3D Manikin Face Modelling And Super-Resolution From Range Images

Yin Yin

Kongens Lyngby 2006
IMM-MSC-2006-70

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk

Abstract

In this thesis, a trial of modelling a manikin face using SwissRanger SR-3000 is implemented. The process includes acquiring data, range data restoration, registration and surface reconstruction.

Several tests are done to evaluate the camera's performance. Then, the noisy and low-resolution range images are restored by MRF by designing intensity information into the prior so that the restored range measurements obtain the high contrast property of the intensity information.

The range images are registered by ICP algorithm. To improve the performance of ICP according to the data, several variants are introduced.

A new surface reconstruction and super-resolution algorithm called 2.5D MRF is originated to combine multiple registered surfaces. This high dimensional MRF merges surfaces by trying to move locally smooth patches together and keep the original values for details. The algorithm is proved to be robust to noise and registration errors.

Finally, a face model combined by 15 registered views via simple averaging and super-resolution of the face combined by 3 views via 2.5D MRF are displayed as the result.

Keywords: 3D modelling, super-resolution, ICP, SwissRanger, 2.5D MRF

Preface

This thesis was prepared at Informatics Mathematical Modelling, the Technical University of Denmark in partial fulfillment of the requirements for acquiring the M.Sc. degree in engineering.

This thesis is about a project of 3D modelling from scans, which covers 3D registration, image restoration topics in Computer Vision and surface reconstruction in Computer Graphics fields.

The project is equivalent to 30 ECTS points and lasts 6.5 months including 1.5 months' vacation. The project started on February, 2006 and will finish on August, 2006.

During the period, the author received a Research Assistantship for the Ph.D. study in the University of Iowa, USA. The study also starts on August, 2006. This project would be a good preparation for the future research.

Yin Yin, August 2006

Acknowledgements

First of all, this thesis is for my dear mother. She encourages me to go to Denmark and study for my M.Sc. degree in DTU. Now, she has been struggling with a cancer for one and a half years. I believe her braveness and the love from me, my father, my aunts, as well as other relatives and friends can help her conquer this deadly disease.

Within the recent half year, my work is contributed by many people in IMM. Most important of them is my supervisor Associate Professor Rasmus Larsen. He provided this interesting project to me and directed me with his plentiful thoughts, technical support and patience throughout the project.

I would also like to thank two faculties in IMM. They are Associate Professor Bjarne Ersbøll and Assistant Professor Jakob Andreas Bærentzen who shared their knowledge in the variant parts of my project. Without them, I can not achieve it.

The last but not least thanks are to Kenneth Haugaard Møller for kindly providing me his Matlab ICP code, Sigurjon Arni Gudmundsson for good cooperation, Sune Darkner for solving my problems on VTK programming, Brian Lading for letting me use his face model and Kristian Evers Hansen from several discussions with whom I approached more to my destinations.

Contents

Abstract i

Preface iii

Acknowledgements v

I Introduction 1

1 The Technology of 3D Modelling from Scan 3

1.1 What This Technology Can Be Used for? 3

1.2 General Steps to Construct 3D Model from Scan 5

1.3 Scanning Method Classification 6

1.4 Application of the 3D Modelling from Scan 7

1.5 Outline of This Thesis 8

II Experiment Instrument 9

2 SwissRanger: A TOF Camera 11

2.1 TOF Technology 12

2.2 Camera Calibration 13

2.3 Camera Measurement Test 14

2.4 TOF vs. Stereo Vision 15

III Range Image Restoration 19

3 2D Image Restoration Methods 21

3.1 Single View Based Restoration 21

3.2 Multiple Views Based Super-resolution 23

3.3 Markov Random Field 26

4 Range Image Restoration via MRF 29

4.1 Forming MRF 30

4.2 Optimization Method – ICM 31

4.3 Restoring Range Images 31

4.4 Summery 33

IV Range Image Registration 35

5 Registration Overview 37

5.1 2D Geometric Transform 38
5.2 2D Image Registration Method 40
5.3 3D Rigid Transform 41
5.4 Iterative Closest Points Registration 42

6 ICP Implementation for Face Registration 45

6.1 Some Problems of Original ICP 45
6.2 ICP with Variants 46
6.3 Simulation Results 47
6.4 Summery 50

V Surface Reconstruction 55

7 From Point Cloud to Surface 57

7.1 Current Situation 57
7.2 Some Popular Algorithms 58
7.3 Other Possibilities 60

8 Surface Refinement by 2.5D MRF 61

8.1 Some Definitions 61
8.2 Mathematical Expression 63
8.3 Optimization by 2.5D ICM 64
8.4 Experiments 65
8.5 Summery 66

VI Results 71

9 Manikin Face Modelling and Super-Resolution 73

9.1 Data Acquisition 73

9.2 Face Extraction 74

9.3 Depth Registration via ICP 75

9.4 Face Modelling 75

9.5 Face Super-Resolution 77

9.6 Summery 77

VII End Remarks 79

10 Future Work and Conclusions 81

10.1 Camera Test 81

10.2 Range Image Restoration 82

10.3 Registration Method 83

10.4 Surface Reconstruction 83

10.5 System Consideration 84

A Compute 3D Rigid Transform from Correspondents 85

A.1 Unit Quaternion 85

A.2 SVD 86

B Interpolating Points in Triangles 87

B.1 Is This a Triangle? 87

B.2 Is the Point in the Triangle? 88

B.3 What Is the Weight Should Be? 89

C Sample Depth and Intensity Images Used in Experiment 91

Part I

Introduction

Chapter 1

The Technology of 3D Modelling from Scan

This chapter is to introduce 3D modelling from scan technology. The usage and the steps of this technology, the scan method classification and the applications are illustrated respectively.

1.1 What This Technology Can Be Used for?

The techniques and hardware for scanning the surface of real object is developing very fast in the recent years, which brings flourish in the related research areas. One of the most prospective areas is 3D surface modelling. Especially, loyally construct the computer readable shape data of object in reality challenges and attracts a large number of people working in the Computer Vision area. Scanning can provide the most direct and automatic measurement of the object's surface. The 3D modelling from scan proves to be very helpful in the following fields:

- **Computer Animation**
  The 3D animation movie has created billions of US dollars' market. But

the budget of such movie is also enormous. A big part of it is to simulate the details of every thing in the scene to let them look real (See Figure 1.1) which is always the nightmares for the computer engineers.

Figure 1.1: A computer created scene full of models.

- **Digital Archiving**
  Doing many experiments may damage the precious fragile relics. The digital archiving aims at moving most of the researches and studies on accurate models of the relics to a computer. Of course, the digital model has infinite durability. Some created digital relics can be seen in Figure 1.2.

Figure 1.2: Some examples of digital relics. The figure is from [10]

- **Quality Control**
  Quality control is important in industry. Sometimes, the flaw of the product is hardly perceivable by eyes, but automatic detection from a 3D digital model is feasible for a computer (See Figure 1.3a).
- **Medical Diagnosis**
  The doctors tend to like watching 3D organs more now. It is reasonable to display real 3D rather than the 2D slices when doing disease diagnosis or surgery plan (Figure 1.3b).
- **Home Entertainment**
  Just by imagining how many PCs in the world and how favorite the 2D digital cameras are, one can conclude that the home entertainment is maybe the most exciting application of this technology. In this area, demanding

a b

Figure 1.3: (a)An computer model for quality control. (b)An 3D cardiac model created by a 3D CT.

a cheaper, faster and smaller equipment is always the first thing. There are some trials, for example, structured light method using a video camera and a projector, Stereo Vision using multiple 2D cameras, but none of them full fills the requirements. Is it possible to build all instruments in just one small low energy cost camera so that it could work as one of the standard computer Accessories? This question interests the researchers and manufacturers, who are making some progress and keeping this field very active.

Depending on the devices, the 3D modelling from scan would be an accurate and efficient tool for the above areas. It is bringing a revolution to the traditional methods.

1.2 General Steps to Construct 3D Model from Scan

To construct a 3D model in the computer, the following steps are generally needed:

1. **Data Acquisition**
   The range information of a real 3D object input into a computer is normally positions of a point set sampled from the surface. A position detection machine can be a laser scanner, an optical camera or some position sensors.
   Under high resolution, cost insensitive case (e.g. some large medical and

industrial CT system) the positions of source and sensor is fixed so that the position of the detected event can be precisely determined. This method is less likely used when a certain amount of flexibility is required. In that situation, a free hand system is preferred.
To obtain a freehand system, like most of the modern medical ultrasound imaging system, the movement of the source/sensor must be detect. To achieve this, one can assume constant velocity of the probe, or use a position sensor, but the most flexible solution is to use image registration and motion estimation method.

2. **Registration**
As stated above, the registration may provide flexible scanning solution. Normally, the range detection machine can only get part of the details of the object. The positions of these pieces of information should be calculated so as to be integrated together to get a whole view the user desired. The global or local position identification is achieved by registration. The registration can be done manually, where the user should tell the system how the two images are exactly related; or automatically, where the system will determine the relationship by itself; or semi-automatically, where the user help the system find the best relation.

3. **Merging**
After registration, the whole model can be patched up by the pieces of scans. The redundant measurement should be removed so that the whole data set will not infinitely increase during continuously scanning. Moreover, the deviated measurement should be deleted as outliers. It could be also plausible to use extra data to produce a super-resolution of the model.

1.3 Scanning Method Classification

The scanning method can be divided into passive and active acquisition [21].

1.3.1 Passive Acquisition

The passive acquisition refers to the sensor which detects the light reflection from the object of ambient light source in passive acquisition mode. Shape-From-Shading[23] and Stereo Vision[17] are two famous passive acquisition methods. The passive acquisition usually suffers by sparse and inaccurate data compared with the active acquisition, so it is hard to be utilized in the high requirement conditions.

1.3.2 Active Acquisition

The active acquisition system generally has its own source. The contact active system use probe sensor like Coordinate Measuring Machine (CMM). The non-contact system emits a serial of signals and the signal is influenced by the target. The influenced signal is then received by the sensor to calculate the target's shape. Some of the successful scanning systems include:

- **Transmit/Emit Computerized Tomography:** X-ray CT (*Computerized Tomography*), PET (*Positron Emission Tomography*) are well-known medical instruments.
- **Active Stereo Vision:** In stead of capturing the ambient light, this system can project a light pattern onto the object to help finding the correspondents.
- **Structured Light:** The structured light methods observe the illumination of the light pattern projected by the system on the object and calculates the shape accordingly.
- **Radar System:** Large system is like SAR (*Synthesis Aperture Radar*). For the common application, the laser is more popular because it is very accurate and with high discrimination in a limited range. Furthermore, the laser instrument becomes smaller and cheaper.
  If the system emits pulses and calculate the time interval between the emission and reception of these pulses, the technology is called *time-of-flight* (TOF). Some of the radar systems are based on TOF technology.

1.4 Application of the 3D Modelling from Scan

Although the 3D scanning and modelling system has not been developed for long, there are some industrial and academic successes:

- The motion capture using contact probe is widely used in computer animation and games creation. More complicated product includes TRITOP developed by GOM.

- 3D CT/ultrasound medical imaging systems are more acceptable by the doctors to improve the diagnosis.
- Passive and active Stereo Vision have been researched for many years, this human-eye-like system is simple and can produce very high resolution images due to the maturity of the 2D camera manufacture. Point Grey's Bumblebee serial is an example of this method in industrial quality control.
- Large Laser equipment is very suitable for some accurate measurement conditions. The high-resolution laser scanned faces are adopted by some of the 3D face database. "Michelangelo Project"[10] which aims at digital archiving the Michelangelo's large statues is also a great attempt for laser scan and modelling. The problems of these systems are that the scanning process is usually long and the system is quite expensive.
- The newly invented range camera SwissRanger and CanestaVision cameras are fast, small and low energy consumption equipments. These cameras are very promising to construct a potable real-time 3D modelling system. However, the accuracy and SNR (*Signal to Noise Ratio*) of these cameras still need to be improved. More deep research should be involved in the future.

1.5 Outline of This Thesis

In the next part of this thesis, I will tell something about my experiment instrument – SwissRanger. Then, the core parts of this thesis follow, which are Range Image Restoration, Range Image Registration and Surface Reconstruction. In these parts, a general knowledge and related work is first introduced, then the method I use and experiment result are illustrated. The last part of this thesis is the outcome for real manikin face modelling and super-resolution, and the end remarks for future work and conclusion will finish this thesis.

Most of the 2D figures in the this thesis are created by Matlab, and most of the 3D figures are by VTK. The registration is programmed with VTK C++ library 5.0, and the restoration and surface reconstruction are done in Matlab 7.0.

