β NN performance is influenced by the distance between webcam and stereo facial model. The figure 4.3.1 shows the different distance between webcam and stereo facial model image. There are two case in the β NN. The case1 is training data in 30 CM of the distance and case 2 is training data in all ranges. The table 4.4 shows the training data in case1 of β NN accuracy rate relating two angle-scale and distance. In the table 4.5, obviously, the distance from 20CM to 30CM had good performance, the distance from 35CM to 40CM is medium performance and the distance from 45CM to 50CM had bad performance. The reason is the stereo facial model closing the webcam and the pixels change a measure of the β NN information is very obvious. Otherwise, the stereo facial model is far away from the webcam and the pixels change a measure of the β NN information extremely awful. The figure 4.3.2

shows the histogram relates to accuracy rate and angle-scale of orientation β NN in case1. In this plot, the accuracy rate accompanies the distance becoming extremely awful. In table 4.4, it is clearly observed that the angle-scale from $\pm 0 \sim \pm 20$ had a good performance and the angle-scale $\pm 21 \sim \pm 30$ was extremely awful. The reason is the angle-scale $\pm 21 \sim \pm 30$ rotation was too extreme, hence the information of the β NN judge the accuracy angle. The figure 4.3.3 shows in the different angle-scale curves in case1, the β NN relating two accuracy rates and distance. Obviously, the curves shows the good performance when the angle-scale within $\pm 0 \sim \pm 20$ and the curves shows the bad performance when the angle-scale within $\pm 21 \sim \pm 30$. The table 4.6 shows the training data including all distance range, the table shows performance better than the training data only distance in 30CM. The figure 4.3.4 shows histogram relates to accuracy rate and angle-scale of orientation β NN of the distance in all ranges. Figure 4.3.5 shows the different angle-scale curves in case2 .Table 4.7 shows γ NN accuracy rate relating two angle-scale and distance. Obviously, γ NN performance are not influenced by the distance between webcam and stereo facial model. The reason is the pixels change a measure of the γ NN information are very obvious in the per-independent angle-scale. The figure 4.3.6 shows histogram relating two accuracy rate and angle-scale of orientation γ NN. Obviously, the performance is very well in different distances. The figure 4.3.7 shows in the different angle-scale curves, theyNN relating two accuracy rate and distance. In this plot, the accuracy rate is not influenced by the different angle- scale curves. Finally, the figure 4.3.8 shows the HAES final result.

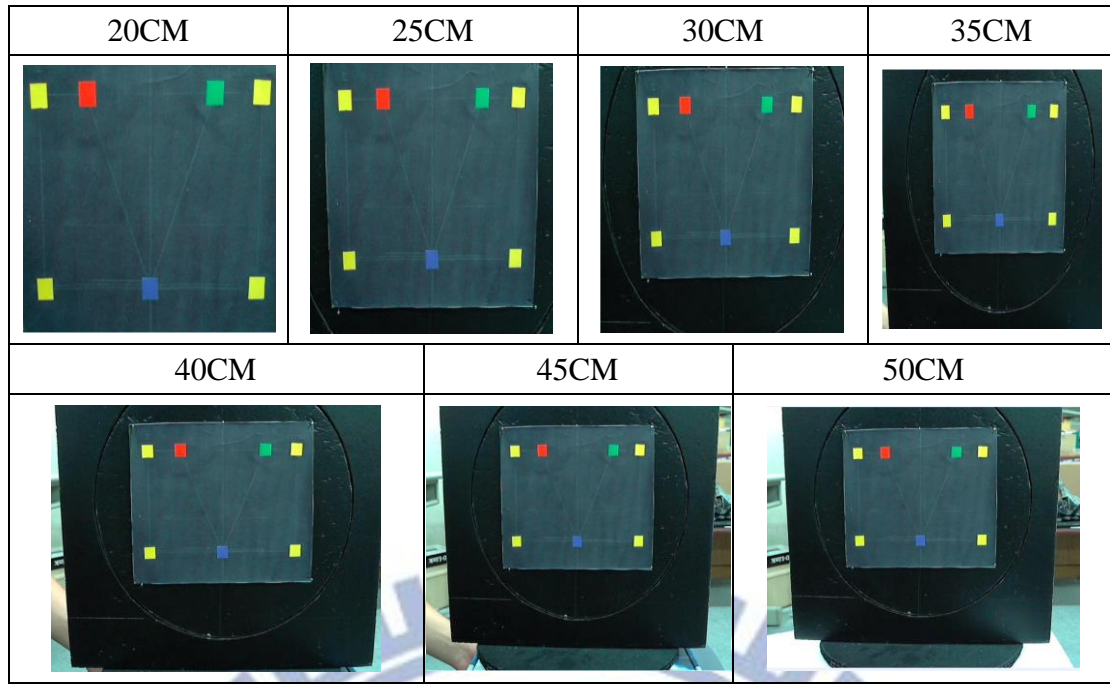


Figure 4.3.1 The different distance between webcam and stereo facial model.

