



國立臺灣科技大學  
高分子工程系  
碩士學位論文

學號：M9704307

---

應用二維人臉影像與三維點雲資料於  
人臉姿態估測之研究

Applying 2D face images and 3D point-clouds on  
estimation of face orientation

研 究 生：黃 勝 達  
指 導 教 授：邱 士 軒 博 士

中 華 民 國 九 十 九 年 七 月 二 十 六 日



M9704307



## 碩士學位論文指導教授推薦書

本校 高分子工程系 黃勝達(HUANG, SHENG-DA) 君

所提之論文：

應用二維人臉影像與三維點雲資料於人臉姿態估測之研究

係由本人指導撰述，同意提付審查。

指導教授：邱士軒

指導教授 邱士軒

2010 年 7 月 28 日



# 碩士學位考試委員審定書



M9704307

指導教授：邱士軒

本校 高分子工程系 黃勝達 君

所提之論文：

應用三維人臉影像與三維點雲資料於人臉姿態估測之研究

經本委員會審定通過，特此證明。

學校考試委員會

委員：

蘇清樹

邱士軒

邱士軒

溫哲彥

指導教授：

邱士軒

學程主任：

邱士軒

系(學程)主任、所長：

中華民國 99 年 7 月 26 日

## 摘要

我們修改 Chiunhsiun Lin(2001)[9]所提出人臉偵測的三角型基底法，並應用於估算二維影像中，人臉在三維空間中的姿態角度。在三維影像空間中，存在著三個旋轉角度資訊，分別為 Roll 角、Yaw 角與 Pitch 角，它們各自依附於 Z、Y 與 X 基準軸上。我們利用三角形區域中雙眼以及嘴巴的特徵，應用三角幾何參數之統計分析，估算二維影像中人臉之三維 Roll-Pitch-Yaw 角，以表示該人臉於三維空間之姿態。

本論文所提出的姿態估測方法，除了二維人臉影像外，亦可應用於三維點雲資料上。我們將人臉三維點雲資料投影至二維空間平面上，經由所投影之二維影像，我們亦可直接應用所提之方法，估算人臉於三維空間之姿態角度。

從我們的姿態估測結果顯示，無論二維影像或是三維點雲資料，Yaw 角與 Pitch 角的估測精度可達 15 度為一個分類基準，而 Roll 角的估測可達 1 度角以內的精度。由實驗結果顯示，我們所提出的姿態估測可應用於二維影像與三維點雲資料中。

**關鍵詞：** 三角形基底；臉部姿態估測；Roll-Pitch-Yaw；點雲資料影像

# Abstract

We propose an orientation estimation method of a human face for 2D images and 3D point-clouds. This method is based upon the face detection method presented by Chiunhsiun Lin (2001). We use a statistical method of triangle-based parameters to analyze the face orientation and obtain three orientation angles: Roll, Pitch and Yaw. Applying the proposed method to 3D point-clouds, we first project the 3D point-clouds onto the 2D image plane, and then we directly estimate the angles as in 2D image cases.

From the experimental results, we can estimate the Yaw and Pitch angles by classifying each 15 degree as a unit from the range of 45~150 degrees, while the Roll angle within 1 degree. This method of orientation estimation could execute on no matter 2D images or 3D point-clouds.

**Key words:** Triangle-based; Facial orientation estimation; Roll-Pitch-Yaw ; Point-clouds images

## 誌謝

本論文能如期完成，主要感謝指導教授邱士軒博士，於學生碩士班這兩年求學過程期間的細心教誨，並提供完善的研究空間讓學生得以完成學業。在這段期間，由於教授的教導使得學生不僅在研究上能順利達成，並且在待人處事上有所成長，僅此由衷致上最誠摯的感謝。特別感謝中央警察大學溫哲彥教授不辭辛勞細心指導學生之研究，提供學生研究方向不遺餘力，學生無限感激。口試期間，特別感謝中央警察大學溫哲彥教授、及系上邱顯堂教授與蘇清淵教授於百忙之中撥冗前來指導學生對本論文提供建議，使學生受益良多，深表感激。



求學期間，感謝研究室朝夕相處的成員給予我幫助與鼓勵。感謝博士班學長李俊輝、陳清棋、陳鴻明、蔡承陳、林國宏、于順穎、張仕政、陳正欽、林信旭、陳坤廷，在職班學長廖成旺、吳清源、唐欣原、王世憲，在求學過程中一直給予協助與鼓勵，使我受益匪淺；畢業學長呂全斌、廖俊鑑、吳典錡、陳嵩岳、陳韋樑、李冠緯、蘇俊豪、詹志元、林唯修、詹振揚、毛漢文，時常回研究室分享工作經驗，深表感激；學長潘舒喬、同學許信紀、陳建廷、林義嵐、陳漢雄，在研究所兩年中，彼此互相砥礪學習與成長；學弟呂俊毅、李紹峰、蔡家榮、李忠謙、黃信傑、吳哲宏、學妹周美君，在研究及撰寫論文期間

給予幫助，在此一併致謝。

感謝我的父親黃添福先生、母親張月英女士，這二十多年來的養育栽培之恩，讓我在研究期間能無後顧之憂，並給予我良好的讀書環境，在我不如意時，給予最大的鼓勵與安慰。感謝姊姊黃莉雯的全力支持，與我共同分擔家裡大小事務，協助我順利完成學業，最後，將本論文之榮譽與喜悅獻給所有愛我及我愛的人。



# Contents

摘要.....	I
Abstract.....	II
誌謝.....	III
Contents .....	V
LIST OF FIGURES .....	VII
LIST OF TABLES .....	X
CHAPTER 1 INTRODUCTION .....	1
1.1 Foreword.....	1
1.2 Relevant research.....	2
1.3 Purpose of research.....	12
1.4 Framework of thesis .....	13
CHAPTER 2 EXPERIMENTAL EQUIPMENT .....	14
2.1 The device of generating point clouds and face images..	14
2.2 Projecting point-clouds images based on OpenGL .....	24
CHAPTER 3 METHOD DESCRIPTION .....	27
3.1 The Angle Estimation of 2D Face Photograph.....	27
3.1.1 The pre-process working .....	27
3.1.2 The triangle-based of face features.....	39



3.1.3	Orientation estimation of triangle-based .....	46
3.2	The Angle Estimation of 3D Point-clouds.....	64
CHAPTER 4 EXPERIMENTAL RESULTS .....		66
4.1	Results of angle estimation of face triangle .....	66
4.1.1	The result of estimation for Yaw angle.....	67
4.1.2	The result of estimation for Roll angle.....	68
4.1.3	The result of estimation for Pitch angle .....	69
4.2	Results of comparison estimating consequence with 2D image and projective image from point-clouds .....	70
CHAPTER 5 CONCLUSION AND PROSPECTS .....		72
REFERENCES .....		74

# LIST OF FIGURES

Figure 1.1	The definition of three face rotation angle: Roll, Pitch, and Yaw [10].....	3
Figure 2.1	LT 3D FaceCam (Model: M300S) .....	15
Figure 2.2	The whole experiment environment.....	16
Figure 2.3	The 13 different orientation of one person.....	17
Figure 2.4	The Beauty3D system index .....	19
Figure 2.5	The flowchart of point-clouds merging steps.....	20
Figure 2.6	The demonstration of reading .ltm files .....	21
Figure 2.7	Addition of manual landmarks and Setting as base model ..	21
Figure 2.8	Pick correspond landmarks to the base one and Registration by points .....	22
Figure 2.9	The merged point-clouds image.....	22
Figure 2.10	Register by area until the average error is between 0.2 ~ 0.3 mm.....	23
Figure 2.11	Output a whole complete point-clouds image.....	23
Figure 2.12	OpenGL program based on Borland C++ Builder 6.0 .....	25
Figure 2.13	The parts of images projecting from 3D point-clouds .....	26
Figure 3.1	Distribution of Gray level image.....	28

Figure 3.2	Distribution of Gray level with binary image .....	29
Figure 3.3	The foreground ( $C_1$ ) and background ( $C_2$ ) of a image.....	30
Figure 3.4	(a)image object (b) 4-neighbors dilation.....	33
Figure 3.5	(a)image object (b) 4-neighbors erosion .....	34
Figure 3.6	(a) Before Opening (b) After Opening .....	35
Figure 3.7	(a) Before Opening (b) After Opening .....	36
Figure 3.8	(a) $3 \times 3$ pixel mask(b) 4-neighbors (c) 8-neighbors.....	37
Figure 3.9	(a) Before labeling (b) After Labeling .....	38
Figure 3.10	The searching flowchart of the face triangle.....	40
Figure 3.11	(a) The triangle-based i j k (b) Three red points( i, j, and k).42	
Figure 3.12	The search area of the third block center k is only limited to the block, E, instead of the whole area of the image.[9] .....	45
Figure 3.13	The parameters of face triangle geometry.....	46
Figure 3.14	The graph of the tendency of changing face orientation....	50
Figure 3.15	The simplify chart of original data.....	51
Figure 3.16	The different zones of orientation estimation .....	52
Figure 3.17	The triangle-based for estimating Pitch angle.....	56
Figure 3.18	The graph of the tendency of changing Pitch angle.....	58
Figure 3.19	The simple chart of original data for Pitch angle estimation	

.....	59
Figure 3.20 The different zones for estimation of Pitch angle .....	62
Figure 3.21 The flowchart of estimating orientations for 3D point-clouds.....	64
Figure 4.1 The results of estimation for Yaw angle.....	67
Figure 4.2 The results of estimation for Roll angle .....	68
Figure 4.3 The results of estimation for Pitch angle.....	69



# LIST OF TABLES

Table 2.1 The hardware format of LT 3D FaceCam.....	15
Table 2.2 The scanner format of LT 3D FaceCam.....	15
Table 2.3 The angles of camera captured correspond to the real face orientation of Yaw angle .....	18
Table 3.1 The part of data record to estimate Yaw angle.....	49
Table 3.2 The standard deviation of original data .....	51
Table 3.3 The zones of different gaps.....	53
Table 3.4 The part of data record for estimating Pitch angle .....	57
Table 3.5 The standard deviation of data for Pitch angle estimation .....	59
Table 3.6 The part of data record for estimating Pitch angle .....	60
Table 3.7 The zones of different gaps for Pitch angle evaluation .....	63
Table 4.1 Results of comparison estimating consequence with 2D image and projective image from point-clouds.....	71

# CHAPTER 1 INTRODUCTION

## 1.1 Foreword

The application for Human Computer Interface (HCI), such as the information about face orientation including of both position and orientation, is a quite important issue. Images of face are covering the approximately features of human identification or estimation of face pose. Generally, human face is one of the most outstanding characteristics from person to person.

Pose estimation is a quite important issue before human identification or retrieval. Base upon the file formats contain 2D and 3D one. How to use traditional 2D image to estimate 3D pose is an interesting study in computer science.

In consequence, a better method of estimation for human face pose is necessary for human face identification or recognition.

## 1.2 Relevant research

Estimation of facial pose is one of the most active research areas in computer vision that have long been studied. This is mainly due to it is could be as pre-working state used as wide range of applications such as person identification [8][11][12][13][36], visual surveillance [13][36], Human Computer Interaction (HCI) [10][15][29][30], and image or video retrieval [22][31][40]. After three decades of research, the state-of-the-art approaches can achieve high recognition rate under controlled settings [23][37]. Before implementing the face recognition, the pre-working state is finding out the human face and the face pose. With the better face information, pose and gaze orientation, the performance of face identification could be presently improved.

In the world, the three axes of rotation were be defined as shown in Figure 1.1 by the authors, Qiang Ji and Rong Hu, in 2002. Latter, there are more and more researches in the study area all following the definition. The three angles are increasing with the direction of counterclockwise by right hand.

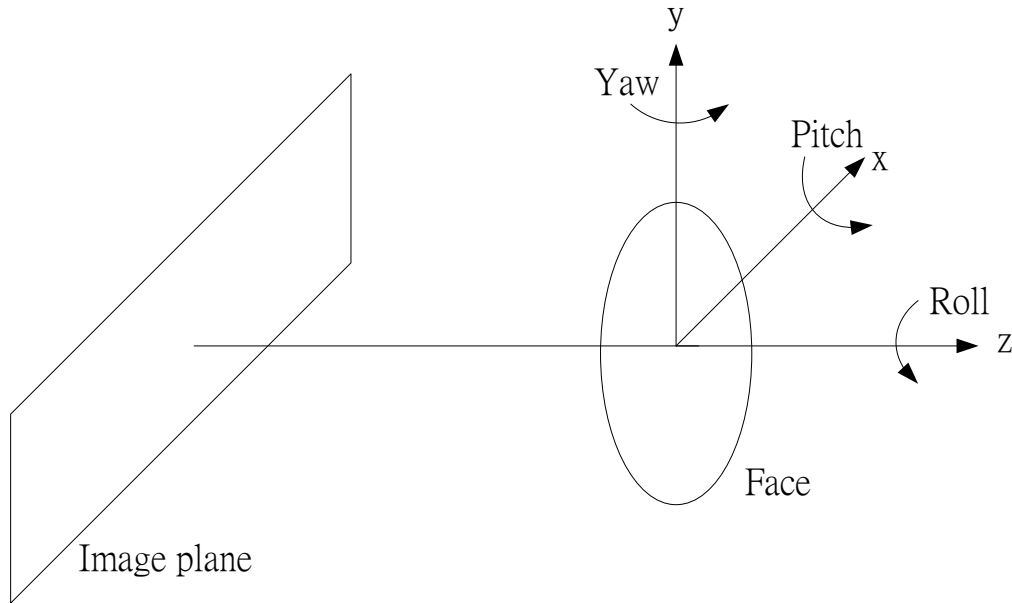


Figure 1.1 The definition of three face rotation angle: Roll, Pitch, and Yaw [10]

In the Ref. [16], Zhao W. and co-authors presented a configuration of a generic face recognition system. Face recognition system is must containing three parts, face detection, feature extraction, and face recognition. At the part of face detection, we should detect where the face region contains in an image and then estimate the head pose. There are many researches in face region finder. However, the most interesting study to us is the estimation of face orientation. Because the orientation of images captured would be occurred rotation of Yaw angle, tilt of Roll angle, or bending or lifting of Pitch angle. Without calibrating the face orientation to initially set, the performance of feature extraction or face



recognition would be failed. Therefore, we could know the importance of estimation for face orientation. Some applications of pose estimation for human identification or retrieval show as bellow.

In the Ref. [8][11][12][13], Yoshino, Masttsuda, and some researchers already presented some method in the area of human face identification. In the Ref. [8], researchers proposed a system for facial image identification. There are a three dimension (3D) physiognomic range finder and a computer-assisted facial image superimposition unit in the system. Two sinusoidal grating projection devices and two CCD cameras are used as a detector. A host computer are including of proprietary software, a flat surface color display, and a color image scanner for inputting two dimension (2D) facial images of a query person. They are used as the superimposition unit of computer-assisted facial image. Use of the 3D facial shape and texture of a person are obtained by the range finder. Use the reciprocal point-to-point difference of anthropometrical points between 3D and 2D facial images to make the comparison between them in, that way, we could know the person is query one or not. Finally, we could see in the process of comparison with reciprocal point-to-point must use the information of face pose.

In the Ref. [11], the scholars proposed a computer-assisted facial image identification system with the reliability of a method of morphometrical matching for identifying disguised faces. The 2D right oblique facial images consist of three target persons disguised with sunglass, cap and gauze mask. By calculating the average perpendicular distance between the facial outlines and the average point-to-point distance of the corresponding landmarks in the 2D image of the disguised face and 3D facial image, hence, they can know the difference of pose and surface contour.

In the 2002, Yoshino presented a review of facial image identification [12]. This paper makes a reviewed about recent advances in facial image identification from the forensic aspects. The literatures describe that the paper included mathematics and computer engineering fields, anthropological techniques. There are three methods about morphological comparison of facial features, facial image anthropometry and face-to-face superimposition generally attempted in the facial image identification. The morphological comparison of facial images is carried out based on the morphological classification of facial components such as facial types, eyebrows, eyes, nose, lips and ears. For implementing of

correct face identification, some problems must be resolved. The problems are like that failure to consider some factors such as the orientation of subject's head, the distance between the subject of the camera system, particularly when attempting video superimposition or biometric analysis, could lead to incorrect conclusions. Therefore, there are more and more 3D physiognomic analysis was developed.

Going though many researchers have been worked hard on the area of face identification, besides that, another concept of 3D facial retrieval system also was presented. In the Ref. [13], Yoshino, Imaizumi, Tanijiri, and so on proposed an automatic adjustment of facial orientation in 3D face image database. The method of morphometric matching for identifying persons from surveillance images which using a computer-assisted facial image identification system has now been introduced to actual casework. In the reference, the authors use about 5000 superimpositions of the 2D face images from 50 subjects which were compared automatically with the 3D face images of 100 people. Though the result of this experiment, it is concluded that the software developed will be applicable to the automatic adjustment of facial orientation in the 3D face image database and this could be anticipated to

improve the efficiency of policing in Japan as 3D imaging systems.