Part II

Experiment Instrument

Chapter 2

SwissRanger: A TOF Camera

The SwissRanger is a TOF technology based range detection camera developed by CSEM Zurich Center. The newly invented model SwissRanger SR-3000 has tiny size (see Figure 2.1) and high frame rate up to 50 fps, which is designed to provide a personal use, free hand scanning method.

Figure 2.1: The SwissRanger SR-3000

The SwissRanger emits a sinusoidally modulated light wave. The reflected wave front is received and sampled, thus the distance of the object can be measured. Each time, two images are produced by SwissRanger. One is the calculated depth information image, the other is the reflected light intensity information image. Both of them are in QCIF format (176×144). The intensity image has higher contrast to distance but it is also highly affected by the material and structure of the object.

The details of the SwissRanger SR-3000 will be described in this chapter. The questions of how this camera measures depth and intensity, how to convert the measurement into Cartesian coordinate, how the integration time affects the output images, and what the advantages and disadvantages are compared with stereo vision will be answered.

2.1 TOF Technology

The time of flight technology counts the light traveling time between emission and reception. The measured time is twice of the distance to travel for light speed.

Instead of directly measuring the time, the SwissRanger uses the modulated infra-red light, and calculate the phase shift.

Assume the emitted signal is

$$e(t) = 1 + cos(wt) \tag{2.1}$$

and the detected signal is

$$s(t) = 2\beta(1 + \cos(wt - \varphi)) \tag{2.2}$$

where t is time, w is the modulated frequency, φ is the phase delay and β is the attenuation factor of the signal.

The modulated signal is demodulated by $g(t) = cos(wt)$, so that the demodulated sample at time τ is

$$c(\tau) = s(t) \otimes g(t)|_{t=\tau} = lim_{n\to\infty}\frac{1}{T}\int_{-\tau/2}^{\tau/2} s(t) \cdot g(t+\tau)dt \tag{2.3}$$

If select the sample point p with 90° phase difference, for example, $\tau_0 = 0°$, $\tau_1 = 90°$, $\tau_2 = 180°$, $\tau_3 = 270°$, then the phase delay

$$\varphi = \arctan(\frac{p(\tau_3) - p(\tau_1)}{p(\tau_0) - p(\tau_2)}) \tag{2.4}$$

The amplitude

$$A = \frac{\sqrt{[p(\tau_3) - p(\tau_1)]^2 + [p(\tau_0) - p(\tau_2)]^2}}{2} \tag{2.5}$$

The offset

$$B = \frac{p(\tau_0) + p(\tau_1) + p(\tau_2) + p(\tau_3)}{4} \tag{2.6}$$

The physical meaning of φ, A and B can be seen from Figure 2.2.

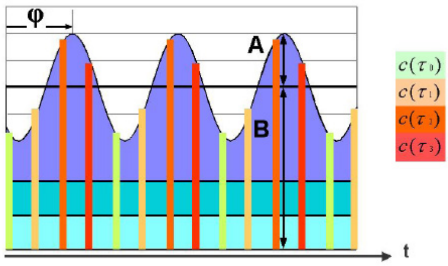

Figure 2.2: The sample process and physical meaning of φ, A and B. The figure is from [1]

The phase delay φ is directly proportional to the distance D:

$$D = \frac{c}{2f_m} \cdot \frac{\varphi}{2\pi} \tag{2.7}$$

where f_m is the modulation frequency and c is the speed of light. From equation 2.7, we can see the theoretic distance limitation for SwissRanger is $0 \sim \frac{c}{2f_m}$, which is approximately $0 \sim 15$m for 20MHz modulation frequency.

For more details of the SwissRanger imaging principle and physical structure, one can refer to [27].

2.2 Camera Calibration

In SwissRanger camera, the CCD cells are equally spaced. Each cell measures the distance from the point on the target to the focus of the camera. The measured distances can be easily transformed to 3D positions in Cartesian coordinate. From Figure 2.3 the measurement D from a CCD cell with horizontal angle θ_x can be transformed to the 2D Cartesian coordinate (x, z) by:

$$x = D\sin(\theta_x) \quad \text{and} \quad z = D\cos(\theta_x) \tag{2.8}$$

and we have the similar result for the 3D:

$$x = Dsin(\theta_x) \quad y = Dsin(\theta_y) \quad z = D\cos(\theta_{xy}) \tag{2.9}$$

where the θ_y is the vertical angle and θ_{xy} is the 3D angle of the CCD cell.

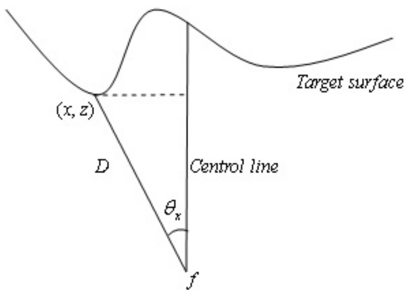

Figure 2.3: Illustration of 2D Cartesian coordinate transform, f is the focus, the measurement D from a CCD cell with horizontal angle θ_x can be transformed to Cartesian coordinate (x, z).

Since θ_x, θ_y and θ_{xy} is fixed for each cell, the transform can be done by multiplying the measured depth image with constant correction factor images for x, y and z (see Figure 2.4), so that the conversion is very fast.

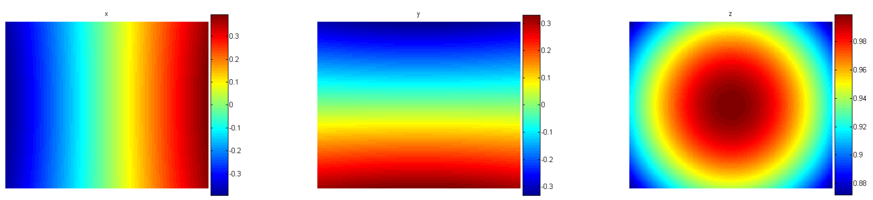

Figure 2.4: From left to right, the correction factor images for x, y and z

2.3 Camera Measurement Test

Like other digital cameras, the image quality produced by SwissRanger is also affected by the size of the CCD cell and the noise which plus the accuracy of the range measurement. They determine the output image resolution.

The iid (*independent and identically distributed*) noise can be decreased by averaging several images of the same measurement, which can be also controlled by changing the integration time of the camera. However too long time will cause motion blur and data overflow. Figure 2.5 shows the depth images of a manikin face captured by integration time $1ms$, $4ms$ and $8ms$. The cold color represents near the camera focus, while the warm color represents far from the camera focus. For $1ms$ integration time, the noise is prominent in the image. For $8ms$ integration time, we can see the wrong measurements on the manikin face, which are caused by the data overflow. This experiment shows $5ms$ around is suitable for measuring an object in the range about $0.5m$.

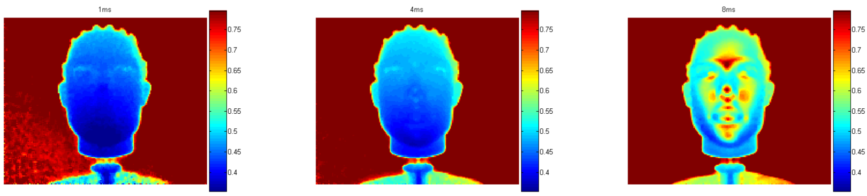

Figure 2.5: From left to right, the integration time is with 1, 4 and $8ms$. The noise is prominent for $1ms$, the wrong measurement appears for $8ms$

Because the CCD cell doesn't have infinite small size, the output of each cell is the average of the signal within the cell. After interpolation, the image is blurred in space. The blurring effect can be described by convoluting a psf (*point spread function*) to the desired sharp image. To estimate this function, an experiment similar with the one used for 2D camera [6] is implemented. An edge between two planes with different distances becomes the measured object in the depth image. The image is shown in Figure 2.6.

In the idea case the line crossing the edge should be a step signal. Assuming the psf is Gaussian function, the standard deviation of this function can be estimated by convoluting with the ideal step signal and comparing with the measured points. Figure 2.7 shows the measured points and Gaussian convoluted curve. The result shows that 0.4 pixel is a good approximation of the standard deviation.

2.4 TOF vs. Stereo Vision

Stereo Vision especially for two photo cameras is another popular way to detect range and has been studied for a long time. Like TOF camera, some Stereo

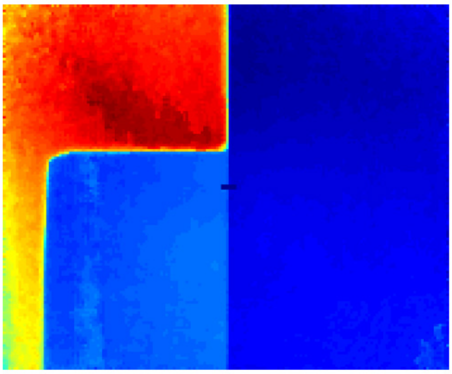

Figure 2.6: The depth image of an edge. The points on the dark line on the center of the image is used for psf estimation.

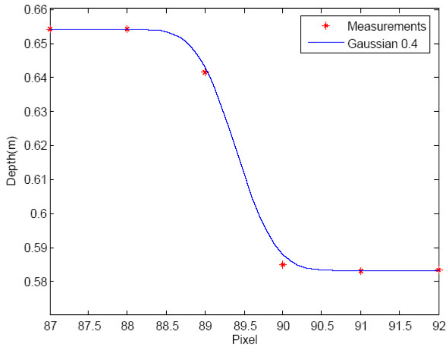

Figure 2.7: The psf measurement. The original points are red stars, the blue curve is a Gaussian psf with standard deviation 0.4 pixel which is a good fit.

Vision industry products are also available. A comparison between these two systems are listed below. Similar comparison can be found in CSEM's technical report[20].

Table 2.1: A comparison between TOF and 2 camera Stereo Vision 3D systems

	Stereo Vision	SwissRanger
Portability	Two photo cameras is needed and they should be placed by a distance to ensure the resolution. An additional illumination source is maybe needed.	The camera with integrated illumination source has comparable size with a normal photo camera.
Computation	Need to search for corespondents usually computationally heavy even for a modern PC.	Phase and intensity calculation are very simple which can be directly integrated onto silicon. Maximum 50 fps is possible.
Accuracy	Sub-millimeter depth resolution can be achieved if high contrast images are available, but may fail when detecting uniform scenes.	Has no problem for uniform scenes but may affected by the reflection angle, material and colors, generally subcentimeter can be provided.
Price	Depending on what quality the cameras are, could be very cheap for family or expensive for industry applications.	Expensive, but cheaper than a high resolution laser scanner. €5000 is for the prototype in IMM. However, the relatively simple structure makes the price have very good potential to reduce after mass-produced.

Part III

Range Image Restoration

Chapter 3

2D Image Restoration Methods

The depth image captured by SwissRanger is low-resolution, and suffered by error measurement as well as noise. Although the target object is 3D, the single depth image is 2D. So, some of the well-developed 2D image restoration methods may have their stages now. In this chapter, the restoration methods based on just one image is first generalized. Then, when multiple views are available, the situation of resulting in a super-resolution image is described.

3.1 Single View Based Restoration

Single image restoration has been researched for a long time. The theory and methods can be found in most fundamental image processing textbooks.

3.1.1 Image Model

In single view image restoration, the degradation process can be simply modelled by a convolution process in spatial domain.

$$\tilde{g}(i,j) = \tilde{h}(i,j) * \tilde{f}(i,j) + \tilde{\eta}(i,j) \tag{3.1}$$

where $\tilde{g}(i,j)$, $\tilde{h}(i,j)$ and $\tilde{\eta}(i,j)$ are the 2D degraded image, spatial degradation function and noise respectively. $*$ means the spatial convolution.

The convolution in Equation 3.1 can be also written as multiplying a matrix H:

$$g = Hf + \eta \tag{3.2}$$

In Equation 3.2, g, f and η are vectors generalized by serializing $\tilde{g}$, $\tilde{f}$ and $\tilde{\eta}$.

3.1.2 Noise Removal Approaches

If the noise is iid, the simple *mean* or *order-statistics filter* (for example *median filter*) usually has good performance. *Adaptive filter* is a more complicated filter, which may change the parameter so as to change the filtering property according to the statistical measure of a local area.

Bandreject filters like *bandpass filters* or *notch filters* are the classic methods used in frequency domain.