Accuracy rate Distance	Accuracy rate					
	0°~5°	6°~10°	11°~15°	16°~20°	21°~25°	26°~30°
20CM	100%	100%	96.5%	100%	100%	96.5%
25CM	96.5%	96.5%	95.5%	99%	95.5%	95.5%
30CM	97%	99%	100%	99%	93%	91.5%
35CM	95%	85.5%	97.5%	91%	71.5%	64%
40CM	96%	92%	92.5%	88.5%	64.5%	63%
45CM	79%	56.5%	51.5%	51%	29.5%	20%
50CM	76.5%	49.5%	48%	47.5%	19.5%	20%
Accuracy rate Distance	Accuracy rate					
	0°~-5°	-6°~-10°	-11°~-15°	-16°~-20°	-21°~-25°	-26°~-30°
20CM	100%	100%	98.5%	100%	100%	97.5%
25CM	97.5%	98%	100%	100%	96%	91%
30CM	100%	100%	99%	100%	92.5%	91%
35CM	96.5%	86.5%	95.5%	89.5%	73.5%	68.5%
40CM	97.5%	90.5%	89.5%	91.5%	60.5%	61.5%

45CM	73.5%	53.5%	56.5%	52.5%	31.5%	19%
50CM	79.5%	47.5	46.5%	48%	21.5%	18.5%

Table 4.5 the β NN accuracy rate relating to angle-scale and distance in case1.

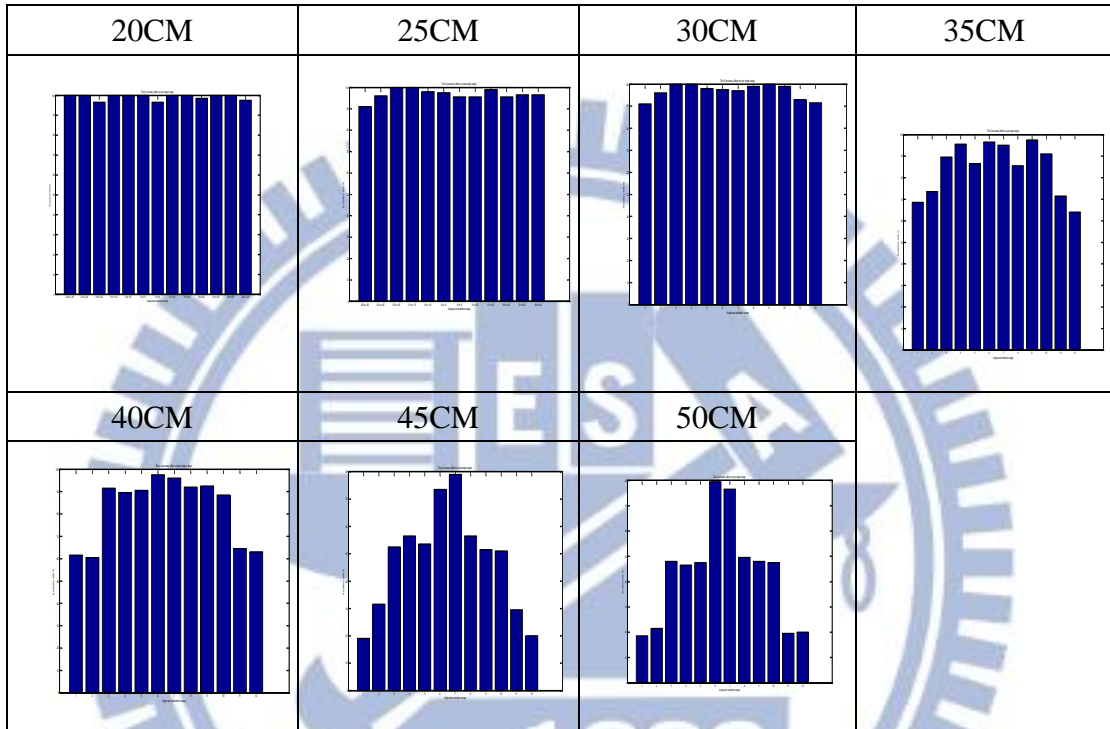


Figure 4.3.2 A histogram relates to accuracy rate and angle-scale of β NN in case1.