In the Ref. [32][33][41], they found out a face retrieval system is quite important for some applications, which like as security monitoring system, human identification system, life safety guard and so on. The human face identification system must check back the database which contains the relational and similar information that belongs to the right people. But, the problem of face pose is containing in the surveillance system. In the Ref. [32], the authors proposed a face retrieval system with 3D face pose estimation. It relies on a large database of registered face images of different people viewed from several perspectives. They use a powerful image retrieval technique to match the input image against database ones, which returns the most similar pose of 3D's. This 3D pose can be refined using matches between input image and database images. From the experiment result given that these labels are not particularly accurate themselves and that the  $15^\circ$  angular error is barely noticeable by a human observer. Then the error modification of the paper would still have some part of issues that could be improved.

Visual surveillance is one part of the wide range application of face recognition. Forensic anthropologists or other “experts” are asked to

express a judgment about the possibility that a person visible in an image can be identified as a known person or not. But in the real situation of the world, the video-surveillance image capture device almost could only get low resolution frames. In order to solve the problem, some scholars proposed a new method apply on personal identification of visual surveillance. A source of face retrieval system is using video rather than using single frame of still images. In the Ref. [41], Shan announced that video data consisted of rich and redundant information, which could be exploited to resolve the inherent ambiguities of image-based recognition. They are like sensitivity to low pose variations, occlusion, and resolution leading to more accurate and robust recognition.

In the Ref. [36], Angelis, Sala, and Cantatore, they proposed a procedure aimed at identification from two-dimensional (2D) images which were contained as video-surveillance tapes by comparison with a three-dimensional (3D) facial model of a query person. The authors apply the concept of “geometrically compatible images” [36]. The idea is using a scanner to reconstruct a 3D facial model of a query person and comparing it with a frame extracted from the video-surveillance sequence which indicates the face of the query one. Therefore, we could see the

human pose in image could be resolved when doing human identification.

Whatever human-face recognition, personal retrieval or visual surveillance, they all have to apply on human identification. However, the procedure of comparison the reciprocal point-to-point distance of the facial images or models, we could find out the face re-orientation was a tough and task problem. The researches of estimation for face pose are such as in Ref. [9][15][25][31][32][34][41]. In the Ref. [9][15], the authors proposed a system for pose estimation consists of two main parts. The first part is to search for the potential face regions that are gotten from the isosceles-triangle criteria, which was based on the rules of “the combination of right eye, left eye, and one mouth”. The second part of the proposed system is the performing the task of pose verification by utilizing face weighting mask function, direction weighting mask function, and pose weighting mask function. Selection of any ten faces which was already been normalized to  $60 \times 60$  pixels, feeding them to train out a face mask that comparing the face mask to other input image. Because the weight values were different, the authors can distinct the different face poses from each input image.

In the Ref. [31], the authors presented an estimation 3D facial pose

in video with just three points. The method that scholars proposed is able to work with only the location of three points –two eyes and one mouth of faces– in each frame. With positions of the three points, the system could estimate out the roll angle and make face tracking. Utilizing the simplified anatomical model which the authors gave, the pitch and yaw angles would be estimated.

Another method of head pose estimation was presented as in Ref. [34]. They use the information of 2D head detection applying ellipse searching and skin color matching to project onto 3D face model with low resolution. The pan angle would hence be estimated by using particle filter.

After reviewing relative researches, we could find that face pose estimation is quite important in security monitoring, automated surveillance system, or forensic evidence. Face recognition with pose is a natural human ability and a widely accepted identification and authentication method. However, one of the main challenges still remaining is the non-rigid structure of the face, which can change permanently over varying time-scales and briefly with facial expressions. In the overview paper [40], the authors discussed the three parts of 3D

face recognition: evolution, approaches, and challenges. As an evolution which the authors said, face recognition methodology has shifted from a pure 2D image-based approach to the use of 3D facial shape. The majority of face recognition methods has been developed by scientists with a very technical background such as biometry, pattern recognition and computer vision. However, numerous studies show that people are not always reliable sources when comparing faces with recollections, and are significantly affected by differences in lighting, familiarity, expression and viewpoint or especially pose. With this in mind, it becomes an almost insurmountable challenge to confirm objectively an identification using human perception alone. In order to achieve a high performance of face pose estimation, we must be thinking out new methods or overcome the sources effect.



### 1.3 Purpose of research

Nowadays, there are many and many devices of capturing 3D data opening up and developing out though the change of times. How to connect the 2D information with the 3D data is an interesting issue for us.

In order to know the relationship of 2D images and 3D ones, we particularly want to understand what kind of information is necessary between 2D and 3D data. For examples, we try to use the 2D face image to find out the information of three axes. The goal of our research is seeking out and estimating the face orientation from 2D images. The orientation estimation contains three rotation axes; one is Pitch angle with X axis, the other is Roll angle with Z axis, and another is Yaw angle with Y axis. With the orientation information of input image for application, we may figure out the original captured situations of camera, being a consult event of face recognition, or forming parameters of similar comparison with people.

## 1.4 Framework of thesis

The thesis consists of five chapters.

The first chapter introduces the present status of the subject and the purpose of the research.

The second chapter introduces experimental equipment and computer-assisted software.

The third chapter specifies the applying method of the orientation estimation of 2D photograph.

In fourth chapter, experimental results and discussions are demonstrated.

In final chapter, we make the conclusion and bring up some prospects.



## **CHAPTER 2 EXPERIMENTAL EQUIPMENT**

### **2.1 The device of generating point clouds and face images**

The capture equipment of the research includes images capture device about point-clouds and face images which is belong to the department of Forensic Science of Central Police University. Calculation of the depth information by using a computer and the difference of projecting light mask when the camera captures images which is fixed on the capture equipment. The captures device is using the light mask and algorithm of time of flight to produce the 3D point-clouds of the face and get the 2D face images. The images and a capturing device of point-clouds information which is called as LT 3D FaceCam (Model: M300S). The machine is illustrated in Figure 2.1. The hardware spec of LT 3D FaceCam is shown as Table 2.1 and the scanner format of LT 3D FaceCam is written as Table 2.2.

Table 2.1 The hardware format of LT 3D FaceCam

Items	Specification
Size	420mm(W)*250mm(H)*150mm(D)
Weight	6.5kg
Power supply	AC110V~240V, 200W
Image capture device	8M Pixels
Speed of capture a image	<0.5 sec.
Light projection device	Structure light of single one color

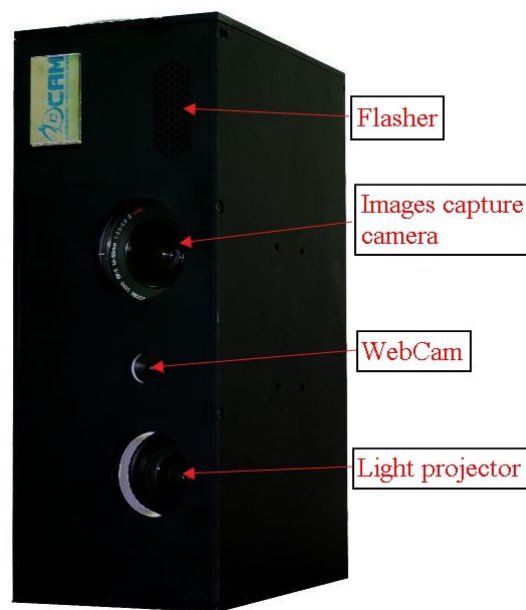


Figure 2.1 LT 3D FaceCam (Model: M300S)

Table 2.2 The scanner format of LT 3D FaceCam

Items	Specification
Measurement of range	300mm(W)*220mm(H)*180mm(D)
Resolution	<1.0 mm
Accuracy	+/- 0.25mm
Texture information	Yes (with> 8M UV Texture Map)
Output of 3D format	LTM, OBJ, STL, WRL
Measurement of range	300mm(W)*220mm(H)*180mm(D)

The experiment environment is indoor environment at the laboratory of the department of Forensic Science of Central Police University. The whole experiment environment is illustrated in Figure 2.2. With this experiment environment, we take pictures about each 15 degrees of Yaw angle from the person right side to left side. With the experiment state, we must take 13 different images and point-clouds information. Every image which size is  $1152 \times 1728$  pixels were shown as Figure 2.3.



Figure 2.2 The whole experiment environment

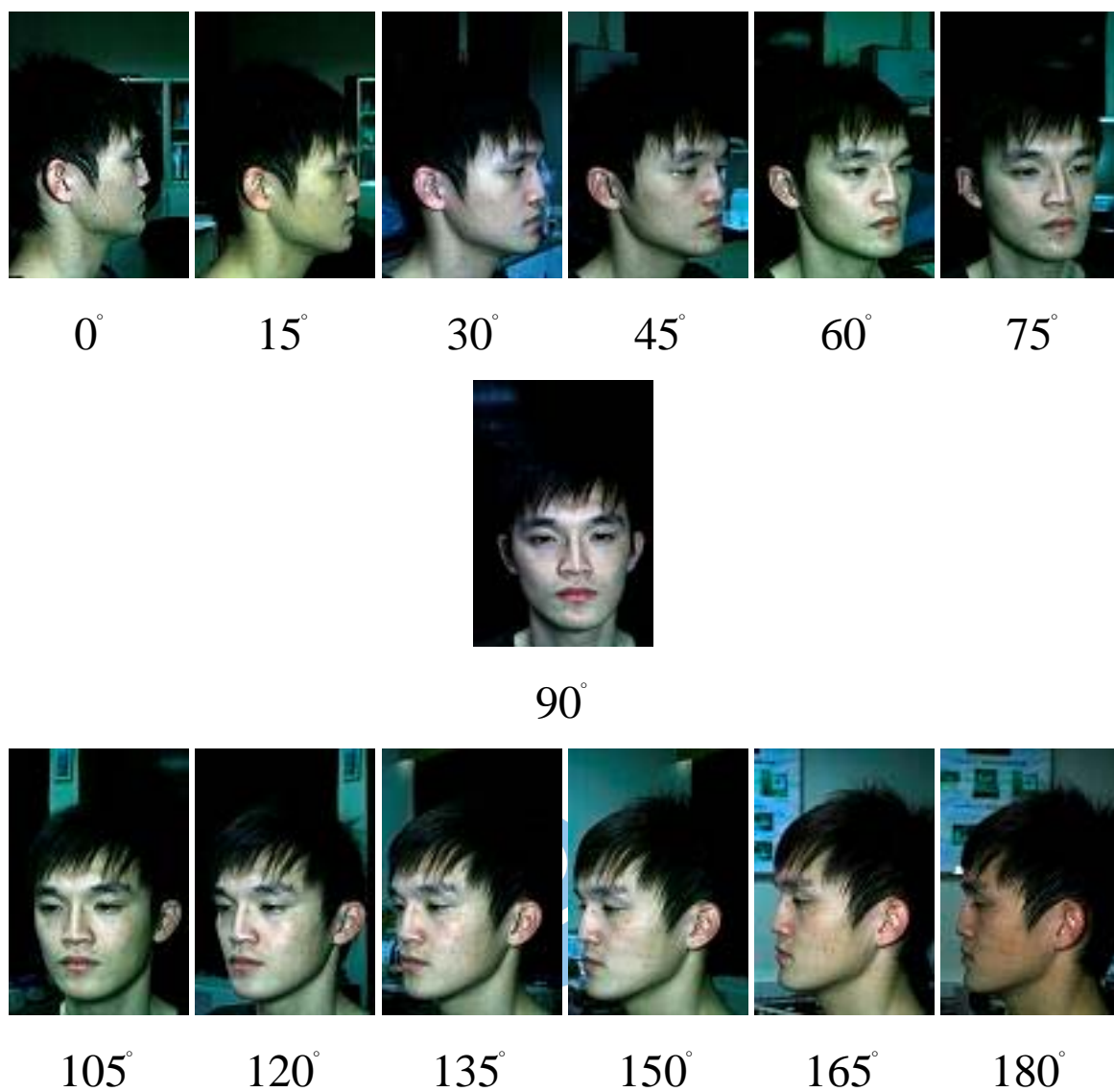


Figure 2.3 The 13 different orientation of one person

Figure 2.3 shows the angles of the camera orientation when capturing to the ground. However, the real orientation of Yaw angle would be seen as the Table 2.3.

Table 2.3 The angles of camera captured correspond to the real face orientation of Yaw angle

Degrees of camera angle (°)	Head_ Orientation ; Degrees of real Yaw angle of face orientation (°)
0	Turn Right , 90
15	Turn Right , 75
30	Turn Right , 60
45	Turn Right , 45
60	Turn Right , 30
75	Turn Right , 15
90	Look Front , 0
105	Turn Left , 15
120	Turn Left , 30
135	Turn Left , 45
150	Turn Left , 60
165	Turn Left , 75
180	Turn Left , 90

After getting 13 images and data of cloud points each person, we use the program which is called as “Beauty3D” that is supported by the company named Logistic Technology Corporation to merge the face frames as reconstruction processing. The index of Beauty3D system is illustrated in Figure 2.4.

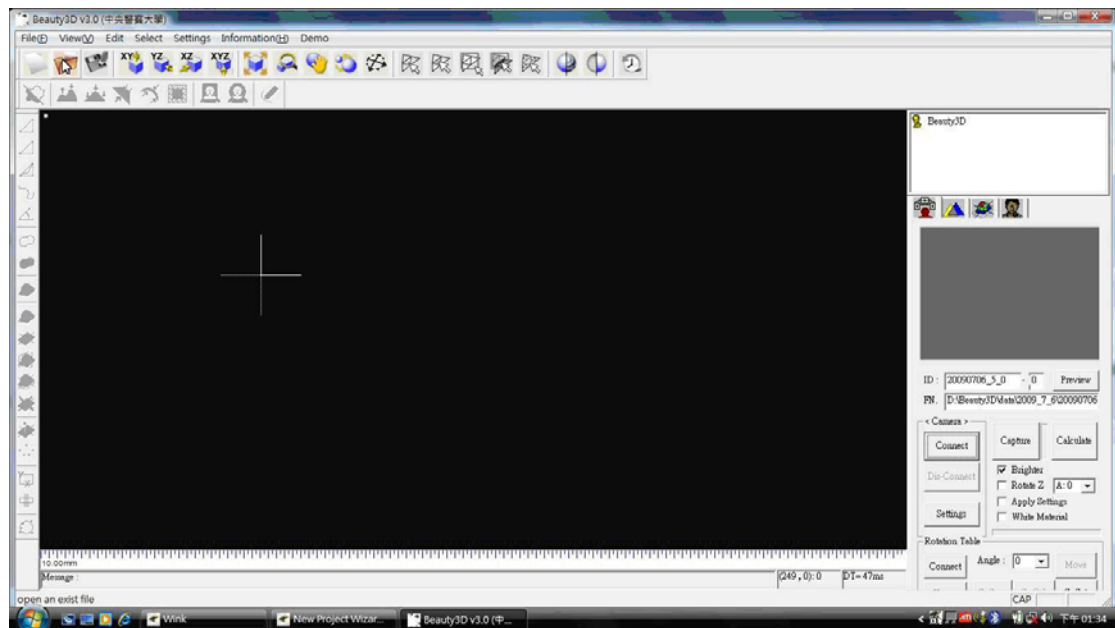


Figure 2.4 The Beauty3D system index

The program which is called as “Beauty3D” makes users more convenient to use an assist tool to merge the 13 different parts of face point-clouds as a point-clouds model of complete personal face. As in Figure 2.5, we could see the flowchart of point-clouds merge steps are following bellow.



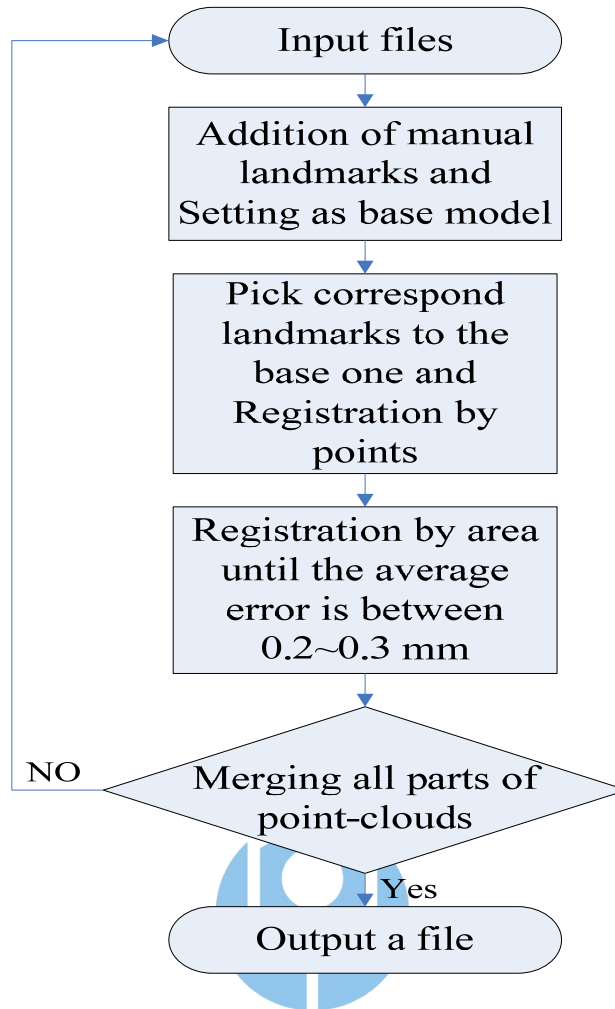


Figure 2.5 The flowchart of point-clouds merging steps

The flowchart of point-clouds merging steps contains five steps. We would show the detail parts and illustrations as following bellow.