3.1.3 Filters Counteracting the Degradation Function

To counteract the effect of degradation function, the most direct approach is using the inverse of the degradation function, so called the *inverse filtering*. The inverse filtering may amplify the noise since it does not take it into consideration. The *Wiener filter* can adjust the inverse process according to the statistical characteristics of noise. Both of these famous approaches can be generalized in one form, so-called *geometric mean filter* [12].

$$\hat{F}(u,v) = \left[\frac{H^*(u,v)}{|H(u,v)|^2}\right]^{\alpha} \left[\frac{H^*(u,v)}{|H(u,v)|^2 + \beta[\frac{S_\eta(u,v)}{S_f(u,v)}]}\right]^{1-\alpha} G(u,v) \tag{3.3}$$

In this equation:

$$\begin{aligned}
\hat{F}(u,v) &= \text{Fourier transform of the estimated undegraded image} \\
G(u,v) &= \text{Fourier transform of } \tilde{g}(i,j) \\
H(u,v) &= \text{Fourier transform of } \tilde{h}(i,j) \\
H^*(u,v) &= \text{complex conjugate of } H(u,v) \\
S_\eta(u,v) &= |N(u,v)|^2 = \text{ power spectrum of the noise} \\
S_f(u,v) &= |F(u,v)|^2 = \text{ power spectrum of the undegraded image}
\end{aligned}$$

and the α and β being positive, real constants which control the property of this filter. Some of the common settings of these parameters are:

$$\alpha = \begin{cases} 1, & \text{the inverse fiter} \\ 0, & \text{parametric Wiener filter} \\ 0 \text{ and } \beta = 1, & \text{standard Wiener filter} \\ \frac{1}{2} \text{ and } \beta = 1, & \text{spectrum equalization filter} \end{cases}$$

3.2 Multiple Views Based Super-resolution

Besides restoration from single image, computer vision researchers began to focus on multiple view based restoration. These views, related by geometric transforms, provide more than one measurements of the target. If the transforms are known, it is possible to use these extra measurements to restore an image with higher resolution than any of the single one. This technique which is called *super-resolution* [19], can surpass the limitation of the hardware and thus is very promising. Super-resolution for the 2D photos is well investigated [6] which will be summarized in this section.

3.2.1 Image Model

Suppose the $n \times n$ high-resolution image is serialized as a $nn \times 1$ vector $\bar{f}$ and the serialized $m \times m$ observed low-resolution image is $mm \times 1$ vector g ($m < n$). The observed image is affected by the position of the camera which is modelled by a $nn \times nn$ geometric transform matrix T; the spatial blurring which is modelled by a $nn \times nn$ convolution kernel matrix H; the limited number of pixels which is modelled by an $mm \times nn$ decimation operator matrix D and the noise is modelled by an $mm \times 1$ additive vector term η. So that the mathematical expression between $\bar{f}$ and one of the observed image g_n is

$$g_n = M_n \bar{f} + \eta_n \tag{3.4}$$

were $M_n = D_n H_n T_n$ which is a $mm \times nn$ matrix.

3.2.2 Interpolation Method

The interpolation method is simplest and straightforward for multiple-view super-resolution. In this method, the repeated measurements are interpolated to recover the wanted point grid. One of the usages is to calculate the *average image*. To get an average image, all the points in the observed images are transformed back to the high-resolution grid. Since this will result that each grid point may have multiple values around, a smooth kernel is used to compute a weighted average that can reduce the noise.

$$f_i^{avg} = \frac{\sum_j w_j g_j}{\sum_j w_j} \tag{3.5}$$

where w_j is the weight which relates the transformed measurement g_j to f_i and its value determined by the smooth kernel.

If choosing the smooth kernel the same as the psf function, the mathematical expression of the interpolation method is

$$f_{avg} = k^{-1} M^T g \tag{3.6}$$

where k is a diagonal matrix and $k_{ii} = \sum_j M_{ji}$ is the sum of the columns of M. From Equation 3.6 we can see that the average image can not result in a shaper image because M^T has the same blurring effect as the psf. Experiment shows the average image is robust to the noise and can be used as a good initial of other super-resolution algorithm[6].

3.2.3 Maximization Method

The exact solution of the problem is to calculate the inverse of the M matrix, so that

$$f = M^{-1} g \tag{3.7}$$

but normally this inverse is difficult or impossible to be calculate due to the existence of noise and ill-posed situation of the matrix.

Instead, one can try to maximize the likelihood of an observed low-resolution image g_n given by the estimated high-resolution image $P(g_n|\hat{f})$. If assume the noise is Gaussian distributed with variance σ^2, the likelihood function can be written as

$$P(g_n|\hat{f}) = \prod_{\forall x,y} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\hat{g}_n(x,y) - g_n(x,y))^2}{2\sigma^2}\right) \tag{3.8}$$

where $\hat{g}_n = M_n\hat{f}$

The log likelihood of 3.8 is

$$\ell(g_n) = -\sum_{\forall x,y}(\hat{g}_n(x,y) - g_n(x,y))^2 = -\|\hat{g}_n - g_n\|^2 \tag{3.9}$$

the maximum of $\sum_n \ell(g_n)$ is found when the first derivative is zero, so that

$$\hat{f}_{mle} = (M^T M)^{-1} M^T g \tag{3.10}$$

Since $M^T M$ is sparse and diagonal symmetric, it is efficient to use conjugate gradient algorithm[26] to find the solution.

3.2.4 Bayesian Method

The maximum likelihood method stated above is to find the maximum of the likelihood function $P(g_n|\hat{f})$. Whereas if we know some information of $\hat{f}$, it would be helpful to find a more desired solution. The way to do that is using Bayesian law to maximize the posterior function.

$$P(\hat{f}|g) = \frac{P(g|\hat{f})P(\hat{f})}{P(g)} \tag{3.11}$$

where $P(g|\hat{f})$ is our likelihood function as before, $P(\hat{f})$ is the prior function which cooperates the prior knowledge of the $\hat{f}$, $P(g)$ is the scaling factor which can be omitted during calculation.

Compared with the maximum likelihood method, the only changes for the Bayesian method is the additional prior term in 3.11. The construction of this term determines how well the Bayesian method can perform. The most popular mathematical model used to describe the prior information of the desired image is MRF (*Markov Random Field*) which will be described below.

3.3 Markov Random Field

The MRF assumes that an individual pixel in an image is only affected by a subset or clique of the pixels in the image. If these pixels are directional organized, the relationship can be expressed by a N dimensional Markov chain, where N is the number of orientations used in the subset.

$$P(f_i|f_{j\neq i}) = P(f_i|f_k, k \in \mathcal{N}_i) \tag{3.12}$$

Here, $\mathcal{N}_i$ is the subset, normally chosen as the adjacent spatial neighborhood of the pixel f_i and usually invariable homogenous meaning that all the pixels have the same neighborhood structure regardless of the positions of these pixels. For simplicity, 4 or 8 adjacent neighbor structure (Figure 3.1) is utilized in 2D image analysis.

Figure 3.1: A point with four and eight neighbors.

3.3.1 Gibbs Random Fields

A *Gibbs distribution* takes the form

$$P(f) = \frac{1}{Z} \exp\,(-\frac{1}{T} \sum_{\forall C \in \mathcal{C}} V_C(f)) \tag{3.13}$$

where Z is a normalize factor called *partition function*, T is a constant called the *temperature* which is normally assumed to be 1, $U(f) = \sum_{\forall C \in \mathcal{C}} V_C(f)$ is the *energy function*, C is one of the pixel clique, $\mathcal{C}$ is all the cliques in the image. V_C is called *potential function* which defines how the pixels are related. The potential function is often chosen as pair-pixel related, meaning that $V_C(f_i) = V(f_i, f_j)$.

If all the f obey Gibbs distribution then they are called *Gibbs Random Field* (GRF).

Hammersley-Clifford theorem states that *F is an MRF on S with respect to N if and only if F is a GRF on S with respect to N*. One of the proof can be referred in [16].

The MRF describes the local property while the GRF focuses on the global. Both on them can be used to provide an easy interpretation of the prior.

3.3.2 Some Common Priors

The best prior can contain as much accurate information of the desired result as possible. So designing a prior which is suitable to the current problem is the best choice. However, sometimes designing a specific prior is not that easy. Some of the priors are proved to be useful for most of the circumstances and being widely accepted. Some of them are:

- **Gaussian Model**
  When the Gibbs distribution is a multivariate Gaussian function, the Prior is called *Gaussian MRF* (GMRF). In this model

  $$\begin{aligned} V(C_x) &= \gamma d_x^2, & V(C_y) &= \gamma d_y^2 \\ V(C_{xy}) &= \gamma d_{xy}^2, & V(C_{yx}) &= \gamma d_{yx}^2 \end{aligned}$$

  where γ is a constant controlling the smoothness. For the first derivative

  $$\begin{aligned} d_x &= f_{x+1,y} - f_{x,y}, & d_y &= f_{x,y+1} - f_{x,y} \\ d_{xy} &= \tfrac{1}{\sqrt{2}}(f_{x+1,y+1} - f_{x,y}), & d_{yx} &= \frac{1}{\sqrt{2}}(f_{x+1,y-1} - f_{x,y}). \end{aligned}$$

  The second derivative of this model could be

  $$\begin{aligned} d_x^2 &= f_{x-1,y} - f_{x,y} + f_{x+1,y}, & d_y^2 &= f_{x,y-1} - f_{x,y} + f_{x,y+1} \\ d_{xy}^2 &= \tfrac{1}{2} f_{x-1,y-1} - f_{x,y} + \tfrac{1}{2} f_{x+1,y+1}, & d_{yx}^2 &= \frac{1}{2} f_{x-1,y+1} - f_{x,y} + \frac{1}{2} f_{x+1,y-1}. \end{aligned}$$

  The GMRF encourages a smoother result.

- **Generalized Gaussian Model**
  The *Generalized Gaussian MRF* (GGMRF) introduces a factor p in the GMRF to make it more flexible. The Gibbs distribution of GGMRF is

  $$P(x) = \frac{1}{Z} \exp\left(-\frac{x^p}{p\sigma^p}\right) \tag{3.14}$$

  GGMRF has heavier tail than GMRF when $1 < p < 2$.

- **Huber Model**
  The *Huber MRF* (HMRF) uses Huber function so that the Gibbs distribution has the form:

$$P(x) = \begin{cases} \frac{1}{Z}\exp\left(-\gamma x^2\right) & \text{if } |x| < \alpha \\ \frac{1}{Z}\exp\left(-\gamma(2\alpha|x| + \alpha^2)\right) & \text{otherwise} \end{cases} \tag{3.15}$$

  HMRF encourages Gaussian smoothness when the pixels have the difference within $|\alpha|$, whereas less penalty for the large difference so as to preserve the edges. So that, the HMRF is also called Edge-preserving MRF.

- **Auto-Model**
  The *Auto-MRF* (AMRF) provides more flexibilities to control the smoothness along any specific direction.

$$P(f_i|f_{N_i}) = \frac{1}{Z}\exp\left(f_i G_i(f_i) + \sum_{i' \in N_i} \beta_{i'} f_i f_{i'}\right) \tag{3.16}$$

  Where G_i is an arbitrary function and the orientation smoothness is controlled by parameter $\beta_{i'}$. If $f_i \in \{0, 1\}$ or $f_i \in \{-1, 1\}$ the auto model is said to be an *auto-logistic* model. Furthermore, if four-neighborhood structure in Figure 3.1 is selected, the model is reduced to the *Ising Model*.

3.3.3 Model Optimization Method

The optimal pixel value is found by maximizing the posterior function 3.11. If the function has the quadratic form, it would be efficient to use *Conjugate Gradient Ascent* method[26]. Otherwise, a general gradient ascent method may be used. It could be also possible to use *Iterative Constrained Modes* (ICM)[3] or *Simulated Annealing* (SA) method.

Chapter 4

Range Image Restoration via MRF

Due to the limitation of the hardware, the depth image is relatively low resolution and noisy. In another hand, the intensity image is with high contrast and contains some information of depth (See Figure 4.1). Furthermore, the Swiss-

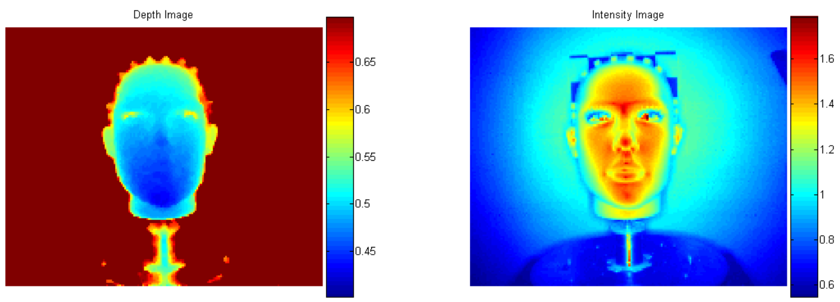

Figure 4.1: The depth (left) and intensity (right) images. The values have been scaled.

Ranger has very fast frame rate. It is possible to acquire multiple images at a very short time interval. Both of the situations can be utilized to increase the resolution of the single depth image.

There are very few researches on multiple view super-resolution of depth image because although we know the 3D objects are related by 3D rigid transform and they are projected onto the 2D plane, it is very difficult to find a transform between two 2D depth images. That will prevent us to write a formula like 3.4.