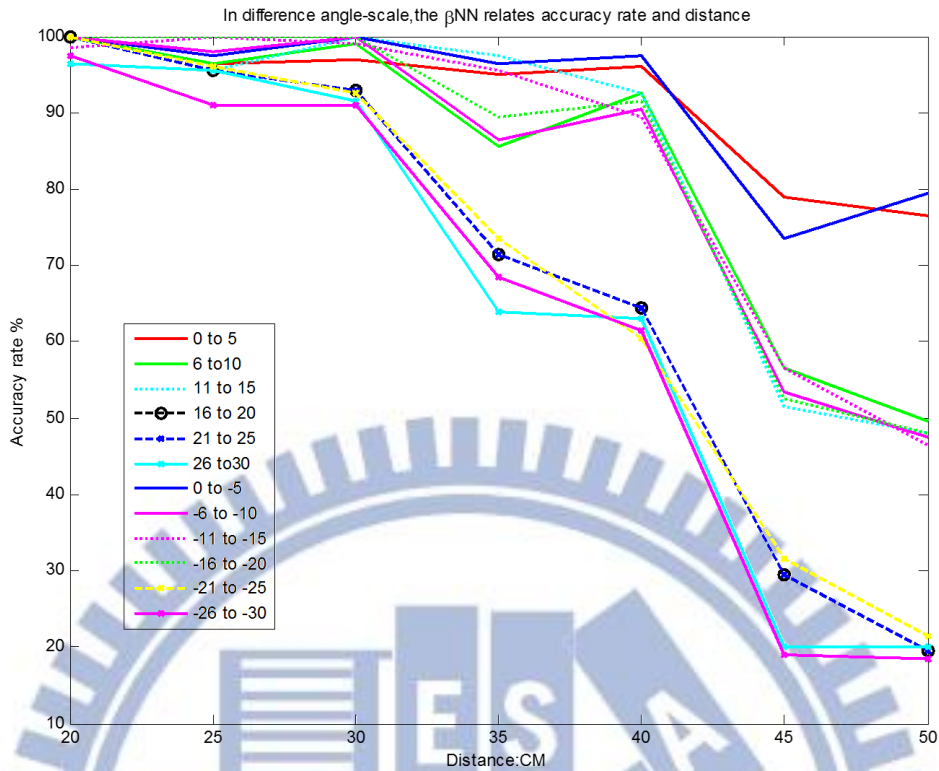


Figure 4.3.3 In different angle-scale curves, β NN relates to accuracy rate and distance in case 1.

Accuracy rate \ Distance	Angle Scales					
	0°~5°	6°~10°	11°~15°	16°~20°	21°~25°	26°~30°
20CM	100%	100%	97.5%	100%	100%	96.5%
25CM	98.5%	98.5%	97.5%	99%	95.5%	95.5%
30CM	96%	99%	100%	99%	95%	91.5%
35CM	97%	89.5%	96.5%	91%	78.5%	68%
40CM	98%	93%	94.5%	91.5%	69.5%	69%
45CM	85%	60%	55.5%	53%	31.5%	23%
50CM	81.5%	55%	53.5%	50.5%	23.5%	26%
Accuracy rate \ Distance	Angle Scales					
	0°~-5°	-6°~-10°	-11°~-15°	-16°~-20°	-21°~-25°	-26°~-30°
20CM	100%	100%	98.5%	100%	100%	96.5%
25CM	98.5%	98%	100%	100%	96%	93%
30CM	100%	100%	99%	100%	92.5%	94%

35CM	97.5%	89.5%	95.5%	91.5%	73.5%	71.5%
40CM	98.5%	93.5%	91.5%	93.5%	60.5%	65.5%
45CM	80.5%	56.5%	59.5%	56.5%	31.5%	23%
50CM	83.5%	51%	51.5%	57%	21.5%	22.5%

Table 4.6 the β NN accuracy rate relating to angle-scale and distance in case2.