Firstly, we would read all “.lrm” files that totally 13 ones of a person. The demonstration is illustrated as Figure 2.6. Beauty3D is mainly a computer-assist tool that helps us to merge all 13 parts of point-clouds and to do distance measurement or to repair the broken parts of point-clouds information.

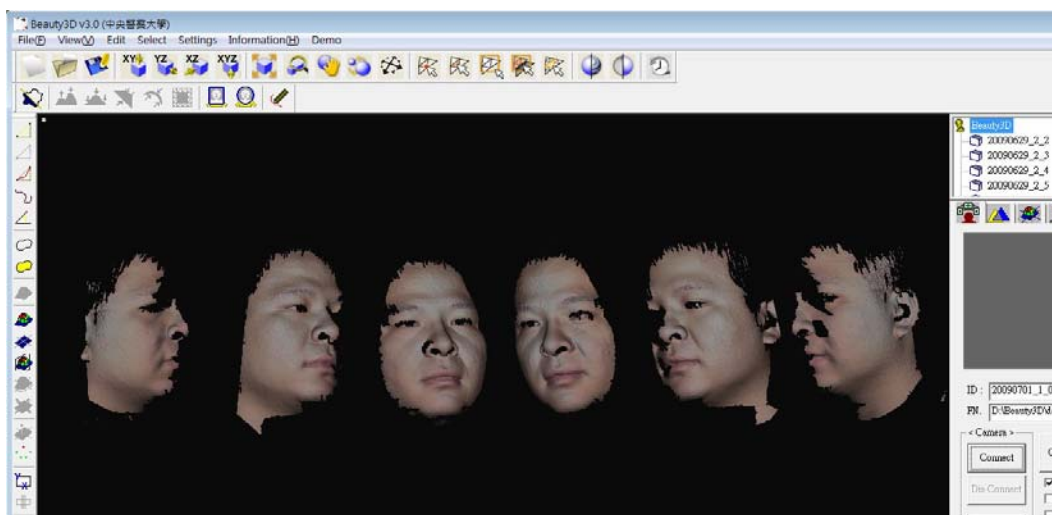


Figure 2.6 The demonstration of reading .ltm files

Secondly, select three or more landmarks by manually in one part of point-clouds image and set certain point-clouds image as a base. The process is shown as Figure 2.7. Red numbers is the landmarks that selecting by manually.

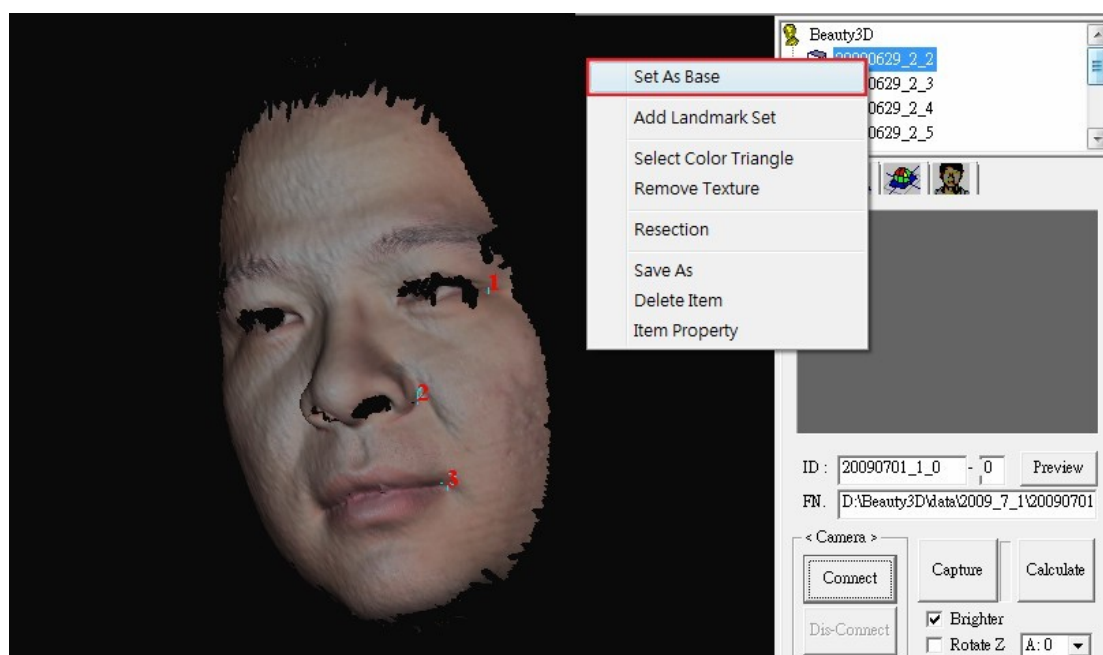


Figure 2.7 Addition of manual landmarks and Setting as base model

Thirdly, pick in sequence correspond landmarks to the base one and register by points, thereby we could get a merged point-clouds image.

The results were demonstrated as Figure 2.8 and Figure 2.9.

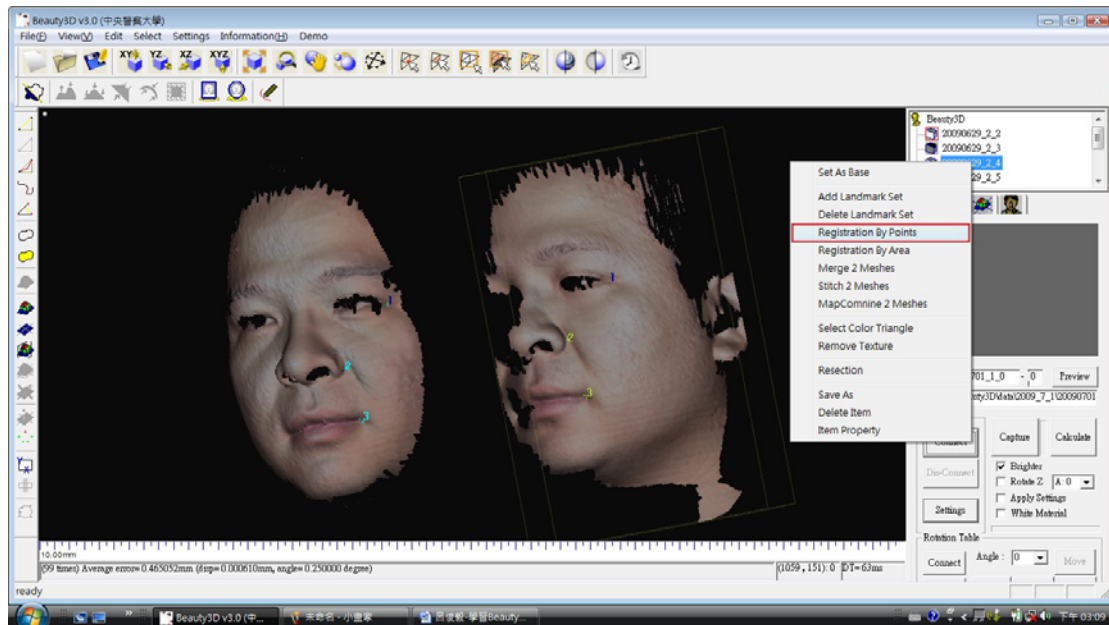


Figure 2.8 Pick correspond landmarks to the base one and  
Registration by points

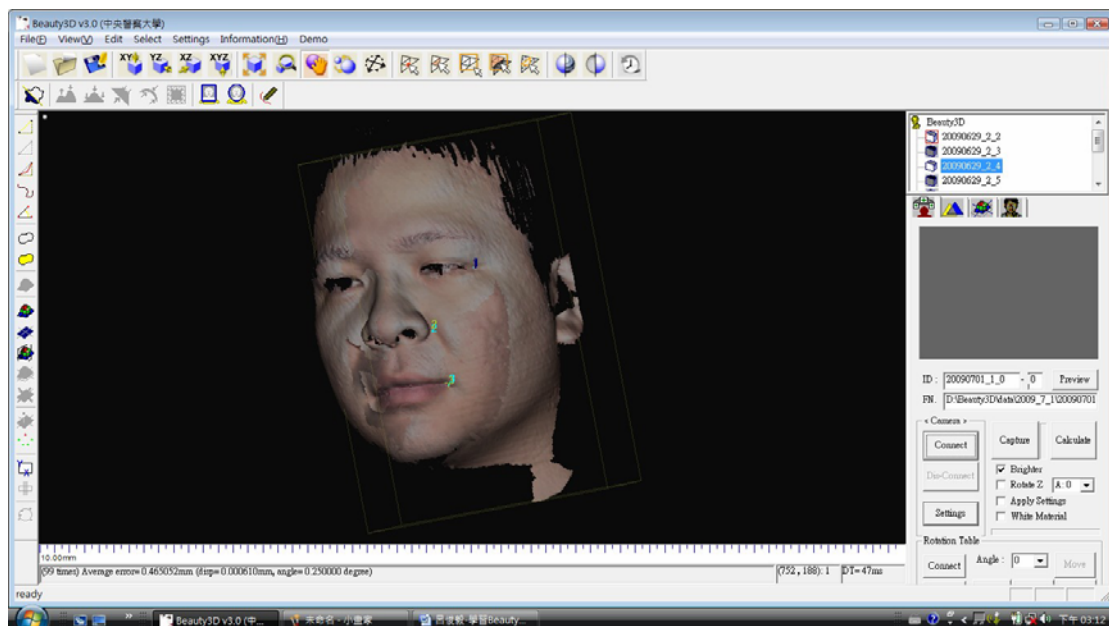


Figure 2.9 The merged point-clouds image

Fourthly, registration by area until the average error is setting within as an ideal range. The ideal range (witch we marked red a rectangle) we required is between 0.2 ~ 0.3 mm.

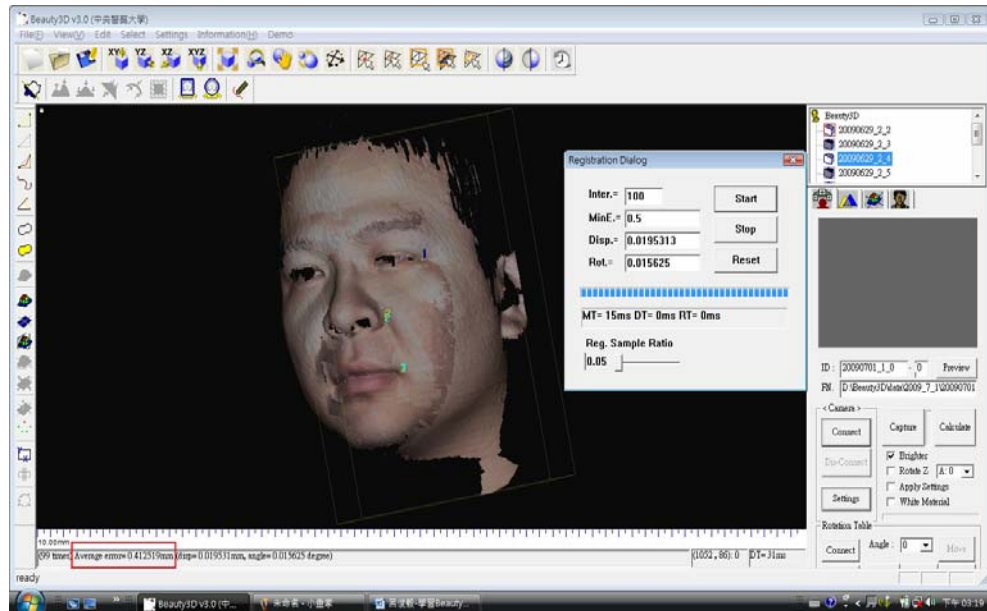


Figure 2.10 Register by area until the average error is between 0.2 ~ 0.3 mm

Finally, after merging all thirteen parts images of the point-clouds of one person, that we could get one whole complete facial model (like as Figure 2.11 ) of point-clouds data of the input images.



Figure 2.11 Output a whole complete point-clouds image

## 2.2 Projecting point-clouds images based on OpenGL

After merging the all point-clouds images by the procedure as section 2.1, we would project the whole complete 3D point-clouds onto the 2D image plane. The computer-assist tool is “OpenGL”. The benefits of OpenGL could be listed as bellow:

- The structure of OpenGL is quite easy and simple that makes users easily to learn it.
- OpenGL can execute not only on Windows but on Linux, FreeBSD, OS/2, Macintosh OS, and BeOS.
- The functions of OpenGL database are more than 100 kinds of libraries.

With the benefits of OpenGL, we implement it on Borland C++ Builder 6.0. Our goal is to develop a program that could change the 3D object size, rotation, shift based on OpenGL. The functions of OpenGL database make us execute the commands that we needed. Figure 2.12 shows that the program OpenGL based on Borland C++ Builder 6.0.

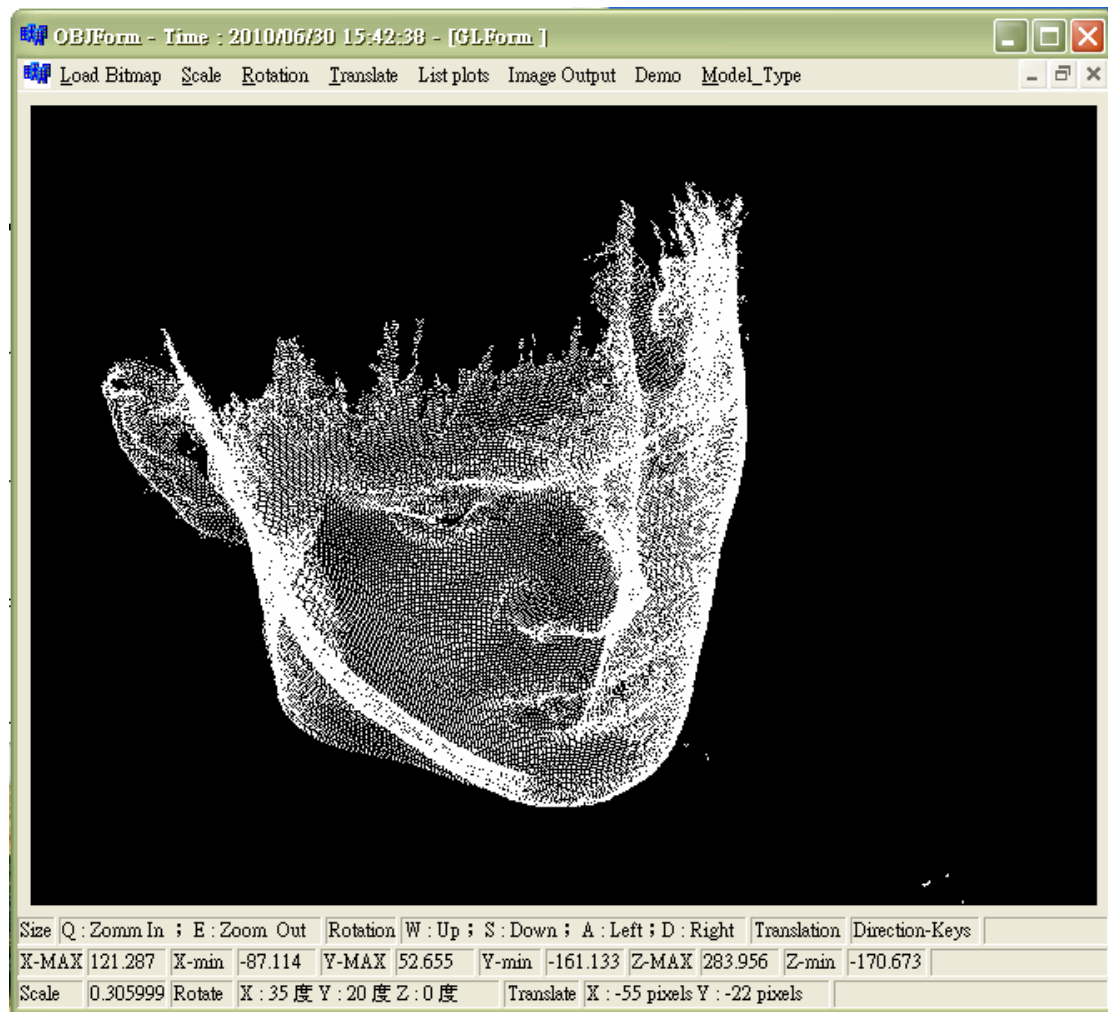


Figure 2.12 OpenGL program based on Borland C++ Builder 6.0

The main application of OpenGL is help us to simulate the as possible as similar situations to the input face photograph. The objects could zoom in/out, rotate with any angle of each axis, and move with different direction that plays an important roll to imitate different possible face orientation of input images. Without knowing what conditions of images would be input our system, we will use the 3D point-clouds data



to project orthogonally onto 2D image plane to imitate the input face orientation as possible. We take the orthogonal projecting images with three conditions of facial orientation: one is the Yaw angle from  $-55^{\circ} \sim 55^{\circ}$  and set  $5^{\circ}$  as a unit; the other one is the Roll angle from  $-15^{\circ} \sim 15^{\circ}$  and set  $1^{\circ}$  as a unit; another one is the Pitch angle from  $-55^{\circ} \sim 55^{\circ}$  and set  $5^{\circ}$  as a unit. There are totally  $23 \times 31 \times 23 = 16,399$  images per one person. The image size is  $640 \times 480$  pixels. Figure 2.13 illustrates part of orthogonal projecting images which belong to one person. The total images of 42 people, the images database would be enlarge as 688,758 images.