In reference [9], the authors proposed a method that utilized a normal high resolution photo as prior to make a single super-resolution depth image. The method is based on designing a new MRF. In this chapter, I will show the structure of this MRF and how to use this structure to restore our depth image.

4.1 Forming MRF

The idea in reference [9] is to exploit the information in a high resolution image to restore low resolution depth image. The assumption of this high resolution image is that the depth discontinuities will also be reflected in this image. The authors use the conventional 2D camera photos because the depth difference may bring the brightness to change. The log likelihood function called *depth measurement potential* is:

$$\Psi = \sum_{i \in L} k(y_i - z_i)^2 \tag{4.1}$$

where y is the restored image that we want to estimate, z is the original depth measurement, k is a positive constant weight, L is all the depth measurements.

The log *depth smoothness prior* is of the form

$$\Phi = \sum_i \sum_{j \in N(i)} w_{ij}(y_i - y_j)^2 \tag{4.2}$$

here N is the neighbor clique and w is the weight connecting the high resolution information:

$$w_{ij} = \exp(-cu_{ij}) \tag{4.3}$$

c is a positive constant and

$$u_{ij} = \|x_i - x_j\|_2^2 \tag{4.4}$$

here x is the high resolution image point, so that the small difference of x will result in large w and smooth estimation. Whereas small w will decrease the functionality of the prior.

Then the normalized posterior probability is

$$p(y|x, z) = \frac{1}{Z}\exp(-\frac{1}{2}(\Psi + \Phi)) \tag{4.5}$$

4.2 Optimization Method – ICM

The maximum of equation 4.5 can be optimized by a general gradient descend method. Furthermore, the quadratic form of ψ 4.1 and ϕ 4.2 makes the Conjugate Gradient descend be an efficient choice. In my experiments, I use another simple optimization method – ICM [3].

In order to use ICM, each time the pixel value is only determined by its four neighbor pixels (Figure 3.1). This is shown in Figure 4.2 where the red and green pixels are neighbors between each other.

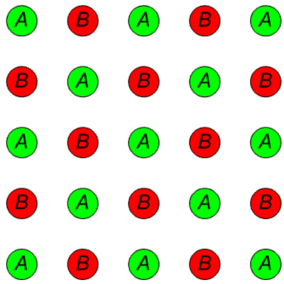

Figure 4.2: An illustration of updating rule of ICM

In each iteration, there are two passes. The first pass is fixing all the red pixels and updating the green pixels according to the red pixel value. The second pass is vice versa, fixing the green pixels and updating the red ones. The updating rule for my application is:

$$\hat{y}_i = \frac{kz_i + \sum_{j \in N(i)} w_{ij} y_j}{k + \sum_{j \in N(i)} w_{ij}} \tag{4.6}$$

4.3 Restoring Range Images

In order to use formulae 4.1 $\sim$ 4.5, a high resolution information is needed. No doubt, the intensity image is a good candidate. The intensity image has sharp jumps at depth discontinuities and low noise at smooth surfaces. Most important, it is produced with depth image as a byproduct, no other equipment is required.

Figure 4.3 shows how the constant parameter k and c influence the result. If k

and c are large the result image has almost no difference with the depth image. On the contrary, the small k and c will result in over-smoothing.

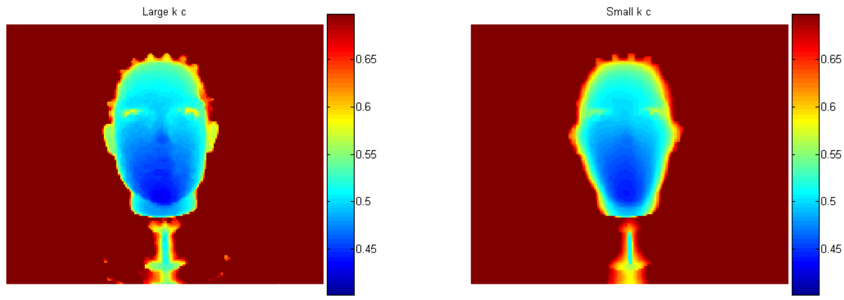

Figure 4.3: The depth (left) and intensity (right) images. The values have been scaled.

Figure 4.4 shows the restored depth image using suitable parameters. The blowups around the nose and top part of the head before and after restoration tell the improvement. The edge becomes sharper and the noise is depressed after restoration.

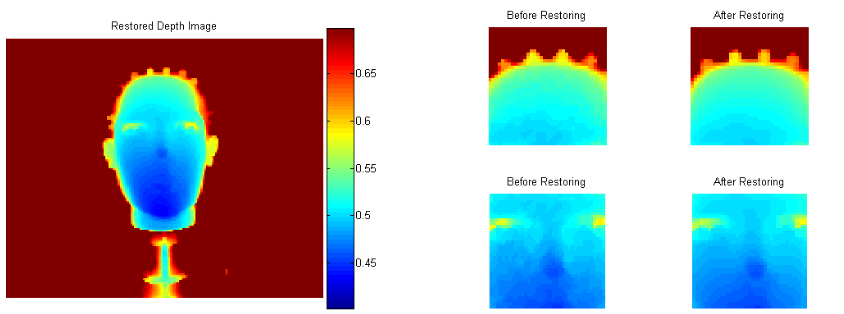

Figure 4.4: The depth image restored with suitable parameters (left). The blowups (right) of the top and central part of the head indicate that the edge becomes sharper and the noise is depressed after restoration.

The noise reduction is more obvious when displaying in 3D. Figure 4.5 is the restorations of three faces.

The plot of iteration vs. sum of absolute changes of restoring Figure 4.4 is

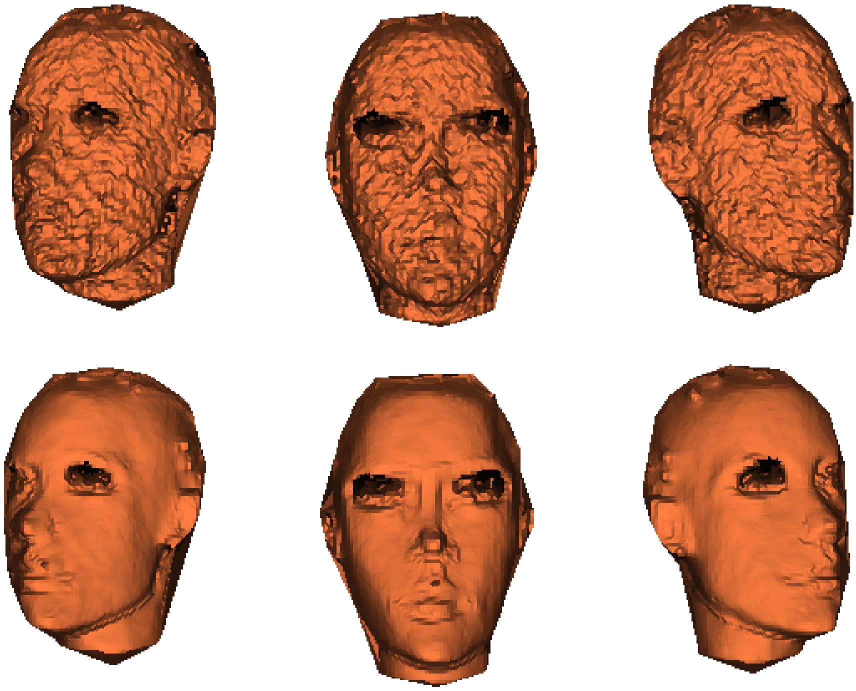

Figure 4.5: The 3D display of the restoration result. The first row is original depth measurements and the second row is the restored measurements.

plotted in Figure 4.6. From this curve, we can conclude that the ICM converges very fast. A bit over 20 iterations seems enough, which is not a problem for the image having QCIF (176×144) format.

4.4 Summery

An experiment of restoring range images is successfully implemented here. The restoration is based on MRF. The prior knowledge is selected from the corresponding intensity images. This low cost solution is proved to be very suitable for the application.

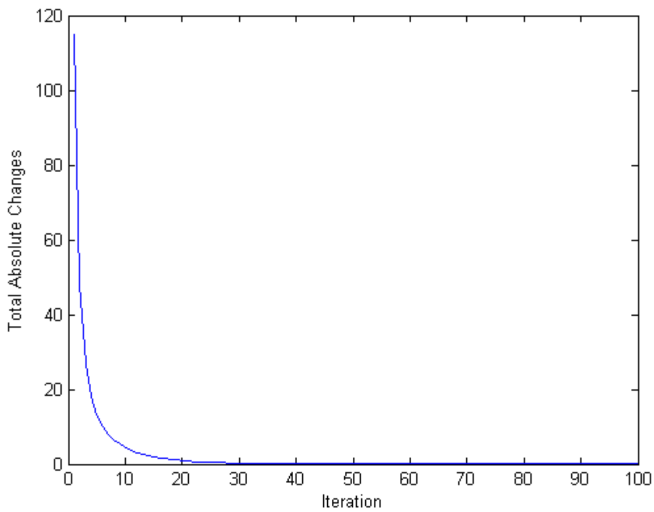

Figure 4.6: The sum of absolute changes for all the iterations when restoring Figure 4.4. The plot shows that the ICM converges very fast.

Part IV

Range Image Registration

Chapter 5

Registration Overview

"*Registration is the determination of a geometrical transformation that aligns points in one view of an object with corresponding points in another view of that object or another object.*"[2]. Registration is widely used in medical image analysis and computer vision areas. The relationship among different poses of the object is described by geometric transforms. The geometrical transformation can map points from space F $((\mathbf{x}, \mathbf{y}, ...)')$ of one view to points from space $\mathbf{V}$ of a second view. The transformation of points in $\mathbf{F}$ written as $1 \times n$ vectors can be expressed by a transformation operator τ :

$$\mathbf{G} = \tau(\mathbf{F}) \tag{5.1}$$

The result $\mathbf{G}$ is the estimation of $\mathbf{V}$.

If the object is rigid, the registration is named rigid registration and the motion of the object can be expressed by a rigid transform. If the object is deformable, the registration should apply nonrigid transform.

In this chapter, some 2D geometric transform and registration methods are first introduced, which is believed to be heuristic to higher dimensions. Then, the mathematical expression of 3D rigid transform is shown followed by the 3D rigid object registration algorithm – ICP.

5.1 2D Geometric Transform

2D image like a painting, a photo, a CT slide is the most common form to represent object in life. The concept of the 2D geometric transform can be easily extended to higher dimensions. Some of the frequently used geometric transforms for 2D include:

5.1.1 Rigid Transformation

The *rigid transform* preserves all distances, straightness and parallelism of lines and all nonzero angles. The rigid transform only allows rotation and translation (see Figure 5.1).

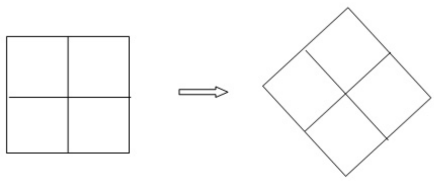

Figure 5.1: The rigid transform

The operator can be expressed as:

$$\mathbf{G} = R\mathbf{F} + \mathbf{b} \tag{5.2}$$

where R is a $n \times n$ orthogonal matrix, meaning that $R^t R = RR^t = I$. Thus $R^{-1} = R^t$. If forcing $det(R) = +1$, the reflection is not allowed.

5.1.2 Nonrigid Transformation

The *nonrigid transformation* can compensate nonrigid shape distortions in the image. Each transformation has its physical meaning and different complexity of the mathematical property.

- Scaling transformation
  The *scaling transform* is similar to the rigid transform, except it allows scaling (see Figure 5.2). It is useful to describe unchanged shapes in

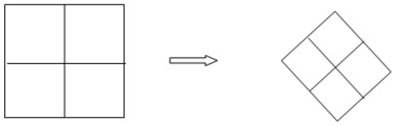

Figure 5.2: The scaling transform

different scenes. The transform equation is:

$$\mathbf{G} = RS\mathbf{F} + \mathbf{b} \tag{5.3}$$

or

$$\mathbf{G} = SR\mathbf{F} + \mathbf{b} \tag{5.4}$$

where S is a $n \times n$ diagonal scaling factor matrix. If S is isotropic, the transformation is called *similarity transform*.

$$\mathbf{G} = sR\mathbf{F} + \mathbf{b} \tag{5.5}$$

- Affine transformation
  The *Affine transform* preserves the straightness and parallelism of lines, but allows angles between lines to change (see Figure 5.3). It is useful

Figure 5.3: The affine transform

to describe the shearing shape in different scenes. The formula of affine transformation is:

$$\mathbf{G} = \mathbf{A}\mathbf{F} + \mathbf{b} \tag{5.6}$$

where $\mathbf{A}$ is a transformation matrix with element a_{ij} which has no restrictions. We can see that the scaling transforms are special cases of affine transform.