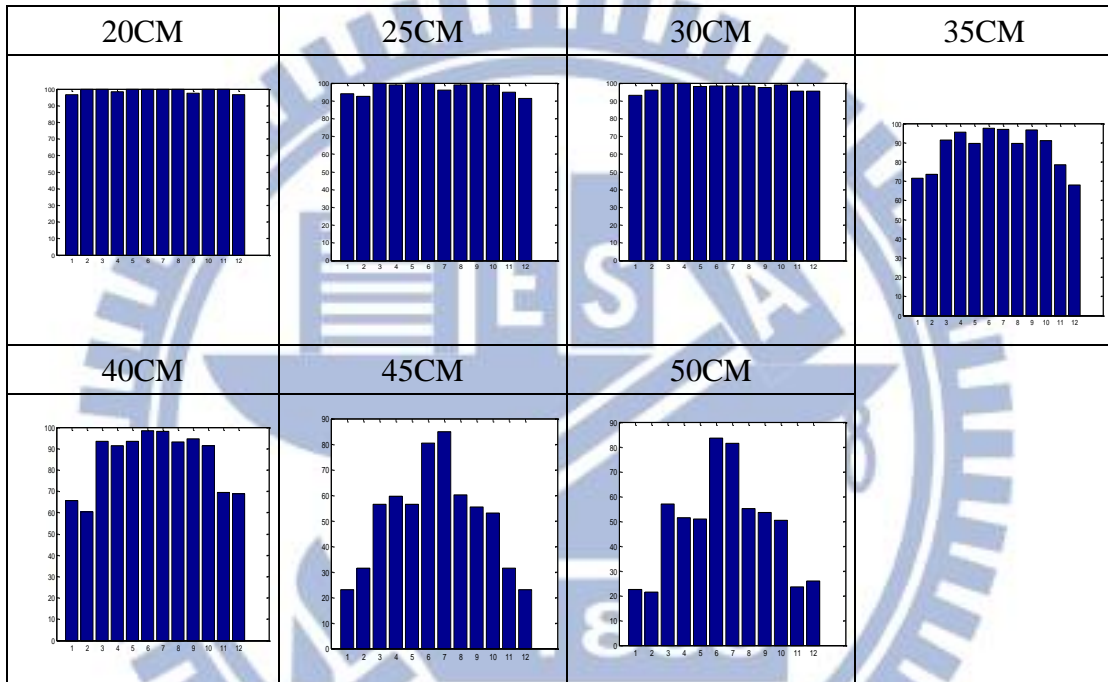


Figure 4.3.4 A histogram relates to accuracy rate and angle-scale of β NN in case2.

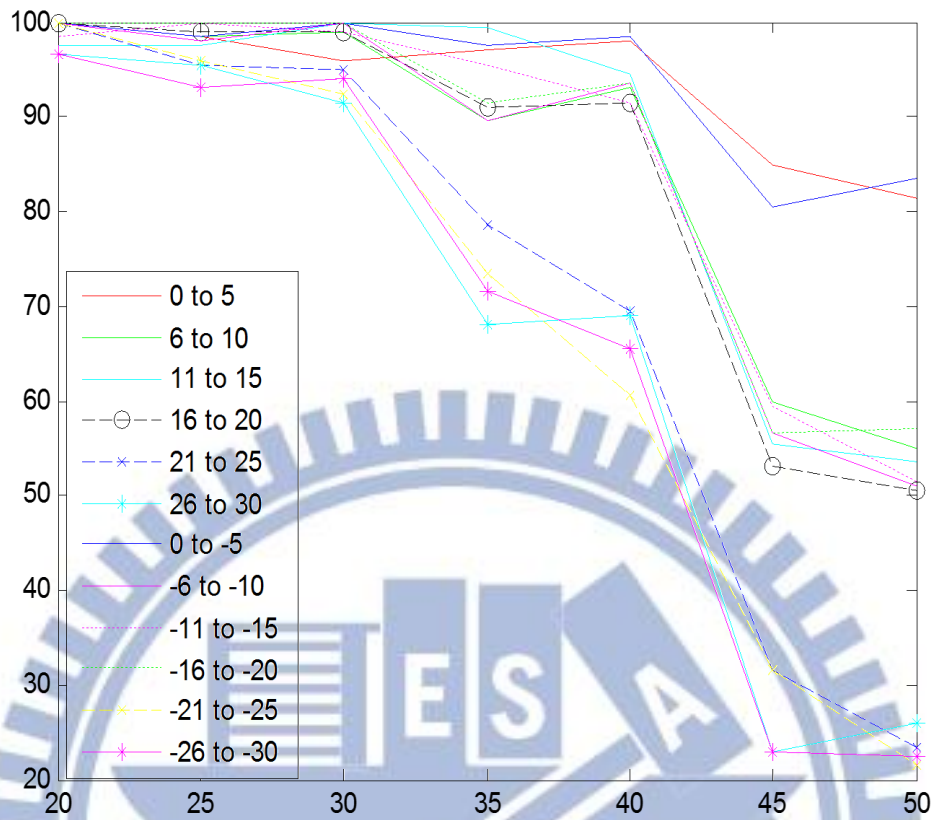


Figure 4.3.5 In different angle-scale curves, β NN relates to accuracy rate and distance in case2.