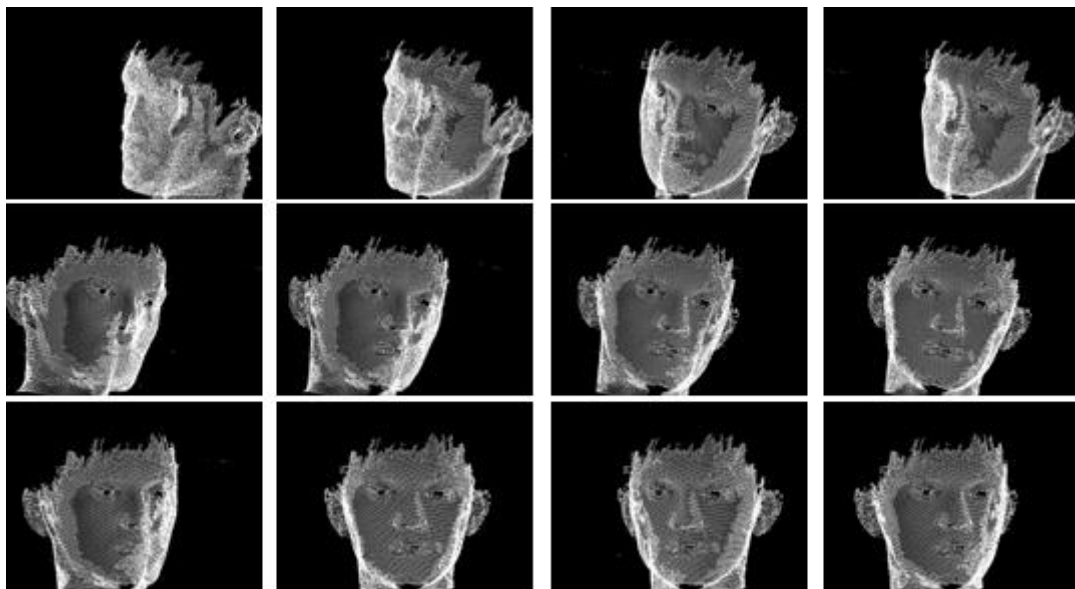


Figure 2.13 The parts of images projecting from 3D point-clouds

## CHAPTER 3 METHOD DESCRIPTION

### 3.1 The Angle Estimation of 2D Face Photograph

In order to reduce the searching times of face retrieval, we firstly estimate the face angle with 2D face image. At the beginning, finding out the triangle of face features is more important. The method of face triangle [9] determination will be showed as following bellow.

#### 3.1.1 The pre-process working

##### 3.1.1.1 The theory of binarlization : Otsu

With digital image processing, there is a lot of color information in color images. In order to decrease the information quantity of color image, we transform the color image to binary images firstly. We call the processing as Binarization.

Binarization is separating gray level of image to two parts. After that, the image will contain only two color levels which are black and white.

The definition of Binarization is following as Equation (3. 1).

$$f(x, y) = \begin{cases} 255 & i(x, y) > T \\ 0 & i(x, y) \leq T \end{cases} \quad (3. 1)$$

$i(x, y)$  : It is the gray level on  $(x, y)$  of input image.

$f(x, y)$  : It is the gray level on  $(x, y)$  of output image.

$T$  : Threshold value



Determine a threshold value to transform the gray image or color image to binary image, which is only two parts of black (0) or white (255). If the gray level of pixel is smaller than the threshold, we set the pixel to black (0). In other words, if the gray level of pixel is larger than the threshold, we set the pixel to white (255). The result is shown as Figure 3.1 and Figure 3.2, the foreground and background can efficiently be segmented.

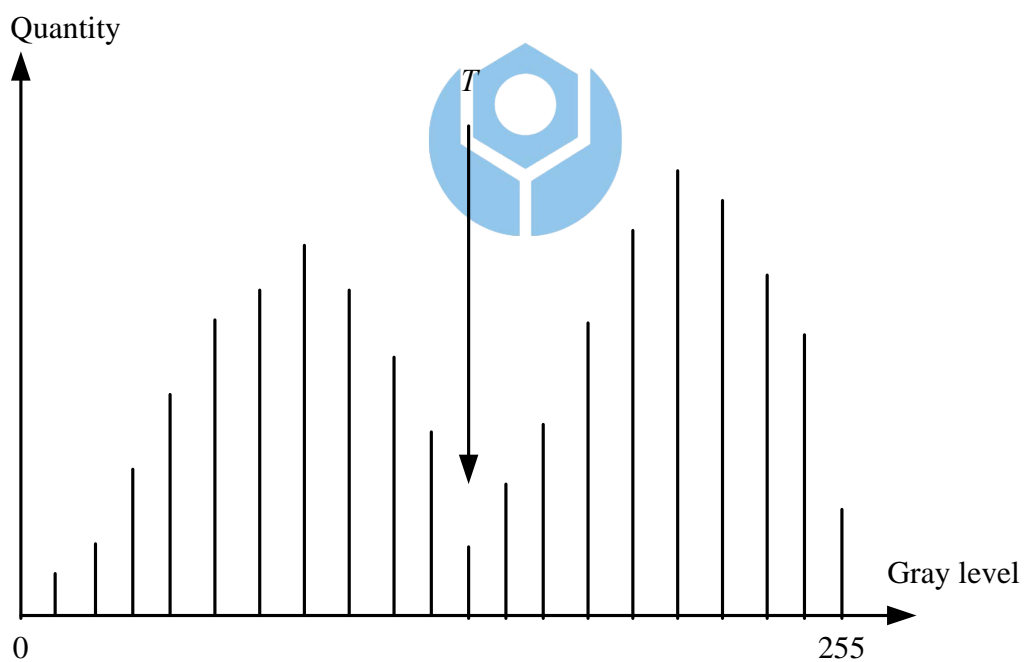


Figure 3.1 Distribution of Gray level image



Figure 3.2 Distribution of Gray level with binary image

There are many methods of transforming gray image to binary image, but in our purpose, we want to find out a suitable threshold value. The determined method of threshold value will affect the next step of image processing. Therefore, we could know that the threshold value determination is quite important to us.

The method of statistical threshold value determination [6] is according by Otsu [1] which presented in 1978. The purpose of the algorithm is finding out an ideal threshold value  $T$  which would make the sum of Within-variance be the minimum. Using the threshold value  $T$  could separate the foreground of image ( $C_1$ ) from the background of

image ( $C_2$ ) as Figure 3.3 showing bellow. In the foreground area ( $C_1$ ), the gray level  $f(x,y)$  must satisfy the range  $0 \leq f(x,y) \leq T$ , and in the background area ( $C_2$ ), the gray level  $f(x,y)$  must satisfy the range  $T+1 \leq f(x,y) \leq 255$ .

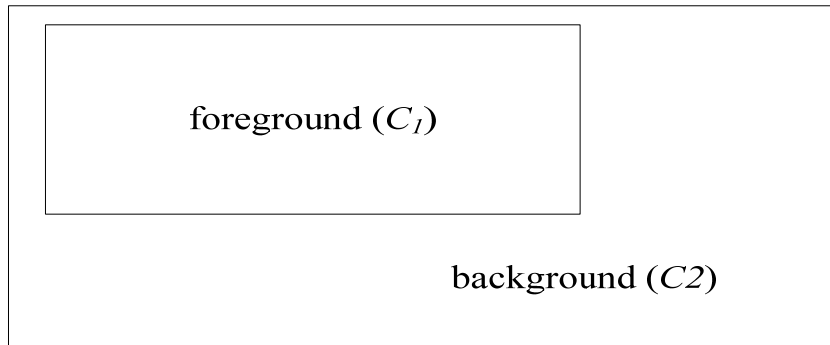


Figure 3.3 The foreground ( $C_1$ ) and background ( $C_2$ ) of a image



The scholar, Otsu, also proposed two conditions. One is the determination of threshold value  $T$  must satisfy the Between-variance of the foreground of image ( $C_1$ ) and the background of image ( $C_2$ ) is the maximum. Another one condition is the determination of threshold value  $T$  must satisfy the sum of the Within-variance in the foreground of image ( $C_1$ ) and the Within-variance in the background of image ( $C_2$ ) is the minimum. One of the two conditions is set up, the threshold value  $T$  must can be determined.

Assumption of the input image size is  $N$ , the numbers  $I$  of gray level value are 256, and  $n_i$  means the appearance time of the gray level  $i$ . The probability of gray level  $i$  could be expressed following bellow:

$$P(i) = \frac{n_i}{N} \quad (3.2)$$

$$0 \leq i \leq I-1 \quad (3.3)$$

The probability of numbers of gray level in the foreground ( $C_1$ ):

$$W_1 = \sum_{i=0}^T P(i) \quad (3.4)$$

The probability of numbers of gray level in the background ( $C_2$ ):

$$W_2 = \sum_{i=T+1}^{I-1} P(i) \quad (3.5)$$

Then, we could calculate out the desired value of the foreground of image ( $C_1$ ) and the background of image ( $C_2$ ):

$$u_1 = \sum_{i=0}^T \frac{P(i)}{W_1} \times i \quad (3.6)$$

$$u_2 = \sum_{i=T+1}^{I-1} \frac{P(i)}{W_2} \times i \quad (3.7)$$

After using  $u_1$  and  $u_2$ , we could calculate out the Within-variance between the foreground of image ( $C_1$ ) and the background of image ( $C_2$ ):

$$\sigma_1 = \sum_{i=0}^T (i - u_1)^2 \times \frac{P(i)}{W_1} \quad (3.8)$$

$$\sigma_2 = \sum_{i=T+1}^{I-1} (i - u_2)^2 \times \frac{P(i)}{W_2} \quad (3.9)$$

The sum of Within-variance in the foreground of image ( $C_1$ ) and the sum of Within-variance in background of image ( $C_2$ ):

$$\sigma_w^2 = W_1 \sigma_1^2 + W_2 \sigma_2^2 \quad (3.10)$$

The sum of Between-variance between the foreground of image ( $C_1$ ) and the background of image ( $C_2$ ):

$$\sigma_B^2 = W_1 (u_1 - u_T)^2 + W_2 (u_2 - u_T)^2 \quad (3.11)$$

$u_T$  denotes the average gray level of the whole original image as the following equation bellowed:

$$u_T = \frac{1}{N} \sum_{i=0}^{T-1} n_i \times i \quad (3.12)$$

Finally, we could determine a threshold value  $T$  that makes the sum of the Within-variance in the foreground of image ( $C_1$ ) and the Within-variance in background of image ( $C_2$ ) is the minimum, and the Between-variance of the foreground of image ( $C_1$ ) and the background of image ( $C_2$ ) is the maximum. Therefore, the threshold value  $T$  is the most suitable one for the input gray-level image.

### 3.1.1.2 Morphology

After transforming the original image to binary image, we could figure out that there would be many noises and holes in the image. In order to reduce the noises and fill the holes, we use the method of Morphology [2] to solving the problems.

#### **Dilation:**

With  $A$  and  $B$  as sets in  $Z^2$ , the dilation of  $A$  by  $B$ , denoted  $A \oplus B$ , is defined as:

$$A \oplus B = \left\{ z \mid \left( \hat{B} \right)_z \cap A \neq \emptyset \right\} \quad (3.13)$$

where,  $B$  is a structuring element and  $A$  is the set (image objects) to be dilated. Figure 3.4 illustrated of 4-neighbors and 8-neighbors dilation (The dilation result are the gray blocks).

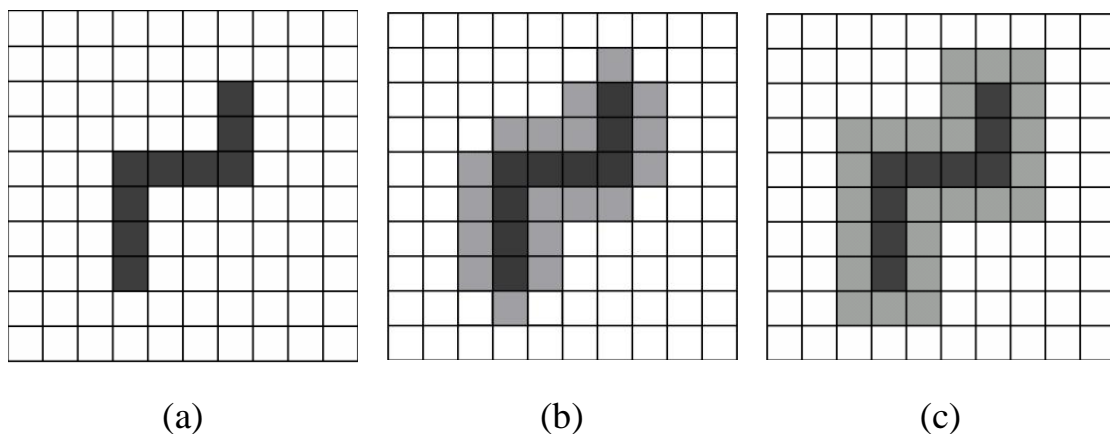


Figure 3.4 (a)image object (b) 4-neighbors dilation  
(c) 8-neighbors dilation

### Erosion:

With  $A$  and  $B$  as sets in  $Z^2$ , the erosion of  $A$  by  $B$ , denoted  $A \ominus B$ , is defined as:

$$A \ominus B = \left\{ z \mid \left( B \right)_z \cap A^c = \emptyset \right\} \quad (3.14)$$

where,  $A^c$  is the complement of  $A$  and  $\emptyset$  is the empty set.

Figure 3.5 illustrated of 4-neighbors and 8-neighbors erosion (The erosion result are the darker gray blocks).

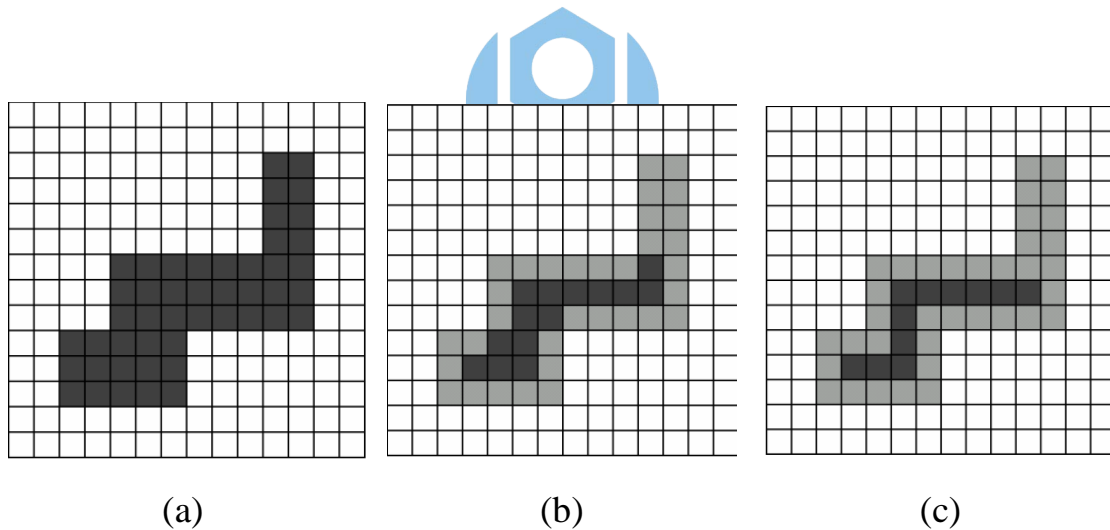


Figure 3.5 (a)image object (b) 4-neighbors erosion  
(c) 8-neighbors erosion

### Opening:

The Morphology of Opening method is combining the Erosion and dilation method which is described above. Opening is first using Erosion once and then do Dilation one time. The benefits of Opening method are reducing noises, smoothing the contour, removing the sharp parts of the shape, and breaking the narrow connected parts. The result of Opening method is illustrated as Figure 3.6.