- Projective transformation
  The *projective transform* preserves the straightness of the lines but not the

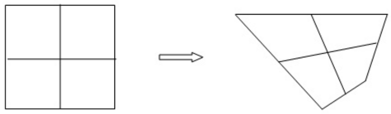

Figure 5.4: The projective transform

parallelism. The parallel lines converge toward vanishing points (see Figure 5.4). It is useful to describe "tilted" scenes. The formula of projective transform is:

$$\mathbf{G} = (\mathbf{AF} + \mathbf{b})/(\mathbf{pF} + \alpha) \tag{5.7}$$

where $\mathbf{p}$ is a $n \times 1$ projection vector. α is usually set to 1.

- Other nonrigid transforms
  The scaling, Affine and projective transformations keep the straightness of lines and hence the planarity of the surface. Other transformations like *curved transformations* do not. Usually they have more complex mathematical formula, which make the parameter estimation hard to converge, and more likely over-fit to the selected control points.

5.2 2D Image Registration Method

The common 2D registration methods include:

- **Point-based methods**
  In point-based method, the correspondents of the points belonging to two images are needed to compute the transformation. These correspondents can be discovered manually or automatically.

- **Intensity-based methods**
  This method uses the information theory. If two images are well registered, then their normalized mutual information should be the maximum. By maximizing the normalized mutual information, the transformation parameter is calculated.[2]

5.3 3D Rigid Transform

If the object to be registered is rigid in 3D, a *3D rigid transform* is needed to describe its movement. This transform is with at least 6 degrees of freedom: three rotations along x, y, z axes: θ_x, θ_y, θ_z; and three translations: t_x, t_y, t_z. Each rotation will result in a left multiplication of the rotation matrix 5.8 to the whole translation matrix.

$$
\begin{aligned}
R_x &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & C_x & -S_x \\ 0 & S_x & C_x \end{pmatrix} \\
R_x &= \begin{pmatrix} C_y & 0 & S_y \\ 0 & 1 & 0 \\ -S_y & 0 & C_y \end{pmatrix} \\
R_x &= \begin{pmatrix} C_z & -S_z & 0 \\ S_x & C_z & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{aligned} \tag{5.8}
$$

with $C_k = \cos\theta_k$, $S_k = \cos\theta_k$, $k \in \{x, y, z\}$. If assuming the object rotates along z,y and x axes orderly,the whole rotation matrix has the form:

$$
R = \begin{pmatrix} C_yC_z & -C_yS_z & S_y \\ C_xS_z + S_xS_yC_z & C_xC_z - S_xS_yS_z & -S_xC_y \\ S_xS_y - C_xS_yC_z & S_xC_z + C_xS_yS_z & C_xC_y \end{pmatrix} \tag{5.9}
$$

5.4 Iterative Closest Points Registration

The ICP algorithm is an automatic 3D surface registration method first proposed by [4] and [7]. The basic idea of this algorithm is to find pairs of closest points between two meshes as the correspondents; then calculate and apply a 3D rigid transform to decrease the total distance of the correspondents. The process is iterative until convergence or the distance is small enough. The ICP algorithm works well when the initial positions of the two meshes are close. This algorithm is fast, simple and accurate when the initial is good which make it the most popular algorithm in the 3D registration field nowadays.

5.4.1 Basic ICP Procedure

The key steps of the ICP are:

- **Finding the correspondents:** If the ICP uses point-to-point distance, the point p_s on source mesh has the correspondent point p_t on target mesh which has smallest Euclidean distance to p_s.
- **Applying the transform:** It is efficient to use *quaternion* (see Appendix A) or *SVD* method to calculate a 3D ridge transform according to the correspondents. This transform, when applying to the points on source mesh, can decrease the average distance between all the pairs of correspondents.
- **Iterative implementation:** The above two steps can be implemented iteratively until convergence or the desired average distance is achieved.

This process is easily understand by a flowchart in Figure 5.5

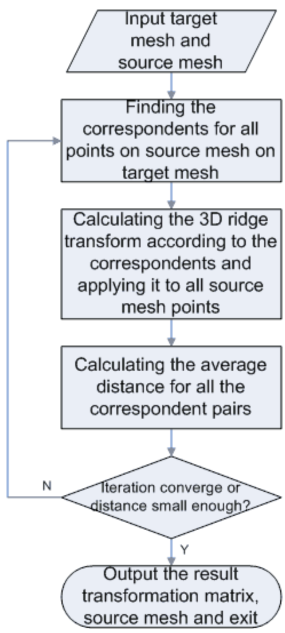

Figure 5.5: The flowchart of basic ICP process

5.4.2 A 2D Curve Registration Example

An illustration of registration two 2D curve using ICP can be seen from Figure 5.6. The blue points are the source, the red points are the target which is a 2D rigid transform of the source and added some Gaussian noise. The colorful lines indicate the correspondents found in each iteration.

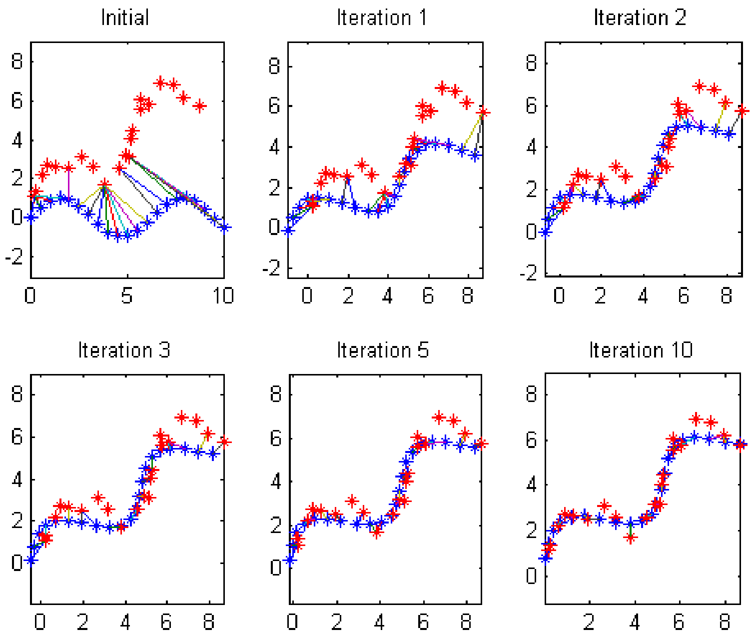

Figure 5.6: An illustration of the ICP registration of two curves, where the blue points are the source, the red points are the target and the colorful lines indicate the correspondents found in each iteration

This experiment shows that the basic ICP algorithm works well for the simple shape. It is robust and converges fast.

Chapter 6

ICP Implementation for Face Registration

The purpose of this chapter is to try to register a simulated face via ICP. I will tell what are the problems to use original ICP and how to introduce the variants to overcome the deficiencies of the algorithm. The experiment result is at the final part.

6.1 Some Problems of Original ICP

As stated in the previous chapter, the concept of ICP algorithm is finding correspondents according to the distance measure, calculating and applying the 3D rigid transform and iterating. It works well for some simple shapes.

Before we proceed to register a simulated face, some of the problems should be first looked at:

- The data size for the normal face application is usually thousands of vertices per face model, finding the correspondents for all these points will be very time consuming.

- The correspondents should only exist on the overlapping part of the source and target mesh. The points on non-overlapping part of the source mesh will find wrong correspondents on the edge of the target mesh, which will introduce registration mistakes.
- The data sometimes is quite noisy that will decrease the performance of the ICP.

Due to these problems, the original ICP may take long time to converge and can be expected to fail to produce the best result.

6.2 ICP with Variants

The solution is to introduce some variants[25] to the original ICP. According to the problems, they could be:

- Only randomly use part of the points from source mesh in each iteration. Too many points will increase computational complexity. Too few points can not capture the feature and will make the result fluctuate. The experiments indicates about $1\% \sim 10\%$ points is a good compromise to speed and accuracy.
- If one of the correspondents is on the edge of the target mesh, this pair of correspondents will not be used in computing the transformation. This process will ensure most of the correspondents are from the overlapping part of the two meshes.
- For each iteration, the average distance of the correspondents is calculated. If the correspondents have relatively large distance comparing to the average distance from previous iteration, this correspondents will not be used.
- Another remedy is to compare the normals of the correspondents, and discard the ones beyond a threshold. This is for not to incorporate the noise points which usually have large distance and irregular normals.

6.3 Simulation Results

The experiments to investigate how the ICP with variants can track the rotation of the face are implemented below. The translation of the face will not affect the registration result because the centers of the target and source mesh can be matched up before registration.

6.3.1 Model I: part of a face

The first 3D model for simulation is a 3D laser high-resolution scan of a part of the face with 6252 vertices (see Figure 6.1).

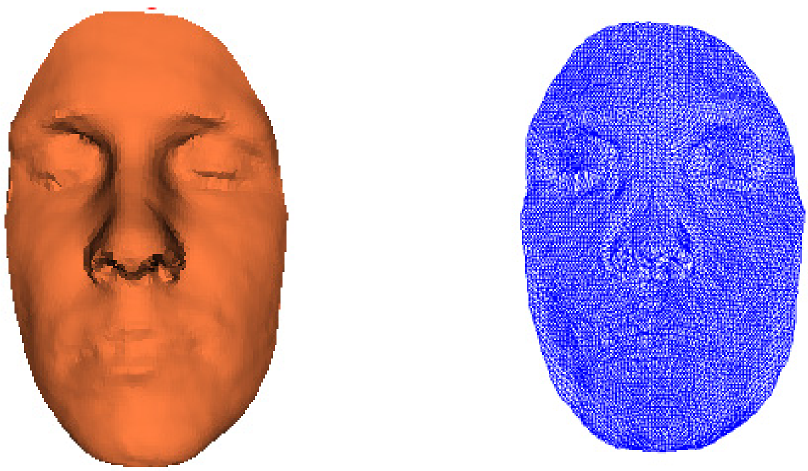

Figure 6.1: The first model for simulation. The left is rendered 3D display. The right is the triangulation of all the vertices.

The target mesh is produced by projecting the model along the z axis into a 71×100 equally spaced grid. The value of the grid points are interpolated by Barycentric coordinate transform(see Appendix B) and then plus some Gaussian noise. This will simulate a depth image (Figure 6.2) with about 3500 vertices.

The source meshes are generated by rotating the model along x,y and z axes from $-30^\circ \sim 30^\circ$ with step 10° and then projecting at the same way as producing target mesh. Some of the source meshes are shown in Figure 6.3

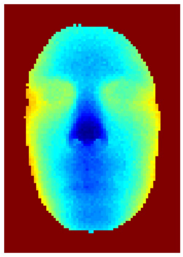

Figure 6.2: The simulated depth image which is produced by projecting the model along z axis to a 2D 71×100 plane then plus Gaussian noise. The target mesh is read from this depth image which have about 3500 vertices.

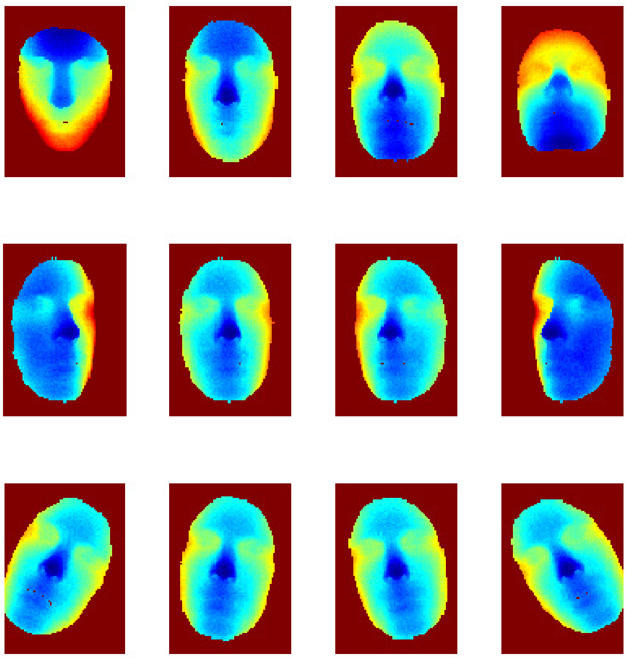

Figure 6.3: Some of the source meshes generated by rotating the model and projecting at the same way as producing the target mesh. From the first row to the last, left to right, are the depth images rotating along x, y and z axes with $-30°$, $-10°$, $10°$ and $30°$ respectively.