Accuracy rate \ Distance	Angle Scale					
	0°~5°	6°~10°	11°~15°	16°~20°	21°~25°	26°~30°
20CM	98.5%	99.5%	98%	99.5%	98%	100%
25CM	99.5%	100%	98.5%	99.5%	99.5%	99%
30CM	98.5%	99.5%	98%	98.5%	100%	99.5%
35CM	98%	98.5%	99%	98.5%	97.5%	100%
40CM	98.5%	100%	100%	98.5%	100%	100%
45CM	99.5%	98%	100%	98%	99%	99.5%
50CM	98.5%	100%	99.5%	98.5%	99%	99.5%

Accuracy rate \ Distance	0°~5°	-6°~-10°	-11°~-15°	-16°~-20°	-21°~-25°	-26°~-30°
20CM	98.5%	99.5%	98.5%	97.5%	99.5%	98.5%
25CM	99.5%	98.5%	99.5%	98%	98.5%	97.5%
30CM	99.5%	99.5%	98.5%	99.5%	98.5%	100%
35CM	97.5%	100%	98.5%	97.5%	99.5%	98%
40CM	100%	99.5%	98.5%	99%	98.5%	98%
45CM	99%	99.5%	100%	99%	99.5%	98%
50CM	99%	98%	99%	99%	98%	100%

Table 4.7 the γ NN accuracy rate relating to angle-scale and distance.

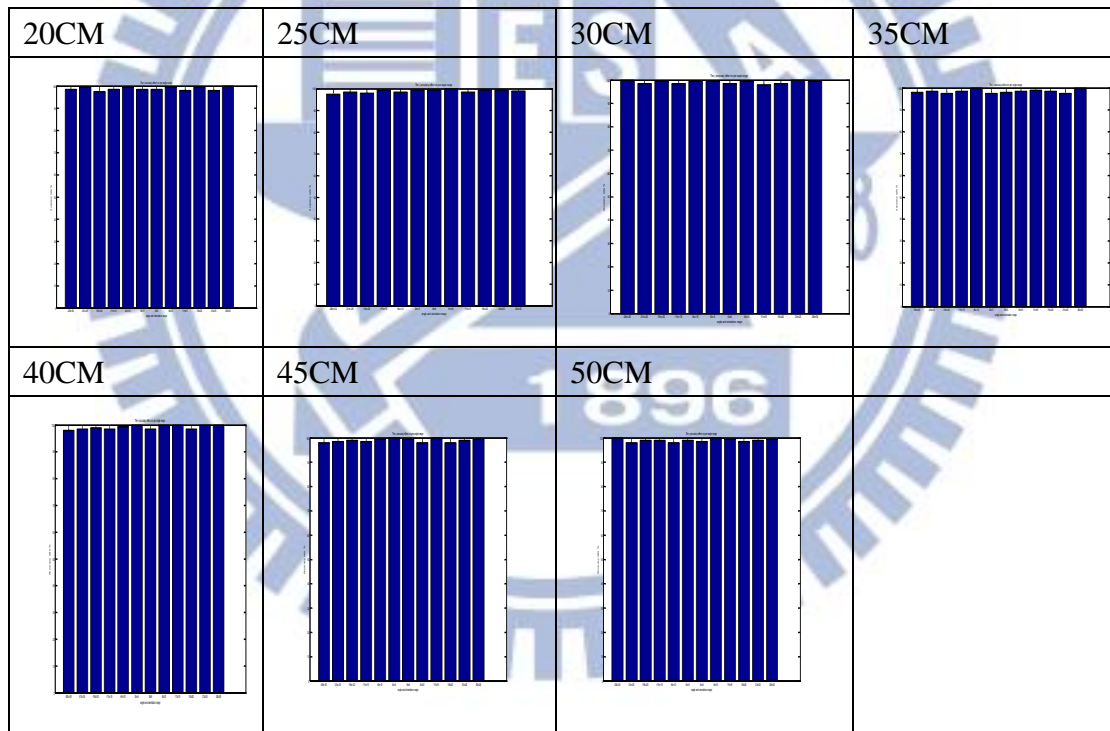


Figure 4.3.6 A histogram relates to accuracy rate and angle-scales of γ NN.