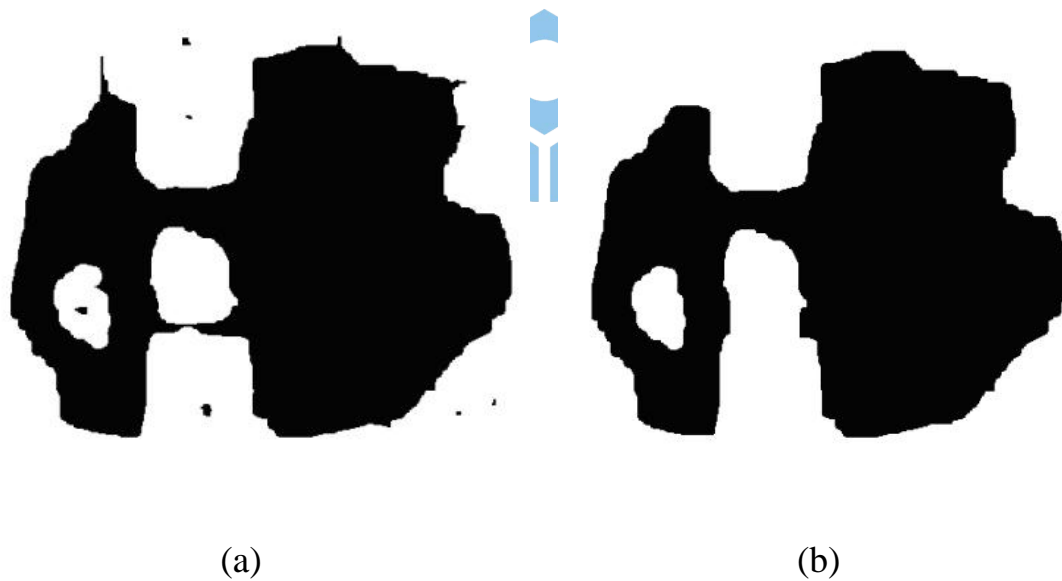


Figure 3.6 (a) Before Opening (b) After Opening



### **Closing:**

The Morphology of Closing method is combining the dilation and Erosion method which is described above. Closing is first using Dilation once and then doing Erosion one time. The benefits of Closing method are filling holes, smoothing the contour, and connecting the narrow break parts. The result of Closing is illustrated as Figure 3.7.

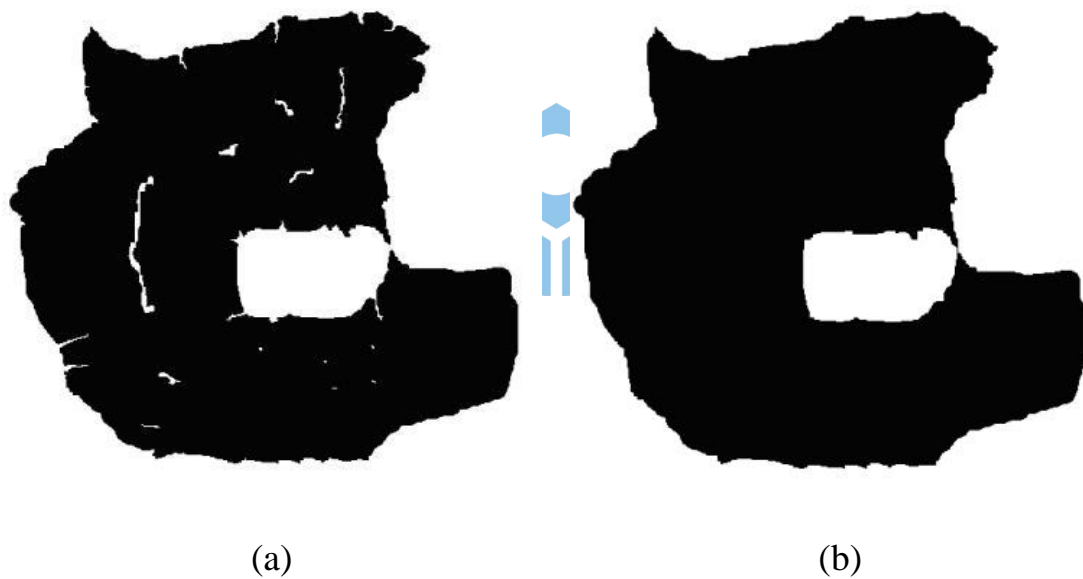


Figure 3.7 (a) Before Opening (b) After Opening

### 3.1.1.3 Labeling

Before introducing the method of labeling [1], we will discuss the difference of 4-neighbors and 8-neighbors. For example, the pixel at center location 5 in the  $3 \times 3$  mask (show as Figure 3.8 (a)) has 4-neighbors location 2, 4, 6, and 8. The 4-neighbors are illustrated as Figure 3.8 (b). The 8-neighbors of the pixel include the 4-neighbors plus location 1, 3, 7, and 9. The 8-neighbors are illustrated as Figure 3.8 (c).

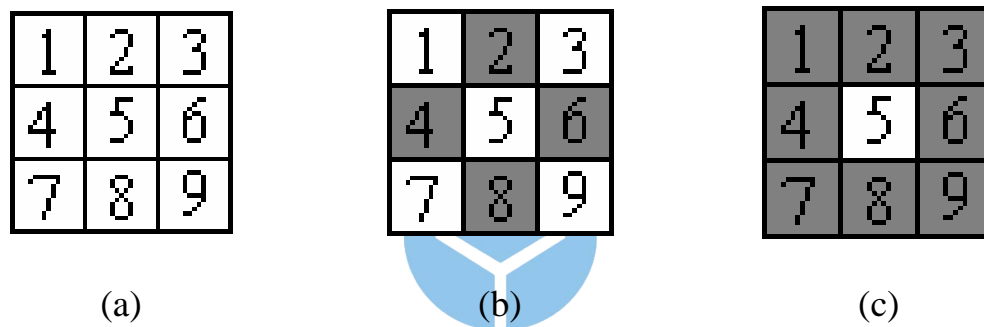


Figure 3.8 (a)  $3 \times 3$  pixel mask (b) 4-neighbors (c) 8-neighbors

Labeling is one of the almost common methods of digital image processing, and it is mainly applying at the binary image. The algorithm is setting the same gray level of pixels which are appeared at the neighbors of the same location then marking the pixels as the same block. For instance, using the black pixels of 4-neighbors as illustrated at Figure 3.9 (b).

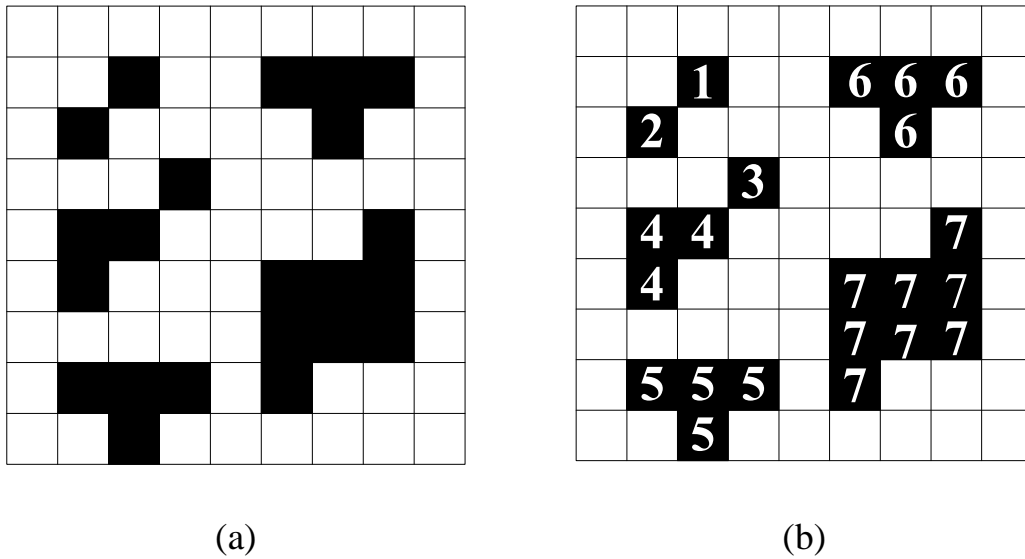


Figure 3.9 (a) Before labeling (b) After Labeling

By using the method of Labeling, we could compute out the different information of each block in our input image. The information of blocks is including the each coordinate of left-top corner, right-down, block-centers, and block-area. Hence, we could use the information to apply on next section as following bellow.

### 3.1.2 The triangle-based of face features

The most important thing of face angle estimation is seeking out facial features that could be set as a standard of angle for estimating orientation. Then, we cite the triangle of face features which is called as “Triangle-based criterion” [9] as our method of estimation for face orientation. Therefore, we can use the information of Labeling which mentioned above, the anthropometric comparison of face geometry and the proportion of face feature triangle to be a measurement criterion of our method.

Triangle-based criterion [9] is combining with the three centers of each Labeling block and the face geometric features that connected the Labeling blocks as a triangle. We could search the possible face triangle in our input photograph and the search flowchart of the face triangle would be indicated as Figure 3.10.

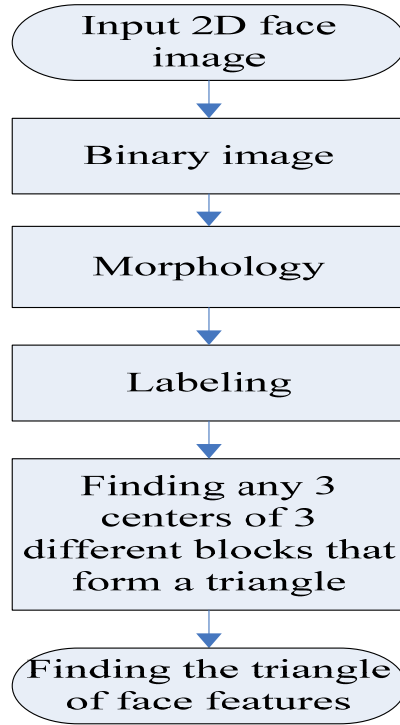


Figure 3.10 The searching flowchart of the face triangle



The main purpose of this process is to find the regions in an input image that might potentially contain faces. There are five steps in the process of searching the Triangle-based criteria.

Firstly, read in a photograph, in our system we transform this input photograph to a gray-level one. The convert function is shown bellow as Equation (3.15).

$$Y = 0.299R + 0.587G + 0.114B \quad (3.15)$$

where,  $Y$  is the index of the gray-level image,  $R$  means the strength of Red color,  $G$  is meaning the Green color of the visual light,  $B$  means the strength of Blue color.

Secondly, we convert the gray-level image to a binary one with the threshold by the method of Otsu. Pixels with gray level of input image  $>$  threshold  $T$  are labeled white, and any pixels with gray intensity of input image  $\leq$  threshold  $T$  are labeled black. Therefore, the output binary image has values 0 (white) for all pixels in the input image with luminance larger than threshold  $T$  and 1 (black) for the other pixels.

Thirdly, use different advantages of Morphology: performing the closing operation (dilation first, then erosion) to fill the holes, and performing the opening operation (erosion first, then dilation) to remove noises.



Fourthly, label all 4-connected components in the image to form blocks and find out the center coordinate of each block that would be form a face triangle. The details of connected component finding can be found in Ref. [2].

Fifthly, detect any 3 centers of 3 different blocks that form a face triangle. We could seek out the potential face regions that contained about “two eyes and one mouth.” As following bellow as Figure 3.11, we would show out the matching rules for finding a face feature triangle. We figured out that the Euclidean distance between left eye (  $i$  ) and right eye

( k ) is about 90-110% of the Euclidean distance between the center of the left or right eye and the mouth. Owing to the effect of imaging and the result of imperfect binarization, thereby, a 25% deviation is given to absorb the tolerance.

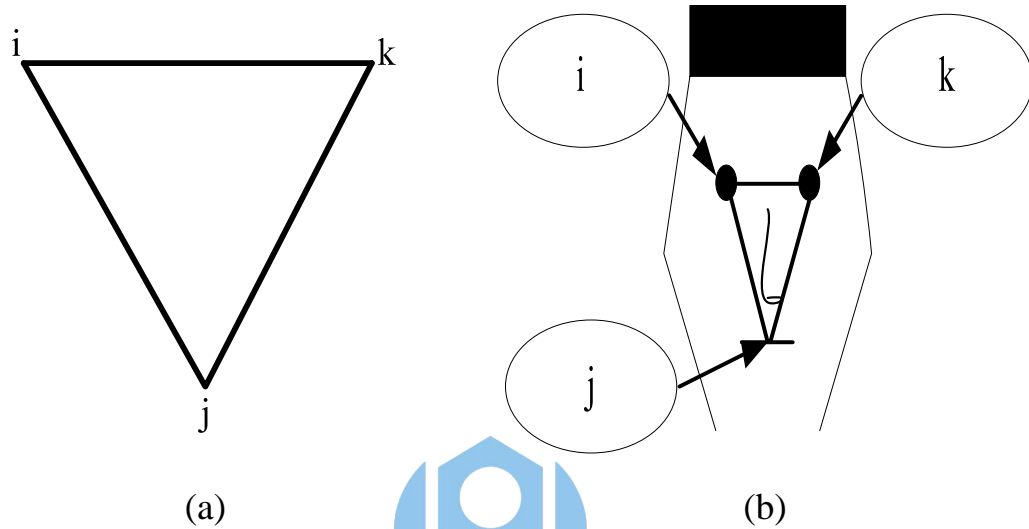


Figure 3.11 (a) The triangle-based i j k (b) Three red points( i, j, and k) satisfy the matching rules, which will form a triangle

The matching rule can thereby be stated as following bellow like Equation (3.16), (3.17), and (3.18).

$$abs(D(i, j) - D(j, k)) < 0.25 \max(D(i, j), D(j, k)) \quad (3.16)$$

$$abs(D(i, j) - D(i, k)) < 0.25 \max(D(i, j), D(j, k)) \quad (3.17)$$

where, “*abs*” means the absolute value. “ $D(i, j)$ ” shows out the Euclidean distance between the centers of block i (right eye) and block j (mouth). “ $D(j, k)$ ” denotes the Euclidean distance between the centers of

block  $k$  (left eye) and block  $j$  (mouth). “ $D(i, k)$ ” represents the Euclidean distance between the centers of block  $i$  (right eye) and block  $k$  (left eye).

Since the labeling process is operated from left to right then from top to bottom, we could get the third matching rule as Equation (3.18).

$$i < j < k \quad (3.18)$$

Therefore, we could find out the regions in an input image that might potentially contain face.





### 3.1.2.1 The speedup of searching for a triangle [9]

When we are looking for any three centers to form a triangle by the matching rules as mentioned previously. The first matching rule is Equation (3.16), the second matching rule is Equation (3.17), and the third matching rule is Equation (3.18) (because the labeling process is operated from left to right then from top to bottom). If the Euclidean distance between the centers of block  $i$  (right eye) and block  $j$  (mouth) is already known, then block center  $k$  (left eye) should locate in the area of 75-125% of Euclidean distance between the centers of block  $i$  (right eye) and block  $j$  (mouth), which will form a circle. Furthermore, since the third matching rule is  $i < j < k$ , the third block center  $k$  is only limited to the up-right part of the circle which is formed by Equation (3.16) and Equation (3.17). In other words, the search area is only limited in the dark area instead of the whole area of the image as shown in Figure 3.12.

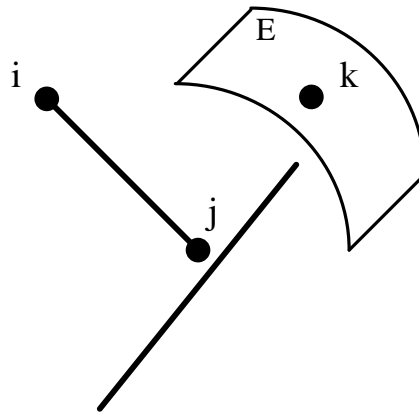


Figure 3.12 The search area of the third block center  $k$  is only limited to the block,  $E$ , instead of the whole area of the image.[9]

As a result, it is not really a selection of  $C_3^n$  (select any 3 combinations from  $n$  blocks). In this way, the triangle-based segmentation process can reduce the background part of a cluttered image up to 91~97%. This process significantly speeds up the subsequent face detection procedure because only 3-9% regions of the original image are left for further processing.

### 3.1.3 Orientation estimation of triangle-based

The goal of our method is estimating the possible orientation of human face by the triangle-based. We use a statistical triangle model to analyze the face orientation and obtain three orientation angles: Roll, Yaw and Pitch (Figure 1.1). The details of our method are shown bellow.

#### 3.1.3.1 Estimation of Yaw Angle

After we find out the face feature triangle, we could discover that the information of triangle would be changed with different Yaw angle. Therefore, we possess the parameters of the triangle geometry (as in Figure 3.13 ) that can be used to estimate angle of face rotation.