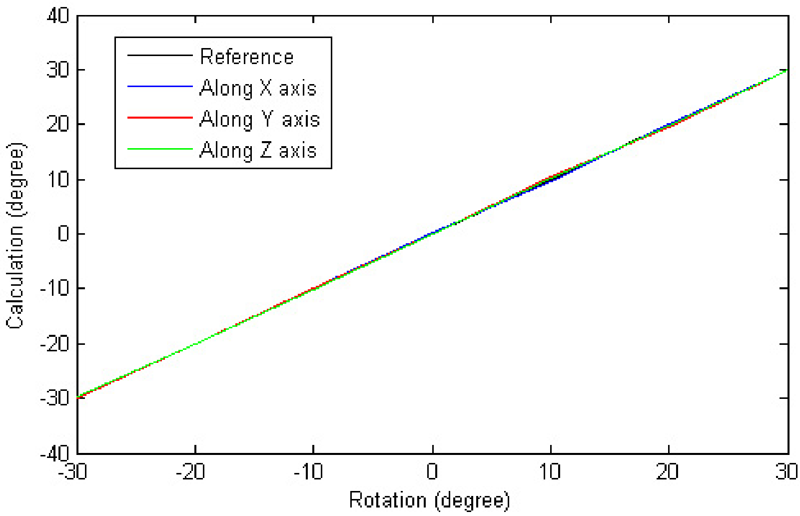

Figure 6.4: The registration result of rotated meshes. The result shows the algorithm has no problem to handle 30° rotation in this case.

Now register the source meshes to target mesh via ICP with variants. The result shows how this algorithm can track the rotation of the face (See Figure 6.4). The rotation angle is calculated according to 3D rigid transformation matrix 5.9. The algorithm has no problem to handle 30° in this case.

Figure 6.5 shows the registration process with the source face having 30° rotation plus a bit translation. The target face is shown in skin color and the source is in blue. The interlaced appearance of the two colors in the final result tells a good match.

6.3.2 Model II: a full face

The second 3D model for simulation is a 3D laser high-resolution scan of a full face with 20904 vertices (see Figure 6.6).

The same processes are made. The 73×100 target depth image resulting about 4500 vertices. Some translated source images and the tracking results are in Figure 6.7 6.8 and 6.9.

As before, a source face with rotation 30° plus some translation is created. The

Figure 6.5: The registration process for source face with 30° rotation plus some translation of Model I. The target face is shown in skin color and the source face is in blue. From left to right, top to bottom: the initial positions, iteration 5, 10, 20, 30 and 50.

registration process is plotted in Figure 6.10. Although the model is quite noisy, it converges to a good result eventually.

6.4 Summery

According to the property of our data, several variants of ICP are added. The experiments prove these improvements make the ICP robust to noise. In simulation, there is no difficulty to register objects within 30° rotation and considerable translation. The drawback of current algorithm is the slowness of the convergence.

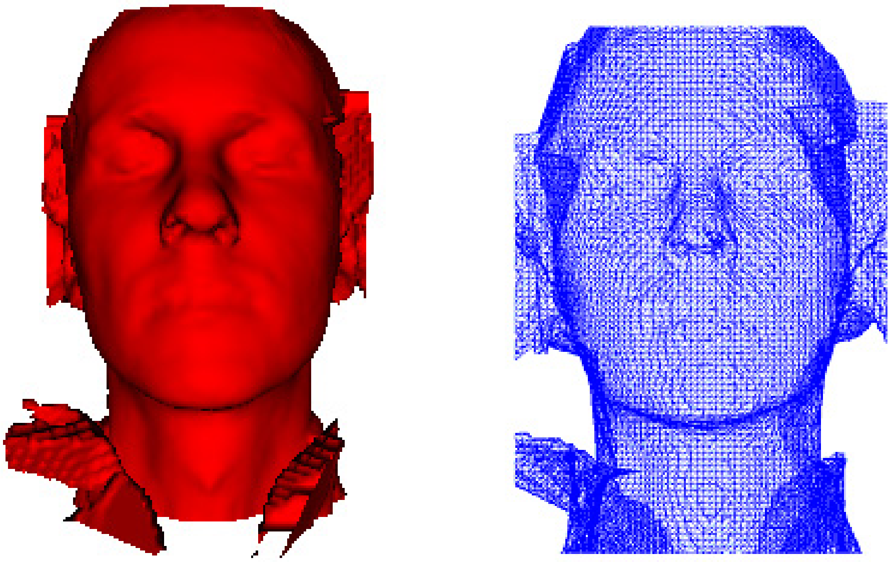

Figure 6.6: The second model for simulation. The left is rendered 3D display. The right is the triangulation of all the vertices.

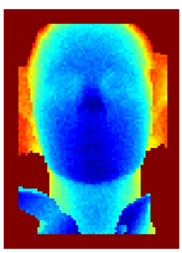

Figure 6.7: The target depth image which produces the target mesh with about 4500 vertices.

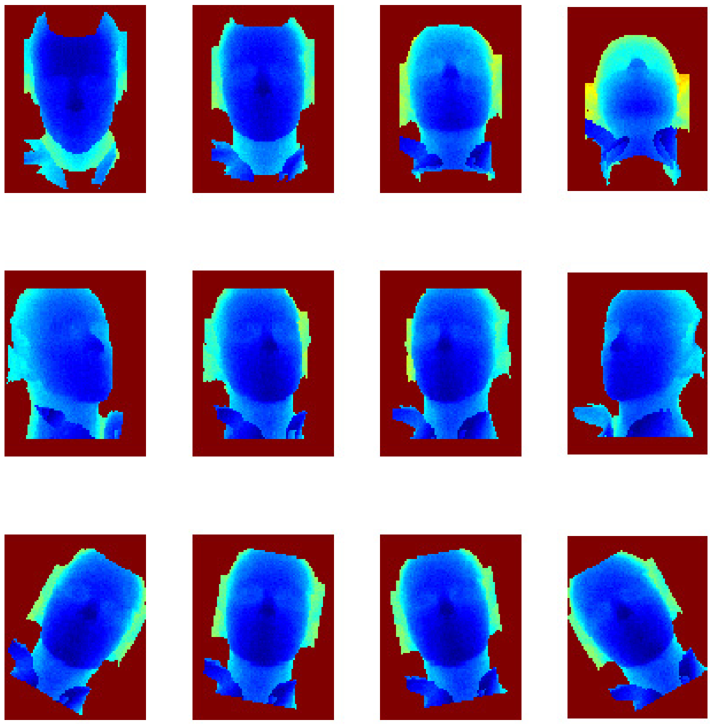

Figure 6.8: Some of the rotated source meshes.

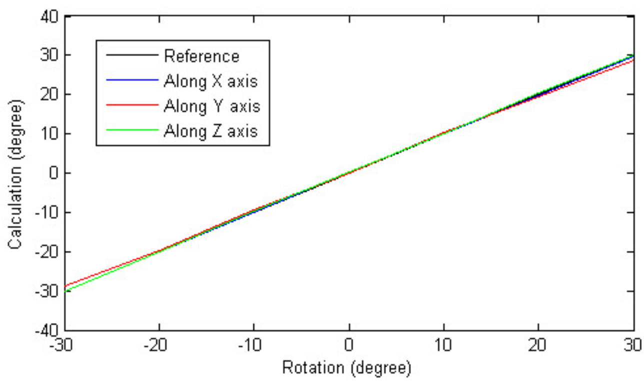

Figure 6.9: The registration result of rotated meshes. There is a little deviating for the y axis rotation, but in all the result is good enough.

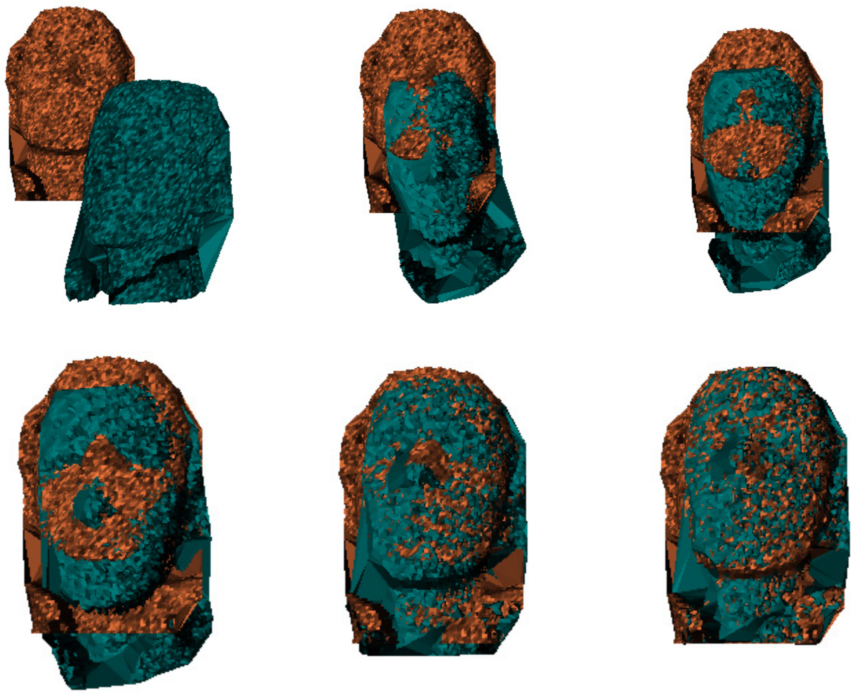

Figure 6.10: The registration process for source face with 30° rotation plus some translation of Model II. The target face is shown in skin color and the source face is in blue. From left to right, top to bottom: the initial positions, iteration 5, 20, 40, 60 and 80.

Part V

Surface Reconstruction

Chapter 7

From Point Cloud to Surface

The registered points are still separated before merging them to one model. Since the scanner discussed in this report can only measure surface of the target, the merged model will be the target's surface. The concept of reconstructing surface from measured points and the previous research will be the content of this chapter.

7.1 Current Situation

The SwissRanger and other laser scanner produce sampled range of the surface to the camera. These samples, forming unorganized point cloud, represent the target's surface. Unfortunately, due to the inaccuracy of the measurement and the registration error, the simple triangulation of all the points will result in bumps and zigzag on the surface. Actually, reconstructing the precise surface from unorganized point clouds, in case of incomplete, sparse and noisy data, is difficult and still not been completely solved [11].

Luckily, with the rapid development of the scanning device, more and more attentions are received in this area during recent years. Some of the previous research will be addressed in the following section.

7.2 Some Popular Algorithms

7.2.1 Averaging Method

- **Simple Averaging**
  The point set can be simply consolidated by replacing the repeated measurements with their mean values. This simple averaging is not robust, and easily produces jumps (see Figure 7.1).

- **Weighted Averaging**
  If assigning different weight to each repeated point, then the averaging process may produce desired result. For example, someone can make the outliers have zero weight so as to remove them. In Dorai's article [5], the author use the formula

$$p_{avg} = \frac{W_s p_s + W_t p_t}{W_s + W_t} \tag{7.1}$$

  to keep a smooth boundary. In this formula, p_{avg} is the averaged point, p_s, p_t are the repeated measurements from source and target meshes and W_s, W_t are the weights related to the distance between point and boundary. The distance is found iteratively, and the weight is increased with the distance. Figure 7.1 shows the result of merging the same surfaces.

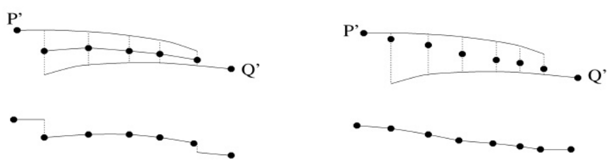

Figure 7.1: The averaging methods. (Left)The simple averaging which produces jumps at boundary. (Right)The weighted average by Dorai's method. Figure is from [5]

7.2.2 Moving Least Squares

The goal of reconstruct surface is usually to find an optimal 2D or 3D curve from point set. This optimization can be done locally, for example, drawing a curve fitting the neighbors of a point and then update the point according to this curve. This method is called *Moving Least Squares* (MLS). In reference [18], the author defines a smooth manifold surface approximated by the MLS.

In his method, to construct the surface around point r, a reference plane $H = \{x| < n, x > -D = 0, x \in \mathbb{R}^3\}$ is first computed by minimizing

$$\sum_{i=1}^{N}(< n, p_i > -D)^2\theta(\|p_i - q\|) \tag{7.2}$$

where n is projection direction for r to H, q is the projection point, p_i is the neighbor point of r and θ is a smooth, positive, monotone decreasing function. Once the reference plane is determined, the height of p_i over H can be calculate as $f_i =< n, (p_i - q) >$. Then the coefficients of a polynomial approximation g is computed by minimize the weighted least squares error

$$\sum_{i=1}^{N}(g(x_i, y_i) - f_i)^2\theta(\|p_i - q\|) \tag{7.3}$$

The symbols in the Equation 7.2 and 7.3 are plotted in Figure 7.2.