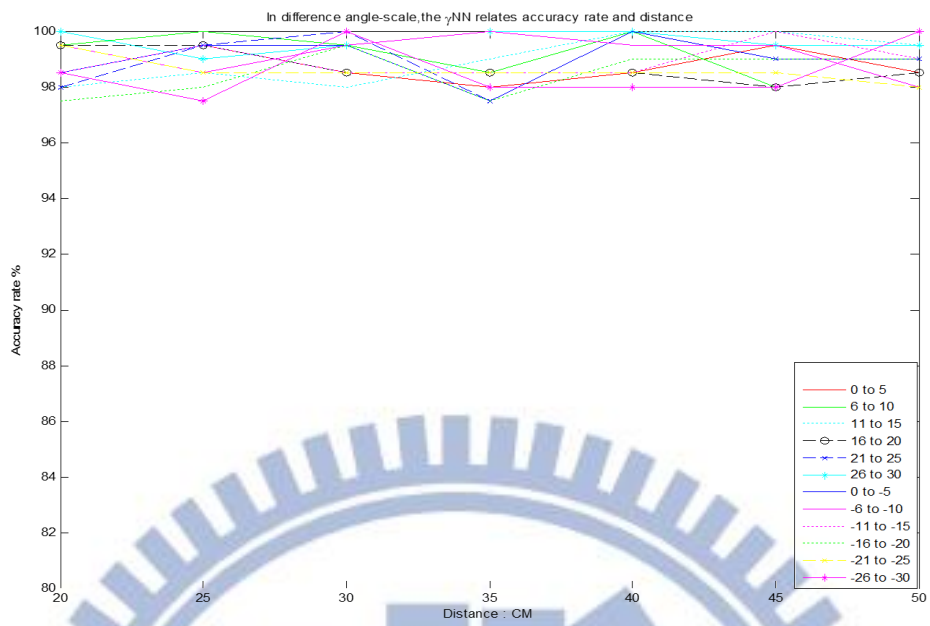


Figure 4.3.7 In different angle-scale curves, γ NN relates to accuracy rate and distance

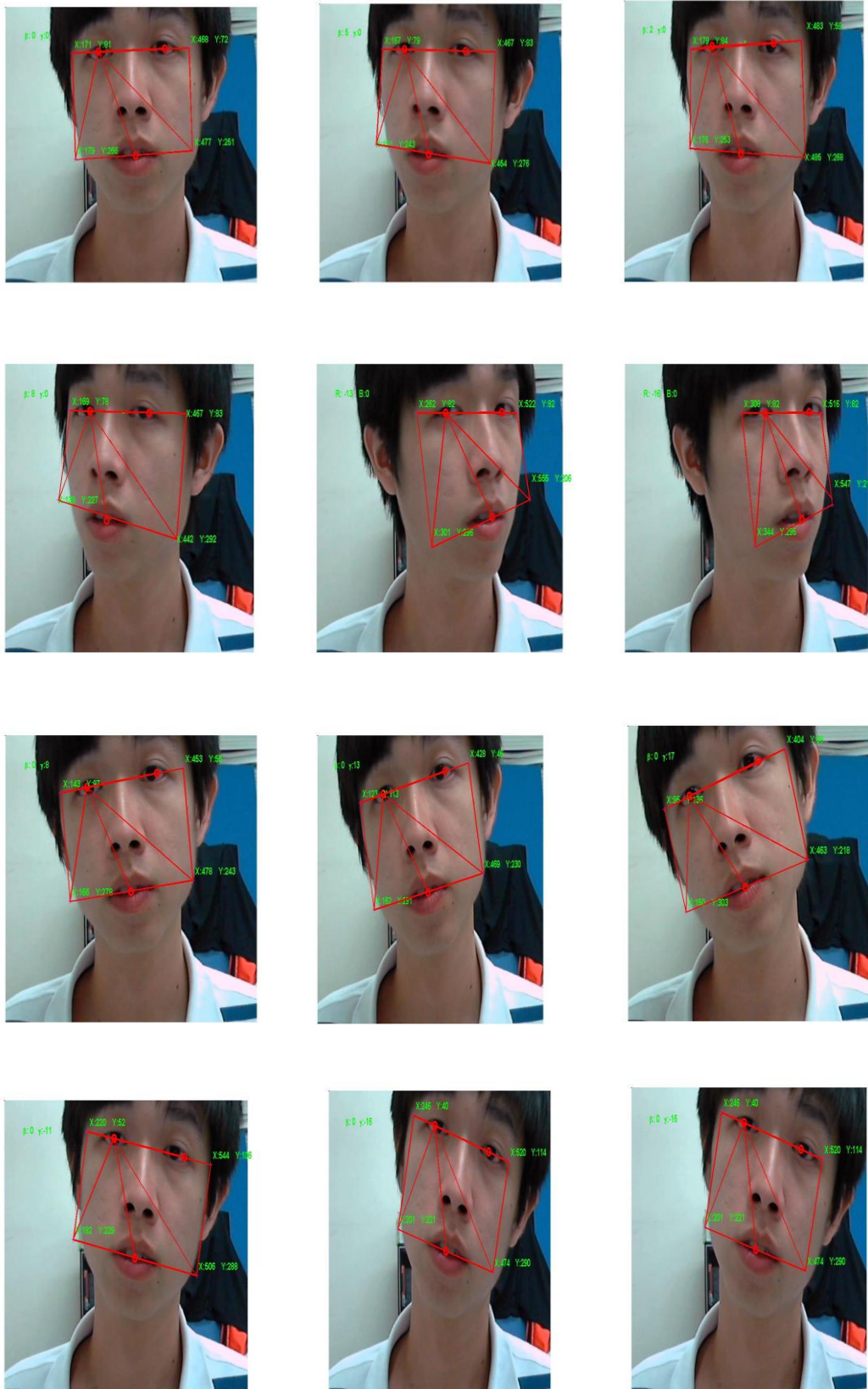


Figure 4.3.8 the HAES final results.