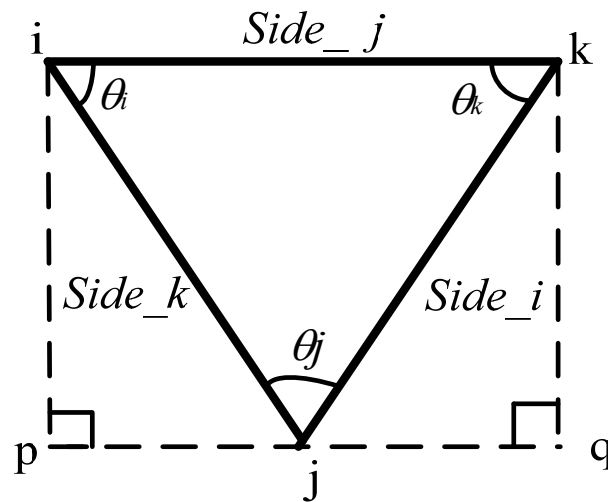


Figure 3.13 The parameters of face triangle geometry

The parameters of face triangle geometry would be calculated by the equations as following bellow.

The lengths of each side are shown bellow.

( $D$  denotes Euclidean distance.)

$$Side\_i = D(j, k) \quad (3.19)$$

$$Side\_j = D(i, k) \quad (3.20)$$

$$Side\_k = D(i, j) \quad (3.21)$$

The angles of each feature point are shown bellow.

$$\theta_i = \cos^{-1} \left( \frac{Side\_j^2 + Side\_k^2 - Side\_i^2}{2 \times Side\_j \times Side\_k} \right) \quad (3.22)$$

$$\theta_j = \cos^{-1} \left( \frac{Side\_i^2 + Side\_k^2 - Side\_j^2}{2 \times Side\_i \times Side\_k} \right) \quad (3.23)$$

$$\theta_k = \cos^{-1} \left( \frac{Side\_i^2 + Side\_j^2 - Side\_k^2}{2 \times Side\_i \times Side\_j} \right) \quad (3.24)$$

The parameters of  $Ratio\_R$  and  $Ratio\_L$  for estimation of Yaw angle would be calculated by the follow equations.

$$Ratio\_R = \frac{D(p, j)}{D(i, k)} \quad (3.25)$$

$$Ratio\_L = \frac{D(j, q)}{D(i, k)} \quad (3.26)$$

After experimenting which be mention above, we find out a phenomenon that the parameters of the face triangle,  $Ratio_R$  and  $Ratio_L$  , would be changed the ratio with rotation of the input photographs. The changing information will be record by each input photograph and do gather static to analysis and estimate the face Yaw angle.



In our experiment of estimation for face orientation, we must take a record of every parameter as a database of each person. The parameters that we want to know are illustrated as Figure 3.13. The part of data record to estimate Yaw angle is shown as Table 3.1.

Table 3.1 The part of data record to estimate Yaw angle

Num. of people	Degrees of camera angle ( $^{\circ}$ )	$\theta_i (^{\circ})$	$\theta_j (^{\circ})$	$\theta_k (^{\circ})$	$Ratio\_R$	$Ratio\_L$
person1	45	68	31	81	0.942	0.058
	60	62	43	75	0.760	0.240
	75	60	50	70	0.672	0.328
	90	61	53	66	0.586	0.414
	105	68	51	62	0.464	0.536
	120	77	44	60	0.321	0.679
	135	84	35	61	0.187	0.813
	150	88	24	68	0.033	0.967
person2	45	67	29	84	0.839	0.161
	60	66	35	79	0.715	0.285
	75	62	38	80	0.708	0.292
	90	66	42	72	0.506	0.494
	105	70	41	69	0.412	0.588
	120	74	37	69	0.323	0.677
	135	82	21	77	0.115	0.885
	150	85	24	72	0.024	0.976
person3	45	65	31	85	0.925	0.075
	60	63	43	74	0.714	0.286
	75	61	49	70	0.632	0.368
	90	64	53	63	0.527	0.473
	105	74	48	57	0.379	0.621
	120	83	36	60	0.278	0.722

However, we use a graph to discuss the tendency of changing face orientation. The graph of Figure 3.14 shows out the distribution of  $Ratio\_R$  and orientation angle.

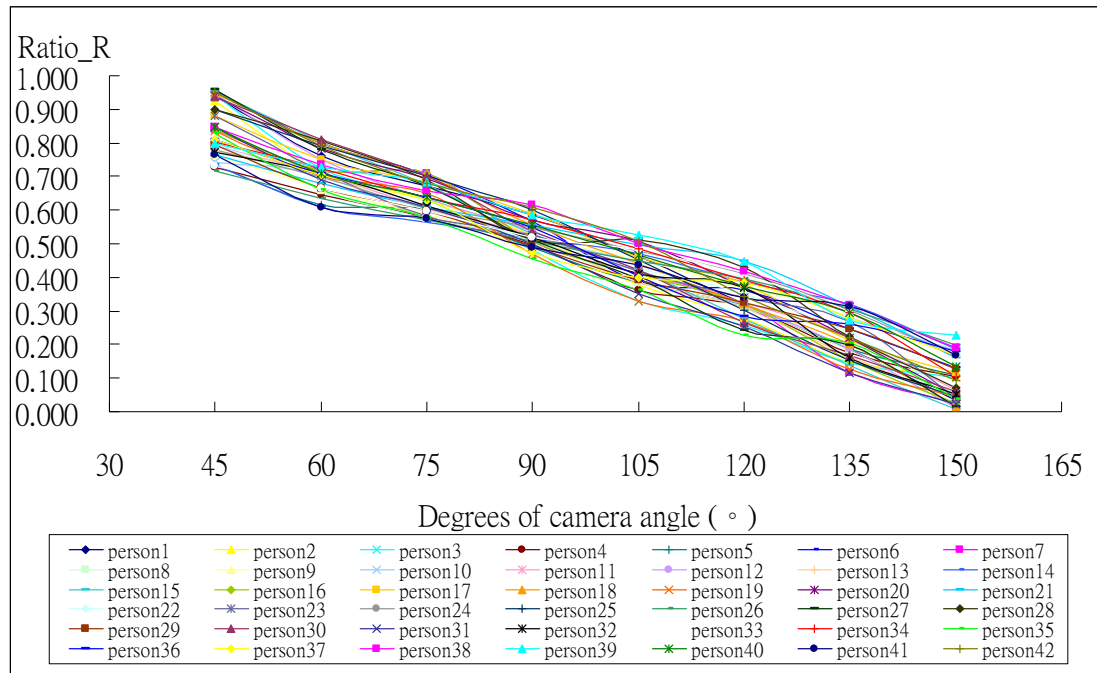


Figure 3.14 The graph of the tendency of changing face orientation

In order to simplify the chart to an easy and a understand one, we apply standard deviation on the data that we could get the sequence of average and standard deviation  $\sigma$  in Table 3.2 and the simplify chart is shown as Figure 3.15.

Table 3.2 The standard deviation of original data

Degrees of camera angle (°)	45	60	75	90	105	120	135	150
Standard deviation( $\sigma$ )	0.078	0.055	0.046	0.038	0.050	0.056	0.063	0.063
average+ $\sigma$	0.917	0.771	0.681	0.568	0.480	0.396	0.285	0.158
average	0.839	0.715	0.634	0.530	0.429	0.340	0.222	0.095
average- $\sigma$	0.761	0.660	0.588	0.492	0.379	0.283	0.159	0.032

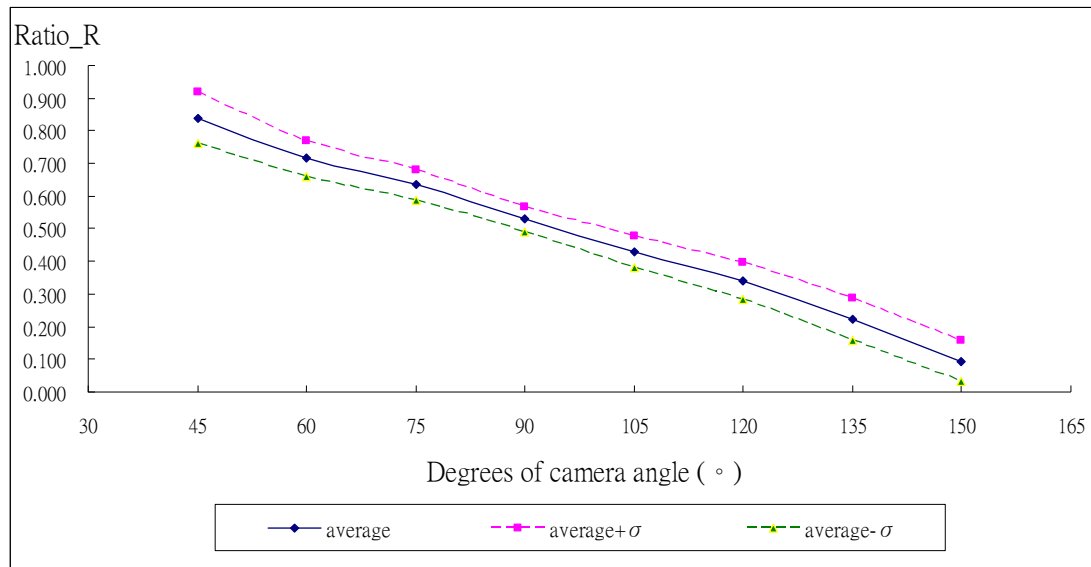


Figure 3.15 The simplify chart of original data

Using the chart as Figure 3.16, we could present some rules to estimate the Yaw angle of input any facial photograph and classify the facial orientation which range it could be. Firstly, we could set the gaps of angle range. The gaps would be determined from the Table 3.2 so that the more detail of the separating rules would be illustrated as Figure 3.16.



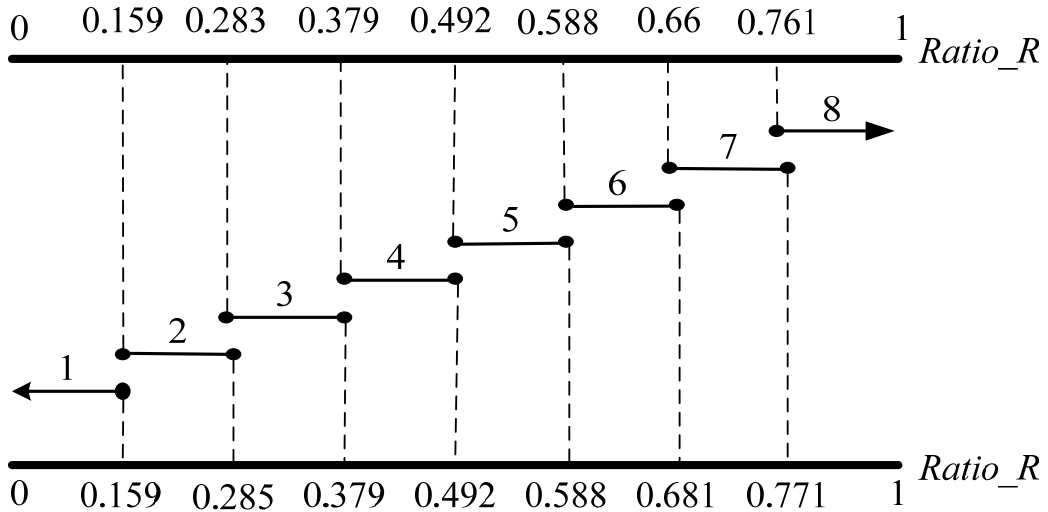


Figure 3.16 The different zones of orientation estimation

The rules of segmentation from each range angle would be listed bellow. If the value of  $Ratio\_R$  is placed as zone 8, we would estimate the face orientation of Yaw angle may be show that “Head\_Orientation: Turn Left, Yaw\_angle : 45 degrees”. In the other words, if the value of  $Ratio\_R$  is placed as zone 7, we would estimate the face orientation of Yaw angle may be show that “Head\_Orientation: Turn Left, Yaw\_angle : 30 degrees”. With the same sequence of head orientation changing, we could reckon the value of  $Ratio\_R$  is placed as zone 1, we would estimate the face orientation of Yaw angle may be show that “Head\_Orientation: Turn right, Yaw\_angle : 60 degrees”. The detail zones of different gaps would be seen in Table 3.3.

Table 3.3 The zones of different gaps

Zone	Head_Orientation; Degrees of real Yaw angle of face orientation (°)	Gaps
1	Turn right , 60°	$Ratio\_R < 0.159$
2	Turn right , 45°	$0.159 \leq Ratio\_R \leq 0.285$
3	Turn right , 30°	$0.283 \leq Ratio\_R \leq 0.379$
4	Turn right , 15°	$0.379 \leq Ratio\_R \leq 0.492$
5	Look front, 0°	$0.492 \leq Ratio\_R \leq 0.588$
6	Turn Left , 15°	$0.588 \leq Ratio\_R \leq 0.681$
7	Turn Left , 30°	$0.660 \leq Ratio\_R \leq 0.771$
8	Turn Left , 45°	$0.761 \leq Ratio\_R$

Using the rules we figured out, the face orientation estimation of input photograph could be applied for the rotation angle of Yaw angle.

### 3.1.3.2 Estimation of Roll Angle

If Pitch and Yaw angles are small, that just like as the situation as near frontal look to the camera, Roll angle could be approximately associated with the 2D rotation of the face in the image, that is, the angle of the line passing through both eyes. In consequence, the estimation of Roll angle would be indicated as Equation (3.27).

$$Roll = \tan^{-1}\left(\frac{Y_k - Y_i}{X_k - X_i}\right) \quad (3.27)$$

where,  $(X_i, Y_i)$  is the coordinate of the right eye.

$(X_k, Y_k)$  is the coordinate of the left eye.



Although Equation (3.27) is valid in most practical situations, we have to observe that it is not precise in all cases. For instance, the face and the camera are at different height level, a horizontal displacement can result in an erroneous roll not equal to 0.

### 3.1.3.3 Estimation of Pitch Angle

Although we mention about the limitation of 2D image, we present a method of statistics to estimate the Pitch angle. In consequence, we could evaluate the accuracy of Pitch angle within 15 degrees and the details of the method would be shown bellow.

If Roll and Yaw angles are small, that just like as the situation as the face is bending or lifting without any rotation or tilt. Pitch angle could be estimated and approximately associated with the looking up or down of the face in the image. The method of estimating Pitch angle would be as similar as estimating Yaw one. The source of 2D face images to estimate the Pitch angle from degrees from 0~180 witch per 15 degrees with horizontally to earth as a unit to capture images. The method of estimation we presented is gathering statistics using the parameters of face-triangle altering with the change of Pitch angle.

With the parameters of face triangle-based (shown as Figure 3.17), we could find out the state of a ratio would be changing when the face is rotating with Pitch angle.

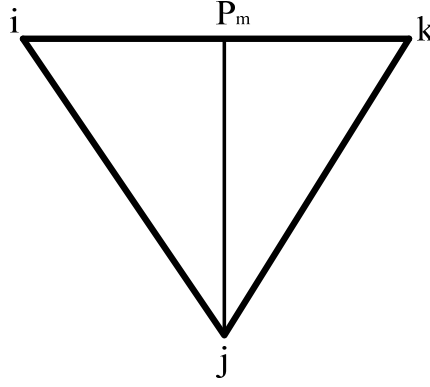


Figure 3.17 The triangle-based for estimating Pitch angle

The parameters of face triangle geometry for Pitch angle estimation would be calculated by the equations as following bellow.

$$D(i, P_m) = D(P_m, k) \quad (3.28)$$

$$Face\_H = D(P_m, j) \quad (3.29)$$

$$Ratio\_U = Face\_H / D(i, k) \quad (3.30)$$



The method of estimation for Pitch angle we proposed is making a statistics of the parameters of face triangle-based. The parameters for Pitch angle estimation are shown as Figure 3.17). And we take a recorder of every parameter for estimating Pitch angle witch shown as Table 3.4.

Table 3.4 The part of data record for estimating Pitch angle

Num. of people	Degrees of Pitch angle ( ° )	Side_j	Face_H	Ratio_U
person1	45	138	60	0.435
	60	140	87	0.621
	75	142	100	0.703
	90	143	121	0.849
	105	145	103	0.710
	120	146	85	0.582
	135	147	65	0.442
	150	149	56	0.376
person2	45	132	56	0.424
	60	133	79	0.594
	75	136	98	0.721
	90	137	119	0.869
	105	137	100	0.730
	120	138	83	0.601
	135	138	62	0.449
	150	140	55	0.393
person3	45	135	60	0.444
	60	135	80	0.593
	75	138	98	0.710
	90	141	118	0.837
	105	141	102	0.723
	120	143	83	0.580
	135	144	66	0.458
	150	145	58	0.400

However, we use a graph to discuss the tendency of changing face orientation. The graph of Figure 3.18 shows out the distribution of  $Ratio\_U$  and orientation of Pitch angle.