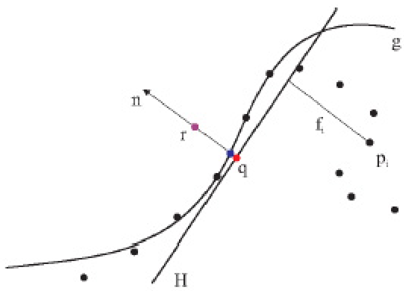

Figure 7.2: The graphic illustration of the symbols used in equation 7.2 and 7.3. The figure is from [18].

If choose θ as a Gaussian

$$\theta(d) = e^{-\frac{d^2}{h^2}}$$

then, the ability of the surface to reveal details can be controlled by h.

7.2.3 Volumetric Method

The *volumetric* method describes the surface by voxels in the volume. In reference [8], the author introduced a simultaneous updating scheme. Each voxel has a cumulative signed distance function $D(x)$ and a cumulative weight $W(x)$, which is updated by i's input range image with

$$\begin{aligned} D_i(x) &= \frac{W_{i-1}(x)D_{i-1}(x) + w_i(x)d_i(x)}{W_{i-1}(x) + w_i(x)} \\ W_i(x) &= W_{i-1}(x) + w_i(x) \end{aligned} \tag{7.4}$$

Here, $d(x)$ is the signed distance from x to the nearest surface along scan line and $w(x)$ is the weight which is computed by linear interpolation of the weights stored at neighboring range vertices (See Figure 7.3).

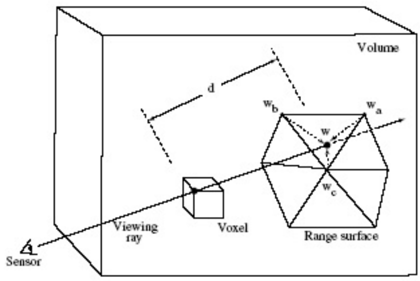

Figure 7.3: The signed distance and weight. Figure is from [8]

The isosurface is extracted from $D(x) = 0$.

7.3 Other Possibilities

If the intrinsic topology of the data is known, reconstruct a parametric surface might be much more accurate and easy. Unfortunately, the details of the date is usually not acquirable in advance. The curve fitting process may partly estimate the structure of the data. For some special usages, where the selection of target objects is limited, then a model recognition process may provide sufficient knowledge. Some of the related works can be found from [15] and [24].

Another possibility is to use high dimensional MRF. In the previous chapters, it has been proved that the MRF is a very powerful tool for 2D image restoration. However, its applications on higher dimension are very rare. This is partly because the difficulties to define neighbors in high dimensional space and the optimization procedure might be very complicated. In the next section, I will briefly introduce an application of MRF to refine surfaces.

Chapter 8

Surface Refinement by 2.5D MRF

In chapter 4, the experiments indicate that restoration result of the depth image using MRF is with low noise and high contrast. Chapter 6 shows the ICP with variants is suitable to find a 3D ridge transform. In this chapter, I will introduce a new method to merge two or more surfaces. These surfaces are at the assumption of having small alignment error and low level of noise. The method is based on high dimensional MRF. This MRF may called *Surface Combining MRF* or *2.5D MRF*. For high dimensional MRF, the most difficulties include defining neighbors, designing suitable mathematical form of the prior and applying optimization approach. My trials of solving these obstacles will be stated below.

8.1 Some Definitions

Before write the mathematical formulae, let me start with introducing some terms used to form my MRF.

- **Overlapping**
  The *overlapping* parts are the parts of two meshes covering the same area.

In my implementation, the points belonging to overlapping part of one mesh are the ones having closet corespondents of another mesh and their corespondents are not on the edge of that mesh.

- **Self Neighbor**
  A point A's *self neighbors* are the points belonging to the same mesh with A and at the self neighbor positions of the A. In my implementation, the mesh is first triangulated by Delaunay triangulation. All the vertices being in the same triangle with A are A's self neighbors. In Figure 8.1, all the green points are the self neighbors of red point.

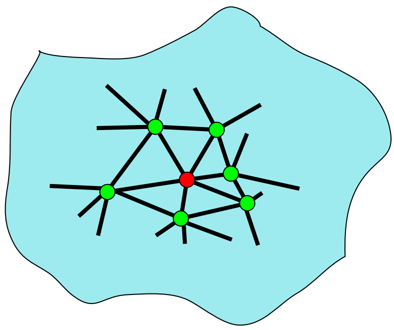

Figure 8.1: An illustration of self neighbor. All the green points connecting the red point by Delaunay triangulation are the self neighbors of red point.

- **Alien Neighbor**
  A point B's *alien neighbors* are the points belonging to the different meshes with B and at the alien neighbor positions of the B. In my implementation, only when B belonging to the overlapping part with another mesh have alien neighbors on that mesh. All the meshes not having B are first triangulated by Delaunay triangulation. The vertices at the closest triangle to B in one mesh is the alien neighbors of B in that mesh. Finding the closest triangle is simplified by finding the closest middle point of each triangle. So, in my experiment, a point B will have no alien neighbors in another mesh if it is not on the overlapping part of that mesh, otherwise will have three alien neighbors. In Figure 8.2, the yellow point finds its three alien neighbors.

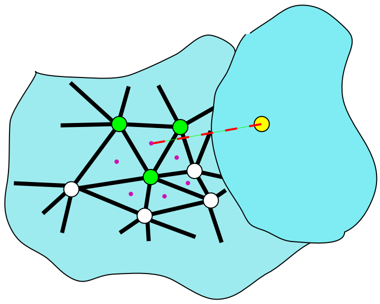

Figure 8.2: An illustration of alien neighbor. The purple points are the middle points of the triangles. The three vertices of the triangle having the nearest middle point to yellow point are the alien neighbors of the yellow point.

8.2 Mathematical Expression

The mathematical expression of my 2.5D MRF is inspired by the MRF used in Chapter 4 (4.1 $\sim$ 4.5).

To restore a surface with point y_i, the depth measurement potential has the same form with 4.1 except that L is all the points on the surface needed to restored.

$$\Psi = \sum_{i \in L} k(y_i - z_i)^2 \tag{8.1}$$

k is called *proportion factor*, which adjusts whether the result is more determined by original measurements or prior.

For the log depth smoothness prior, the difference with 4.2 is that $\tilde{N}(i)$ here is only the clique of alien neighbor, so that all $\tilde{y}_j$'s in 8.2 are alien neighbors of y_i. For one point, there will be none or there alien neighbors as stated above.

$$\Phi = \sum_i \sum_{j \in \tilde{N}(i)} w_{ij}(y_i - \tilde{y}_j)^2 \tag{8.2}$$

If choose the weight w_{ij} as in 4.3, and c positive, $\forall_{i,j} u_{ij} \geq 0$, then w is in the range $\exp(-c \max(\forall_{i,j} u_{ij})) \sim 1$

$$w_{ij} = \exp(-cu_{ij}) \tag{8.3}$$

At this situation, the c can be called as *discrimination factor*, which adjusts the appearance caused by the difference of u.

The formation of u is important to decide what the prior will be in this MRF. Based on the assumption of our depth measurement, I expect the locally smooth part of a mesh can combine with other overlapping part seamlessly, because mostly they are separated by registration error. Whereas the part with details should be more likely to keep the origin measurements, because it helps to increase the resolution.

To evaluate the local smoothness at a point y_i, I choose the maximum angle difference between the normal of the measurement z_i with the normals of its self neighbors z_j. So that

$$u_{ij} = u_i = \max_{\forall j} < \vec{n}_{z_i}, \vec{n}_{z_j} > \tag{8.4}$$

here, $\vec{n}$ is the normal vector and

$$\begin{aligned} w_{ij} &= w_i = \exp(-cu_i) \\ \Phi &= \sum_i w_i \sum_j (y_i - \tilde{y}_j)^2 \end{aligned}$$

Finally, the log posterior probability is the summation of $-\Psi$, $-\Phi$ and a constant.

$$\log p(y|x,z) = -\Psi - \Phi + c' \tag{8.5}$$

8.3 Optimization by 2.5D ICM

One of the advantages of separating alien and self neighbors is that the 2D ICM algorithm can be easily extended into 2.5D. In stead of updating odd and even grid consecutively, the *2.5D ICM* updating points from one mesh to another with the updating rule

$$\hat{y}_i = \frac{kz_i + w_i \sum_{j \in \tilde{N}(i)} \tilde{y}_j}{k + w_i num(\tilde{N}(i))} \tag{8.6}$$

here, $num(\cdot)$ means the number of. If the errors between the meshes are small, the alien and self neighbor cliques and the weight can be thought to be fixed during iterations. Then, they can be computed in advance, so that the total computational complexity of this 2.5D ICM is $Q\mathcal{O}(N)$. Here Q is the number of iterations and N is total number of points. This is not a very big computation amount.

8.4 Experiments

To display the surface of a combined point set, four methods are provided here.

1. **Direct Display**
   Each sub-point set is triangulated and rendered in the same view. This method can not provide an integrated 3D model, but feasible to observe the difference between the point sets. The displayed surface is the patches of the outermost ones for each point set which generally is not the correct result. We have met this method in Chapter 6 before.

2. **Display Appended Point Set**
   This method is just simply appending all the points, and triangulating as a whole. It is ok for error free situation, but usually suffered by zigzag pattern on non-ideal cases.

3. **Simple Averaging**
   As described in the previous chapter, this method can partly consolidate the points, but not very robust.

4. **2.5D MRF**
   This is the method I derived in this chapter. It is designed to remove noise on smooth part and recover details on rough part.

8.4.1 Noise Robustness Test

First I transform one of the image in Figure 6.3 back with correct parameters. Figure 8.3 shows the target, correctly registered source, and direct display of the both.

Figure 8.4 shows the appended display, simple averaging and 2.5D MRF. Clearly, compared with original face in Figure 6.1, 2.5D MRF removes most of the noise and partly recovered the mouth which is hard to see in any other image.

The test on Model II for more prominent noise (See Figure 8.5 and 8.6) proves that 2.5D MRF ensures a reasonable solution even if a disastrous scanner is at hand.

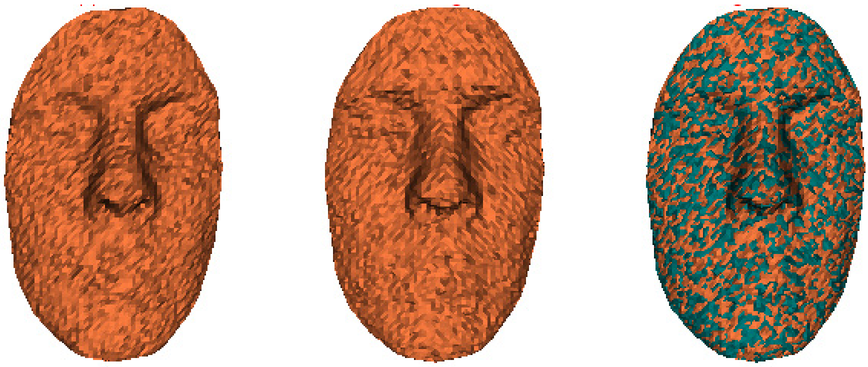

Figure 8.3: Noise robustness test for Model I figure I. From left to right, the target, registered source and direct display of both faces where the registered source is in blue.

8.4.2 Registration Error Robustness Test

Now the noise-free Model I has been applied with a registration error of one degree and small translations for x, y and z axes. From Figure 8.7 we can see that even a small error will ruin the appended display. Both simple averaging and 2.5D MRF are robust for the small errors. It is hard to increase the resolution by simple average, but 2.5D MRF can generate finer grid (see Figure 8.8) so as to make a super-resolution when the error is tolerable.

When the registration error is enlarged for Model II. The 2.5D MRF seems more suffered by the misalignment of the details (see Figure 8.9). It is expected, because the algorithm can not discriminate where is the misalignment to remove or where is the additional useful details. The misalignment should be handled as much as possible in the registration process.

8.5 Summery

An 2.5D MRF is proposed and implemented here. This MRF aims to combine multiple surfaces with low noise and small mis-registration error. The basic idea is to move the point as close as possible to its alien neighbors if it is considered to be locally smooth and keep the original value when it is a detail. The local smoothness is evaluated by the maximum angle difference between the normals of the point and its self neighbors'.