Chapter 5

Conclusions and Future Works

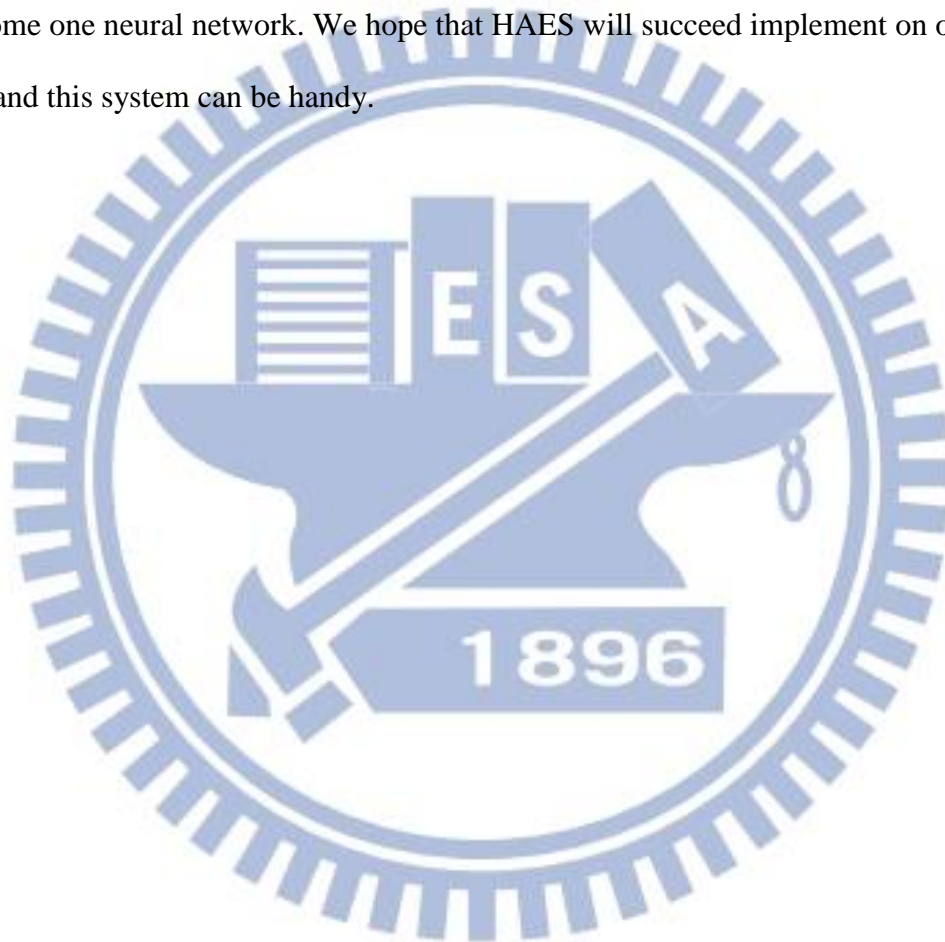
This thesis proposes an intelligent human attitude estimation system (HAES) designed to detect the face orientation and angle based on geometric facial features, which are both eyes and the mouth, two important facial features in HAES. There are three steps to complete the intelligent head attitude estimation. First, facial feature detection method for detecting faces in color images is proposed [23]. It is based on a robust skin region detector which provides face coordinates. Then using some simple rules derived from anthropological characteristics, eyes are selected within the face region, the mouth is selected based on the relation from the eyes and mouth characteristics. This method achieves a successful rate of 91.6% on the HAES application. It can also detect eyes and the mouth rotation. Second, build up a stereo facial model to simulate the head attitude which is able to adjust the face orientation and angle by seven detected points marked on the face model. Then record the seven detecting points on each image referring to a specific face orientation and angle, which will be used in neural network learning. Third, the HAES is completed by intelligent neural networks under supervised learning.

The HAES intelligent neural networks include both β NN and γ NN . Both β NN and γ NN relation to accuracy and distance is shown in experiment results. Table 4.3.1 and the table 4.3.2 show the accuracy rate as high as 97.3% within 30CM between webcam and human head.

This system is useful in many computer vision tasks such as mobile phone user identification, the user can be whether observing the screen or not, or man-machine interface used in detecting the state of vehicle driver.

The proposed intelligent HAES detects the facial features in accuracy rate equal

to 91.6%, which is accepted in thesis experiment but the accuracy rate is not enough for the commercial products which requires to be as high as 99%. Future work includes developing other detection facial features methods or algorithms for upward the accuracy rate. The HAES is restricted to the distance because the dots per inch are not enough to determine face rotation angle and orientation of the webcam, thus raising the dpi in the sequence of images will improve it. Finally, combine two neural networks to become one neural network. We hope that HAES will succeed implement on our daily life and this system can be handy.