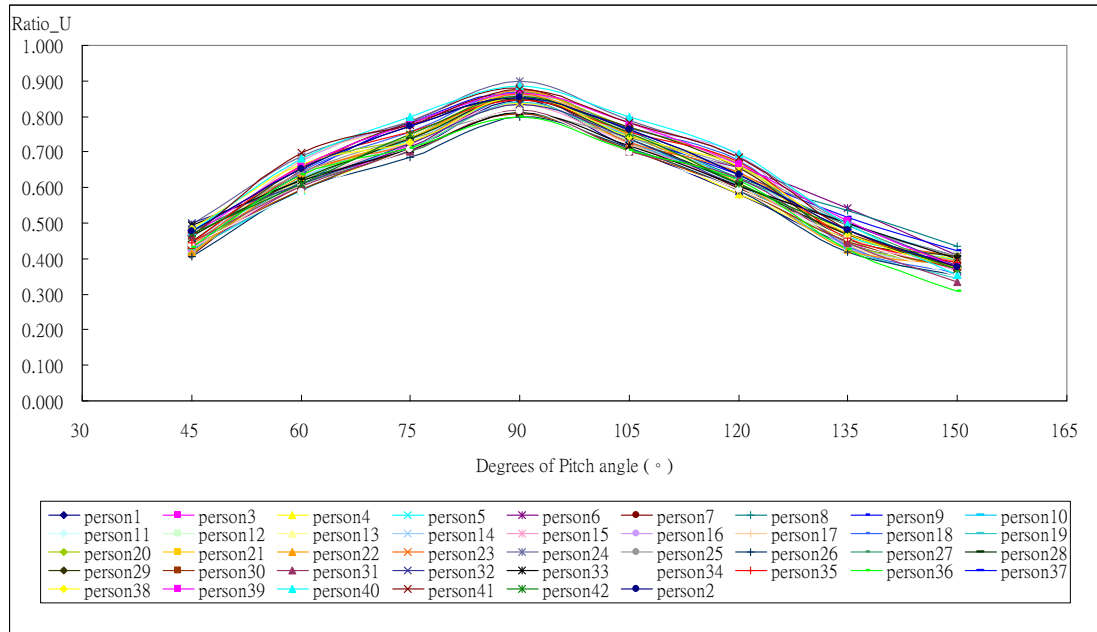


Figure 3.18 The graph of the tendency of changing Pitch angle

In order to simplify the chart to an easy and understandable one, we apply standard deviation on the data that we could get the sequence of average and standard deviation  $\sigma$  in Table 3.5 and the simplify chart is shown as Figure 3.19.

Table 3.5 The standard deviation of data for Pitch angle estimation

Degrees of camera angle vertically to ground( $^{\circ}$ )	45	60	75	90	105	120	135	150
Standard deviation( $\sigma$ )	0.024	0.030	0.029	0.025	0.025	0.032	0.031	0.024
average+ $\sigma$	0.471	0.661	0.769	0.875	0.768	0.664	0.497	0.406
average	0.447	0.631	0.739	0.850	0.742	0.632	0.466	0.382
average- $\sigma$	0.422	0.602	0.710	0.826	0.717	0.599	0.435	0.358

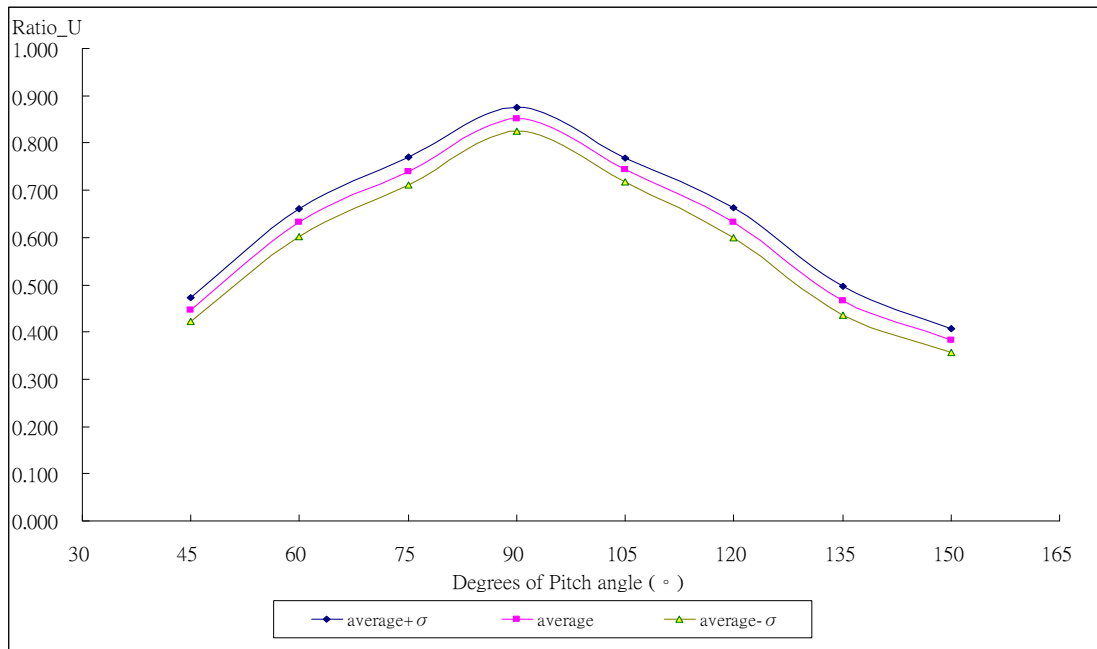


Figure 3.19 The simple chart of original data for Pitch angle estimation

From the chart as Figure 3.19, we could see the almost symmetric distribution of statistics data. Although we could use the characteristic of orientation angular distribution, with the same index of  $Ratio\_U$ , there would be the confused area, such as the angle of  $45^{\circ}$ ,  $135^{\circ}$  ;



60°,120°;75°,105°, of this chart to us. We thereby try to use another index to distinguish the overlapping area of the same  $Ratio\_U$ .

The condition of estimating Pitch angle is restricted to a small angle of Yaw's and Roll's. Therefore, we can use the index (such as Table 3.6) which is the area of face triangle (abbreviation :  $A_T$ ) to distinguish the face is under the state of bending or lifting. From the data of Table 3.4, an interesting phenomenon of the length,  $D(i, k)$ , is increasing with angle of Pitch becoming big. With the situation, we try to calculate out and then average the areas of each face triangle under the same condition of face orientation. The result shows us that although in the confused area, there would be difference of triangle size.

Table 3.6 The part of data record for estimating Pitch angle

Num. of people	Degrees of Pitch angle ( ° )	Side_j	Face_H	Ratio_U	Area of Triangle
person1	45	138	60	0.435	4140
	60	140	87	0.621	6090
	75	142	100	0.703	7089
	90	143	121	0.849	8676
	105	145	103	0.710	7468
	120	146	85	0.582	6205
	135	147	65	0.442	4778
	150	149	56	0.376	4172

Table 3.6 shows us the area of triangle within confused area, such as the angle of  $45^\circ, 135^\circ$ ;  $60^\circ, 120^\circ$ ;  $75^\circ, 105^\circ$  is difference. Therefore, we try to find out a suitable threshold to separate the angle of confused part by the average area from every area of triangle with the same Pitch angle in our database. Hence, we could get three thresholds from three confused part. The three thresholds are  $T_1 = 6833$  for the confused part of  $75^\circ \& 105^\circ$ ,  $T_2 = 5809$  for the confused part of  $60^\circ \& 120^\circ$ , and  $T_3 = 4222$  for the confused part of  $45^\circ \& 135^\circ$ .

Using thereby three thresholds and the chart as Figure 3.19, we could present some rules to estimate the Pitch angle of input any facial photograph and classify the facial orientation which range it could be. Firstly, we could set the gaps of angle range. The Zones would be determined from the Table 3.5, so that the more details of the separating rules would be illustrated as Figure 3.20.

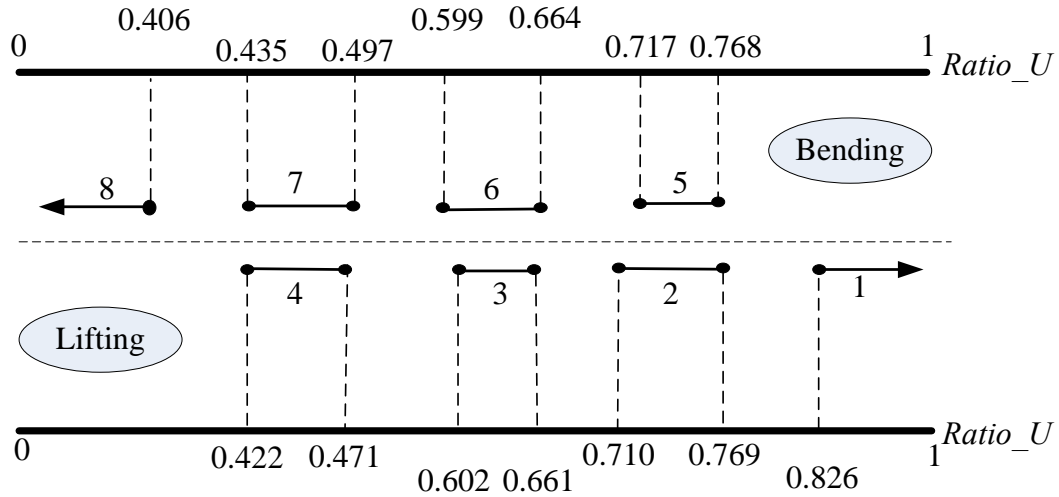


Figure 3.20 The different zones for estimation of Pitch angle

The rules of segmentation from each range angle would be listed as bellow. If the value of  $Ratio\_U$  is placed as zone 1, we would estimate the face orientation of Pitch angle may be show that “Head\_Orientation : Look front, Pitch\_angle : 0 degrees”. In the other words, if the value of  $Ratio\_U$  is placed as zone 2, we would estimate the face orientation of Pitch angle may be show that “Head\_Orientation : Lifting, Yaw\_angle : 15 degrees”. With the same sequence of head orientation changing, we could reckon the value of  $Ratio\_U$  is placed as zone 8, we would estimate the face orientation of Pitch angle may be show that “Head\_Orientation : Bending, Pitch\_angle : 60 degrees”. The detail of zones in different gaps would be seen in Table 3.7.

Table 3.7 The zones of different gaps for Pitch angle evaluation

Zone	Head_Orientation ; Degrees of real Pitch angle of face orientation (°)	Gaps
1	Look front, 0°	$0.826 \leq Ratio\_U$
2	Lifting , 15°	$0.710 \leq Ratio\_U \leq 0.769 \ \& \ A_r < T_1$
3	Lifting , 30°	$0.602 \leq Ratio\_U \leq 0.661 \ \& \ A_r < T_2$
4	Lifting , 45°	$0.422 \leq Ratio\_U \leq 0.471 \ \& \ A_r < T_3$
5	Bending , 15°	$0.717 \leq Ratio\_U \leq 0.768 \ \& \ A_r > T_1$
6	Bending , 30°	$0.599 \leq Ratio\_U \leq 0.664 \ \& \ A_r > T_2$
7	Bending , 45°	$0.435 \leq Ratio\_U \leq 0.497 \ \& \ A_r > T_3$
8	Bending , 60°	$Ratio\_U \leq 0.406$

Using the rules we figured out, the face orientation estimation of input photograph could be applied the third rotation axis of Pitch angle.

## 3.2 The Angle Estimation of 3D Point-clouds

The 2D face orientations were already estimated by the method of estimation we presented. For estimating of the three orientations of 3D point-clouds, we use the method of projecting images from the 3D point-clouds which we got in section 2.2. We use the similar method of estimation for 2D images to estimate three possible angles of Roll, Yaw and Pitch. The flowchart for estimation of 3D point-clouds was illustrated as Figure 3.21.

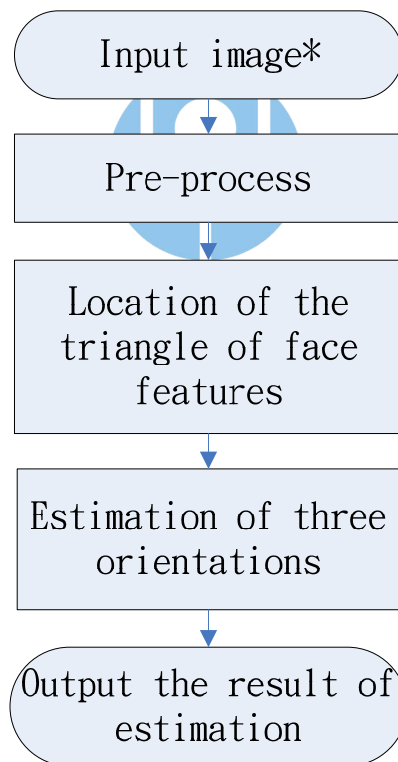


Figure 3.21 The flowchart of estimating orientations for 3D point-clouds

There are five steps in the method we presented for estimating of face orientation in the 3D point-clouds. Firstly, we input an image\* which is projective image (details in section 2.2) from 3D point-clouds. Secondly, utilize the work of pre-process in section 3.1.1 to get information of black blocks. Thirdly, the marked blocks would feed into the triangle-based finder (pre-introduction in section 3.1.2) to figure out the face triangle. Fourthly, the triangle area covered two eyes and one mouth provides the coordinates of three feature points, therefore, we could calculate out the parameters for estimation of face orientation. Finally, we could get the result of orientation estimation and then comparison the result and the one from estimating 2D image. Therefore, we could see the difference of the two results from 2D images and 3D point-clouds.

## CHAPTER 4 EXPERIMENTAL RESULTS

The experiment devices contains not only we mentioned before in the chapter 2 but also a personal computer which computing the methods of digital image processing we showed above. The software is developed with Borland C++ Builder 6.0 and all programs could implement on Window XP. Then all of the experiment results would be shown bellow. We would show out the orientation estimation of 2D facial photograph. In the part of section 4.1.1, the result of Yaw angle of 2D facial photograph could be shown and discussed. Within this part of section 4.1.2, the result of estimation of Roll angle of input facial photograph would be listed out and discussed. At the other section 4.1.3, we might discuss the result of estimation of Pitch angle.

### 4.1 Results of angle estimation of face triangle

In this part, we would show out the result of orientation estimation for 2D facial photograph. Because that we use the facial feature triangle as a characteristic mark. The different and changeable parameters would interest us how the values change with the different face orientation. All the relevant estimation of face orientation would be mentioned at section 3.1.3. There are totally 42 people in our database; 4females and 38 males.

### 4.1.1 The result of estimation for Yaw angle

In the Figure 1.1, we could figure out Yaw angle is one of the rotating axis which is belong to y-axis. With the face that turning right or turning left, the Yaw angle would be changed. Though doing gather static with the variability of changing face orientation, the result of estimation for Yaw angle is shown bellow as Figure 4.1.

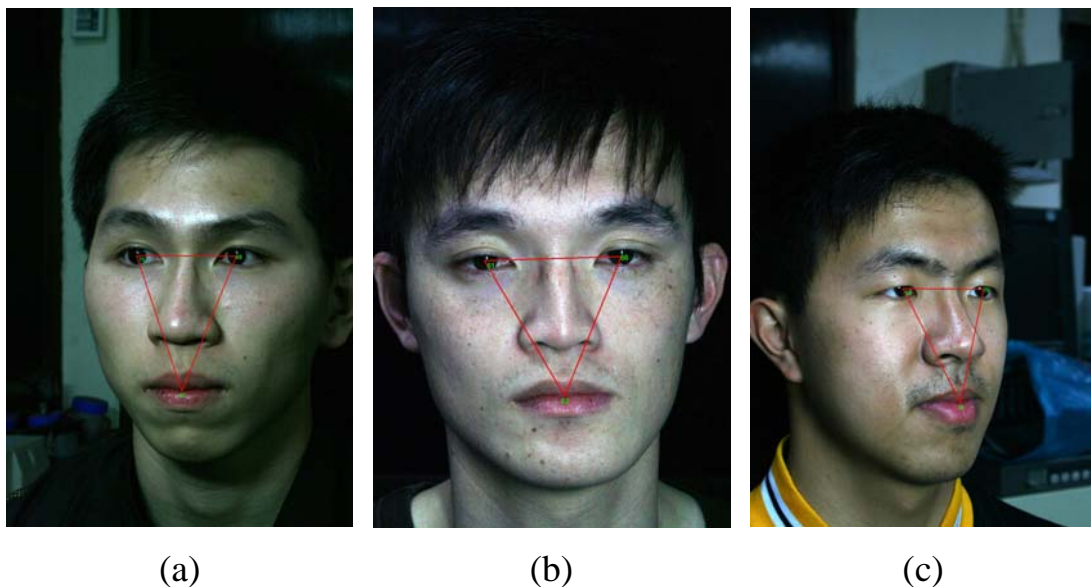


Figure 4.1 The results of estimation for Yaw angle

(a) Turn right , 15 degrees (  $Ratio\_R = 0.453$  )

(b) Look front, 0 degrees (  $Ratio\_R = 0.586$  )

(c) Turn left , 30 degrees (  $Ratio\_R = 0.712$  )

Using the face orientation estimation method we presented, the rotation of Yaw angle could be classified by 15 degrees per one unit.