Reference

- [1] A. Yuille, D. Cohen, and P. Hallinan, "Feature extraction from faces using deformable templates," in *Proc. IEEE Computer Soc. Con Computer Vision and Pat. Recog.* pp. 104-109, 1989.
- [2] P. Hallinan, "Recognizing human eyes," in *SPIE Proc.: Geometric Methods in Computer Vision*, vol. 1570, pp. 214-226, 1991.
- [3] M. Ahmed Fadzil, and H. Abu Bakar, "Human face recognition using neural networks," in *IEEE Int. Con Image Proc.* vol. 3, pp. 936-938, 1994.
- [4] T. C. Chang, and T. S. Huang, "Facial feature extraction from color images," *12th IAPR Int. Con Pattern Recognition*, vol. 2, pp. 39-43, October 1994.
- [5] J. Ohya, Y. Kitamura, F. Kishino, and N. Terashima, "Virtual space teleconferencing: real-time reproduction of 3d human images," *Visual Commun. and Image Representation*, vol. 6, no. 1, pp. 1-25, March 1995.
- [6] N. Rahman, K. Wei, and J. See, "RGB-H-CbCr Skin Color Model for Human Face Dectcion," *Proc. of the MMU International Symposium on Information & Communications Technologies*, 2006.
- [7] L. Bretzner, I. Laptev, and T. Lindeberg, "Hand Gesture Recognition using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering," *Proc. The 5th IEEE Internat. Conf. on Automatic Face and Gesture Recognition*, pp.423–428, May 2002.
- [8] Q. Peng, and X. Zhang, "Sensitive Image Recognition Technology Based on Eigenvectors," *Academic Journal of Southwest Jiaotong University*, pp.13–18, Jan, 2007.
- [9] Q. Zhang, S. Li, and H. Xiao, "Extracting regions of interest in medical images based on visual attention mechanism," *Application Research of Computers*, vol.

26 , Dec. 2009, pp. 4803-4805.

- [10] V. Vezhnevets, V. Sazono, and A. Andreeva, "A Survey on Pixel- Based Skin Color Detection Techniques," *Proc. Graphicon-2003*, Sep.2003, pp. 85–92.
- [11] K. C. Yow, and R. Cipolla, "Feature-based Human Face Detection," *Image and Vision Computing*, vol. 15, no. 9, pp. 713-735, 1997.
- [12] S. Abdallah, A. Lynn Abbott, and A.Mohamad, "A New Face Detection Technique using 2D DCT and Self Organizing Feature Map," *Proceeding of World of science, Engineering and technology*. Vol. 24, May. 2007, pp. 15 - 19.
- [13] J. Wu, X. Zhang, and F. Zhang, "The experiment research of edge detection in digital image," *Microcomputer Information (CONTROL & AUTOMATION)*, vol. 20, No. 5, pp: 106-107, 2004.
- [14] L.G. Roberts, "Machine Perception of Three-Dimension Solids," *Optical and Electro-Optimal Information Processing, England : Cambridge*, pp:99-19, 1965.
- [15] Sobel L, "Camera Models and Machine Perception," *CA: Stanford University*, pp:121, 1999.
- [16] D.C. Marr, E. Hildret, and Hildreth, "Theory of Edge Detection," *Theory of Edge Detection, vol.B275*, pp:187-217, 1980.
- [17] J. Prewitt, "Enhancement and Extraction," *Picture Process, New York : Academic Press*, pp:75-1970, 1970.
- [18] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-8, No. 1, pp: 679-699, 1986.
- [19] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey, " *Proc. of the IEEE*, vol. 83, no. 5, pp. 705-740, May 1995.

- [20] K. Aizawa, and T. S. Huang, "Model-based image coding: advanced video coding techniques for very low bit-rate applications," *Proc. of the IEEE*, vol. 83, no. 2, pp. 259-271, Feb. 1995.
- [21] C. Huang, and C. Chen, "Human facial feature extraction for face interpretation and recognition," *Pattern Recognition*, vol. 25, no. 12, pp. 1435-1444, 1992.
- [22] D. Reisfed, and Y. Yeshuran, "Robust detection of facial features by generalized symmetry," in *Proc. 11th Int. Conf. on Pat. Recog.* pp. 117-120, 1992.
- [23] D. Sidibe, P. Montesinos, and S. Janaqi, "A simple and efficient eye detection method in color images," *Author manuscript, published in International Conference Image and Vision Computing New Zealand*, version 1-8 Apr 2009.
- [24] W. Gonzalez, and S. Eddins, *Digital Image Processing Using Matlab, Image Processing*.
- [25] B. Hogarth, "Drawing the Human Head," *1st ed., New York: Watson-Guption*, 2000.