### 4.1.2 The result of estimation for Roll angle

In the Figure 1.1, we could figure out Roll angle is one of the rotating axis which is belong to z-axis. With the face that occurs to tilt, then Yaw angle would be changed. In the Figure 4.2, we can see the samples of estimation of Roll angle.

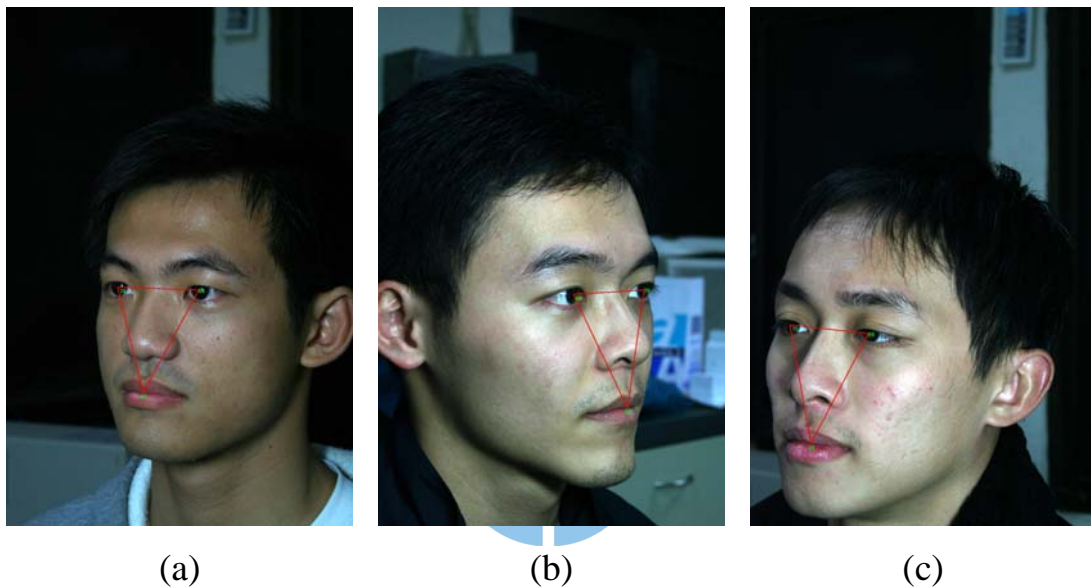


Figure 4.2 The results of estimation for Roll angle

(a) -2.11 degrees

(b) 3.41 degrees

(c) -6.33 degrees

Using the estimation of Roll angle in section 3.1.3.2, the rotation of Roll angle could be calculated out even smaller than one degree.

### 4.1.3 The result of estimation for Pitch angle

In the Figure 1.1, we could figure out Pitch angle is one of the rotating axis which is belong to x-axis. With the face that occurs to bend or lift, then Pitch angle would be changed. In the Figure 4.3, we can see the samples of estimation of Pitch angle.



Figure 4.3 The results of estimation for Pitch angle

(a) Bending; 15 degrees

(  $Ratio\_U = 0.766$ ,  $A_r = 7911$  )

(b) Look front; 0 degrees (  $Ratio\_U = 0.839$  )

(c) Lifting ; 30 degrees

(  $Ratio\_U = 0.642$ ,  $A_r = 5553$  )

Using the estimation of Pitch angle in section 3.1.3.3, the rotation of Pitch angle could be calculated out within 15 degrees.




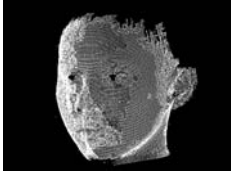

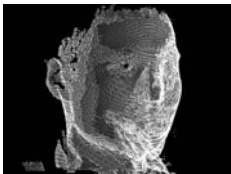
## **4.2 Results of comparison estimating consequence with 2D image and projective image from point-clouds**

In this part, we would show out the two kinds of results with estimation for human face orientation from the 2D images and 3D point-clouds. (shown as Table 4.1)

From the experimental results, we can estimate the Yaw and Pitch angles by classifying each 15 degree as a unit from the range of 45~150 degrees, while the Roll angle within 1 degree. No matter the applied data are 2D images or 3D point-clouds.



Table 4.1 Results of comparison estimating consequence with 2D image and projective image from point-clouds

Items	Images	$Ratio_R$	$Ratio_U$	$A_T$	Orientation		
					Yaw	Pitch	Roll
(a)		0.60	0.60	5553	L.F., 0°	L., 30°	0.79°
(b)		0.59	0.59	5731	L.F., 0°	L., 30°	0.82°
(c)		0.39	1.06	5284	T.R., 15°	L.F., 0°	0.36°
(d)		0.4	1.08	5188	T.R., 15°	L.F., 0°	0.98°
(e)		0.67	1.02	5614	L.F., 30°	L.F., 0°	1.88°
(f)		0.68	1.05	5147	T.R., 30°	L.F., 0°	1.75°

where, L.F., T.R., T.L., B., and L. witch means Look front, Turn right, Turn left, Bending, and Lifting.

## CHAPTER 5 CONCLUSION AND PROSPECTS

In this thesis, a method of face orientation estimation for 2D images and 3D point-clouds is proposed. The results of estimation for face orientation show that as follow bellow.

1. We could achieve the goal of accuracy about 15 degrees of Yaw angle and Pitch angle.
2. The result of estimation of Roll angle could be exquisite around 1 degree.
3. We can execute the method of face orientation estimation on no matter 2D images or 3D point-clouds.



In the future, with estimation of the orientation for human face, we may use the information to reset the orientation to an original situation, figure out the original captured states of camera. Even, we can imitate the same condition of the input image at the moment by point-clouds images, that the court of forensic would use them as the reference.

And, we can try to improve our method to apply on multi-people about the estimation of faces orientation at the same time or to use the orientation information for re-orientation, even, to develop a method of face retrieval. In addition, improve the method estimation for face orientation on the people with the face features were hided from view is also a task challenge, because the features missing would affect the performance of face region finder. At the worst result, we could not find out a face-feature triangle in the beginning. Another task is, however, estimation of face orientation in unconstrained real-life environments remains challenging for most practical applications. At the summary, there are still some works that we can work on.

## REFERENCES

- [1] Otsu N. , “A threshold selection method from gray level Histogram,” *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 9, No. 1, pp. 62-66 (1979).
- [2] Gonzalez R. C. and Woods R. E., *Digital Image Processing*, Addison Wesley, New York (1992).
- [3] Gee A. H. , and Cipolla R. , “Determining The Gaze Of Faces In Images,” *Image and Vision Computing*, Vol. 12, No. 10, pp. 639-647 (1994).
- [4] Jain R., Kasturi R., and Schunck B.G., *Machine Vision*, International Editions, McGRAW-HILL, New York, (1995).
- [5] Horprasert T. , Yacoob Y. , and Davis L. S. , “Computing 3D Head Orientation from a Monocular Image Sequence,” *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pp. 242-247 (1996).
- [6] Cherist M. , Said J. N. , and Suen C. Y. , “A recursive thresholding technique for image segmentation,” *IEEE Trans. on Image Processing*, Institute of Electrical and Electronics Engineers, Vol. 7, No. 6, pp. 918-921 (1998).

- [7] Yoshino M. , Matsuda H. , Kubota S. , Imaizumi K. , and Miyasaka S. , “Assessment of Computer-assisted Comparison between 3D and 2D Facial Images,” *Japanese journal of science and technology for identification*, Vol. 5, No. 1, pp. 9-15 (2000).
- [8] Yoshino M. , Matsuda H. , Kubota S. , Imaizumi K. , and Miyasaka S. , “Computer-assisted facial image identification system using a 3-D physiognomic range finder,” *Forensic Science International*, Elsevier BV, Vol. 109, Issue 3, pp. 225-237 (2000).
- [9] Lin C. , and Fan K. -C. , “Triangle-based approach to the detection of human face,” *Pattern Recognition Letters*, Elsevier BV, Vol. 34, Issue 6, pp. 1271-1284 (2001).
- [10] Ji Q. , and Hu R. , “3D Face pose estimation and tracking from a monocular camera,” *Image and Vision Computing*, Elsevier BV, Vol. 20, Issue 7, pp. 499-511 (2002).
- [11] Yoshino M. , Noguchi K. , Atsuchi M. , Kubota S. , Imaizumi K. , Thomas C. David L. , and Clement John G. , “Individual identification of disguised faced by morphometrical matching,” *Forensic Science International*, Vol. 127, Issue 1-2, pp. 97-103 (2002).
- [12] Yoshino M. , “Recent Advance in Facial Image Identification,”



- Japanese journal of science and technology for identification*, Vol. 7,  
No. 1, pp. 1-17 (2002).
- [13] Chung K. L. , Image Processing and Computer Vision (in Chinese),  
4<sup>th</sup> Edition, Dong-Hua Pub, Taipei (2008).
- [14] Yoshino M. , Imaizumi K. , Tanijiri T. , and G. J. , “Automatic  
Adjustment of Facial Orientation in 3D Face Image Database,”  
*Japanese journal of science and technology for identification*, Vol. 8,  
No. 1, pp. 41-47 (2003).
- [15] Lin C. , and Fan K. -C. , “Pose classification of human faces by  
weighting mask function approach,” *Pattern Recognition Letters*,  
Elsevier BV, Vol. 24, Issue 12, pp. 1857-1869 (2003).
- [16] Zhao W. , Chellappa R. , Phillips P. J. , and Rosenfeld, “Face  
Recognition : A Literature Survey,” *ACM Computing Survey*,  
Association for Computing Machinery, Vol. 35, Issue 4, pp. 399-458  
(2003).
- [17] Adams G. L. , Gransky S. A. , Miller A. J. , Harrell W. E. , Hatcher J.  
D. and D. C. , “Comparison between traditional 2dimensional  
cephalometry and a 3dimensional approach on human dry skulls,”  
*American Journal of Orthodontics and Dentofacial Orthopedics*,

Comparative Study, Vol. 126, Issue 4, pp. 397-409 (2004).

[18]Hommond P. , Hutton T. J. , Allanson J. E. , Hennekam R. C. M. ,  
Holden S. , and Patton M. A. , “3D Analysis of Facial Morphology,”  
*American Journal of Medical Genetics*, Vol. 126A, Issue 4, pp.  
339-348 (2004).

[19]Ansari A. N. , and Abdel-Mpttaleb M. , “Automatic facial feature  
extraction and 3D face modeling using two orthogonal views with  
application to 3D face recognition,” *Pattern Recognition*, Pergamon  
Press, Vol. 38, Issue 12, pp. 2549-2563 (2005).

[20]Song H. , Lee S. , Kim J. , and Sohn K. , “Three-dimensional  
sensor-based face recognition,” *Applied Optical*, Optical Society of  
America, Vol. 44, Issue 5, pp. 677-687 (2005).

[21]Kittler J. , Hamouz M. , Tena J. R. , Hilton A. , Illingworth J. , and  
Ruiz M. , “3D assisted 2D face recognition : Methodology,” *Lecture  
Notes in Computer Science(including subseries Lecture Notes in  
Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer  
Verlag, 3773 LNCS, pp. 1055-1065 (2005).

[22]Yoshino M. , Taniquchi M. , Imaizumi K. , Miyasaka S. , Tanijiri T. ,  
Yano H. , and David C. , “A new retrieval system for a database of

- 3D facial images,” *Forensic Science International*, Vol. 148, Issues 2-3, pp. 113-120 (2005).
- [23]Phillips P. J. , Flynn P. J. , Scruggs T. , Bowyer K. W. , Chang J. , Hoffman K. , Marques J. , Min J. , and Worek, “Overview of the Face Recognition Grand Challenge,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 947-654 (2005).
- [24]Goos M. I. , Alberink I. B. , and Ruifrok A. C. , “2D/3D image (facial) comparison using camera matching,” *Forensic Science International*, Elsevier BV, Vol. 163, Issues 1-2, pp. 10-17 (2006).
- [25]Chai X. , Shan S. , Qing L. , and Gao W. , “Pose Estimation Based on Gaussian Error Models,” *Lecture Notes in Computer Science*, Springer Verlag, LNCS 3832, pp. 136-143 (2006).
- [26]Ghoddousi H. , Edler R. , Haers P. , Wertheim D. , and Greenhill D. , “Comparison of three methods of facial measurement,” *International Journal of Oral and Maxillofacial Surgery*, Churchill Livingstone, Vol. 36, Issue 3, pp. 250-258 (2007).
- [27]Savakis A. , Erhard M. , Schimmel J. , and Hnatow J. , “A multi-camera system for real-time pose estimation,” *Proceedings of SPIE – The International Society for Optical Engineering*, Vol. 6560,

art. No. 656006(2007).

[28]Wang J. -G. , and Sung E. , “EM enhancement of 3D head pose estimated by point at infinity,” *Image and Vision Computing*, Elsevier BV, Vol. 25, No. 18, pp. 1864-1874 (2007).

[29]Kinoshita K. , and Lao S. , “A Fast and Robust 3D Head Pose and Gaze Estimation System,” *IEEE International Conference on Automatic Face and Gesture Recognition*, art. no. 4813466 (2008).

[30]Martins P. , and Batista J. , “Accurate single view model-based head pose estimation,” *IEEE International Conference on Automatic Face and Gesture Recognition*, FG 2008, art. No. 48133369 (2008).

[31]Garcia-Mateos G. , Ruiz A. , Lopez-de-Teruel P. E. , Rodriguez A. L. , and Fernandez L. , “Estimating 3D facial pose in video with just three points,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, art. No. 4563050 (2008).

[32]Nemanja G. , and Slobodan I. , “Facial pose estimation for image retrieval,” *Vincent Lepetit, Pascal Fua, CH-1015 Lausanne, Switzerland* (2008).

[33]Breitenstein M. D. , Kuettel D. , Weise T. , Van Gool L. , and Pfister H. , “Real-Time Face Pose Estimation from Single Range Images,”

*IEEE International Conference on Automatic Face and Gesture Recognition*, art. No. 4587807 (2008).

[34]Canton-Ferrer C. , Casas J. R. , and Paradas M. , “Head Orientation Estimation using Particle Filtering in Multiview Scenarios,” *Lecture Notes in Computer Science(including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 4625 LNCS, pp. 317-327 (2008).

[35]Jimenez P. , Nuevo J. , Bergasa L. M. , and Sotelo M. A. , “Face tracking and pose estimation with automatic three-dimensional model construction,” *Institution of Engineering and Technology Computer Vision*, Institution of Engineering and Technology, Vol. 3, Issue 2, pp. 93-102(2009).

[36]Angelis D. D. , Sala R. , and Cantatore A. , “A new computer-assisted technique to aid personal identification,” *International Journal of Legal Medicine*, Springer Berlin / Heidelberg, Vol. 123, pp. 351-356 (2009).

[37]Mayo M. , and Zhang E. , “3D face recognition using multiview keypoint matching,” *IEEE International Conference on Advanced Video and Signal Surveillance*, Genoa, Italy (2009).

- [38]Zhang X. , and Gao Y. , “Face recognition across pose : A review,”  
*Pattern Recognition*, Vol. 42, No.11, pp. 2876-2896 (2009).
- [39]Murphy-Chutorian E. , and Trivedi M. M. , “Head Pose Estimation in  
Computer Vision : A Survey,” *IEEE Transactions of Pattern Analysis  
and Machine Intelligence*, Vol. 31, No.4, pp. 607-626 (2009).
- [40]Smeets D. , Claes P. , Vandermeulen D. , and Clement JG. ,  
“Objective 3D face recognition: Evolution, approaches and  
challenges,” *Forensic Science International*, FSI-6001, pp.  
1-8(2010).
- [41]Shan C. , “Face Recognition and Retrieval in Video,” *Studies in  
Computational Intelligence*, Springer Berlin / Heidelberg , Vol. 287,  
pp. 235-260(2010).

# 國立臺灣科技大學博碩士論文授權書

(本授權書裝訂於紙本論文內)

本授權書所授權之論文為黃勝達〔M9704307〕在國立臺灣科技大學高分子系 98 學年度第 2 學期取得碩士學位之論文。

論文題目：應用二維人臉影像與三維點雲資料於人臉姿態估測之研究  
指導教授：邱士軒

茲同意將授權人擁有著作權之上列論文全文〔含摘要〕，依下述授權範圍，以非專屬、無償授權本校圖書館及國家圖書館，不限地域、時間與次數，以紙本、微縮、光碟或其他數位化方式將上列論文重製典藏，並提供讀者基於個人非營利性質之線上檢索書目、館內閱覽、或複印。

授權人

黃勝達

邱士軒

簽章

(請親筆正楷簽名)

黃勝達

邱士軒

備註：

1. 授權人不因本授權而喪失上述著作之著作權。
2. 本授權書請授權人簽章後，裝訂於紙本論文內。

中 華 民 國 99 年 7 月 29 日