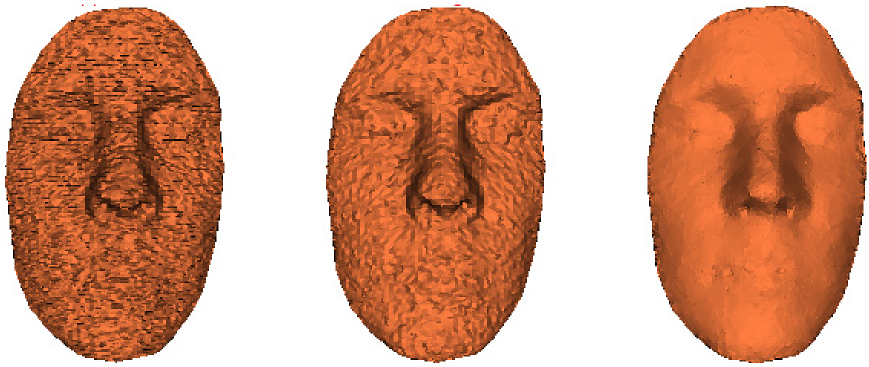

Figure 8.4: Noise robustness test for Model I figure II. From left to right, the appended display, simple averaging and 2.5D MRF. Clearly the last one is the best which removes most of the noise and recovers part of the mouth.

The experiments show that this surface combining MRF is very powerful to handle noisy situations and can make super-resolution when the registration error is small.

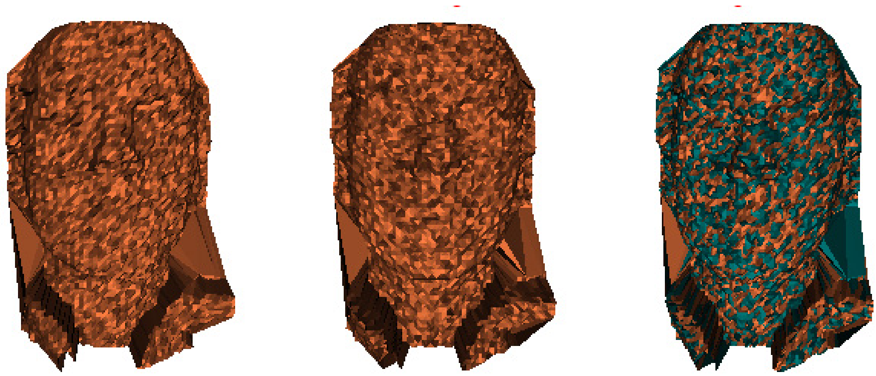

Figure 8.5: Noise robustness test for Model II figure I. From left to right, the target, registered source and direct display of both faces where the registered source is in blue.

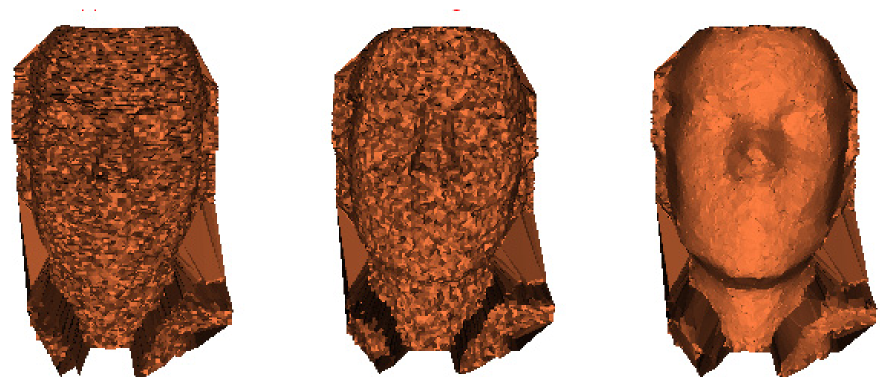

Figure 8.6: Noise robustness test for Model II figure II. From left to right, the appended display, simple averaging and 2.5D MRF. The 2.5D MRF ensures a reasonable solution even when the noise is prominent.

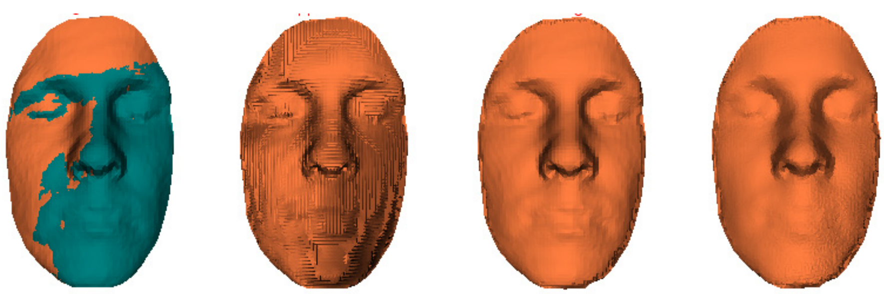

Figure 8.7: Registration error robustness test for Model I. From left to right, the direct display, appended display, simple averaging and 2.5D MRF.

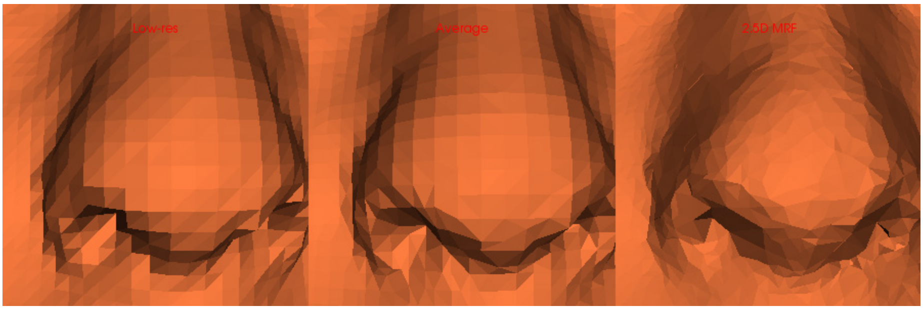

Figure 8.8: The blowups of the nose for Model I. From left to right, the original low-resolution surface, the simple averaging and 2.5D MRF. The 2.5D MRF can generate finer grid than simple averaging, so as to make a super-resolution.

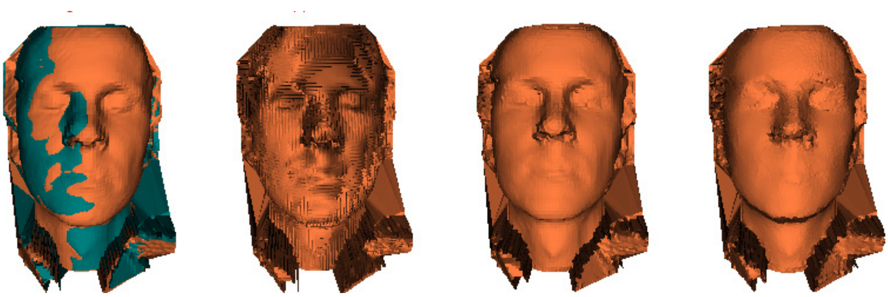

Figure 8.9: Registration error robustness test for Model II. From left to right, the direct display, appended display, simple averaging and 2.5D MRF. The 2.5D MRF is more suffered by the misalignment of the details.

Part VI

Results

Chapter 9

Manikin Face Modelling and Super-Resolution

Based on the techniques stated in the previous chapters, an experiment to create a manikin face model and super-resolution surface are implemented here. The process includes scanning the model, extracting the face, restoring the image via MRF, registering the depth measurement via ICP, modelling face via simple averaging and creating super-resolution surface via 2.5D MRF.

9.1 Data Acquisition

The target object is a manikin head placed on a rotation table (see Figure 9.1). The head is about $0.5m$ far from the camera, which have the depth resolution $2.5 \sim 6mm$ according to the manual [1], resulting losing many details of the face.

The SR-3000 is fixed and captures depth images while the head rotates with the rotation table. Each time the head rotates approximately 1 degree. 150 images are captured which cover the whole face and most part of the head. Appendix C shows some of the depth and corresponding intensity images.

Figure 9.1: Photos of the front and side views of the real manikin head.

9.2 Face Extraction

Before registration, the preprocessing to extract the manikin face from the noisy depth image must be done first, which can be divided into the following steps:

- **Cutting Rotation Table**
  The rotation table is connected to the manikin head, but it is not interested. In this experiment, the rotation table is always located at the bottom of the image. It can be easy removed by cutting part of the image.

- **Separating the background**
  The background is supposed to be far from the camera, so a threshold applying to the range image will be sufficient.

- **Removing the Isolated Noise**
  When the integration time is not long, there will be some noise. The pattern of the noise is usually isolated or forms small patches. The removal algorithm is similar with the one used in reference [5]. The size of each connected regions in the image is computed, and the ones less than 20% of the largest one will be removed as noise part.

- **Restoring the Measurement**
  The high-resolution and low noise depth measurement is restored by MRF which is detailed described in Chapter 4.

- **Solving Edge Extension**
  As stated in Chapter 2, the psf function of the camera is about 0.4 pixel standard deviation Gaussian function. The psf function will make smooth descents around sharp edges causing the object stretch toward the background. My solution is deleting 3 pixels' thick edge from the outermost of the object which seems a good compromise between removing the extension and keeping the information.

Figure 9.2 shows a preprocessed face, which is also used as the target in most of the succession experiments.

Figure 9.2: The front and side views of one of the preprocessed depth measurement of the manikin images used as target in most of the succession experiments.

9.3 Depth Registration via ICP

The method of registration of the multiple views is based on the ICP algorithm with variants as stated in Chapter 6. To accelerate the speed of searching the nearest correspondents, KD tree [14] is used.

Although optimal results have been got for simulation in Chapter 6, the experiment here shows the registration error is likely to increase dramatically when the rotation is more than 10 degree. Furthermore, the side views are much harder to be matched than the front views because of the lack of features.

Figure 9.3 and 9.4 are the processes of two views with about 5 and 10 degrees registered together. The mismatch on the noise is obvious in Figure 9.4.

9.4 Face Modelling

Now 15 views with 5° rotation difference are registered and synthesized into a face model covering a broader part of the face. The points of this model are generated by simple averaging, which are displayed in Figure 9.5. In order to ensure the combining process incremental and order independent which means the updating is simultaneous and unbias to each scan, every point is assigned a weight w with initial value 1 and the updating of a point P is following the rule

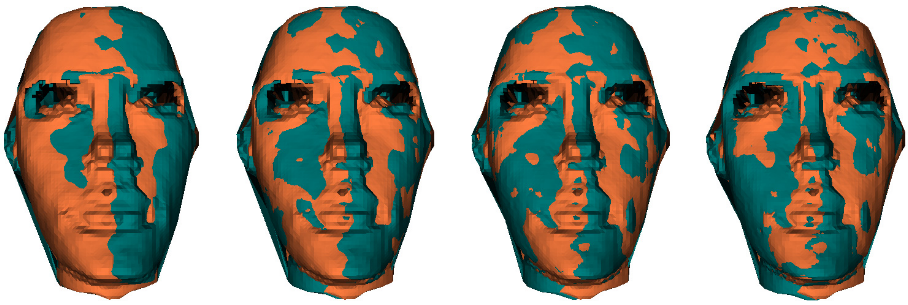

Figure 9.3: The registration process of front views with about 5 degree rotation. From left to right: the initial positions, 10, 20, and 50 iterations.

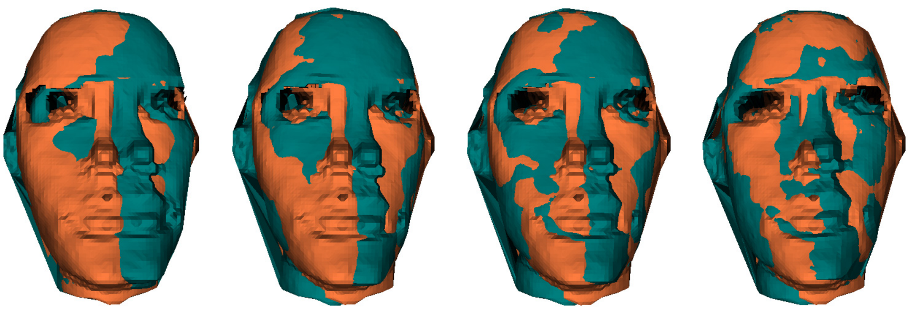

Figure 9.4: The registration process of front views with about 10 degree rotation. From left to right: the initial positions, 10, 20, and 50 iterations. The mismatch around the nose is obvious.

9.1 where the i is the scan number.

$$P_i^{avg} = \frac{(w_{i-1}P_{i-1}^{avg} + P_i^{new})}{w_{i-1} + 1} \quad \text{and} \quad w_i = w_{i-1} + 1 \tag{9.1}$$

To render and display this point set, one may need to use volumetric method as mentioned in Chapter 7. For simplicity. the Matlab functions for volumetric rendering is used. From Figure 9.5 and 9.6, we can see the deadly accumulation of errors.

Figure 9.5: Face model synthesized by 15 registered views, each with 5° rotation. The points are generated by incremental simple averaging method.

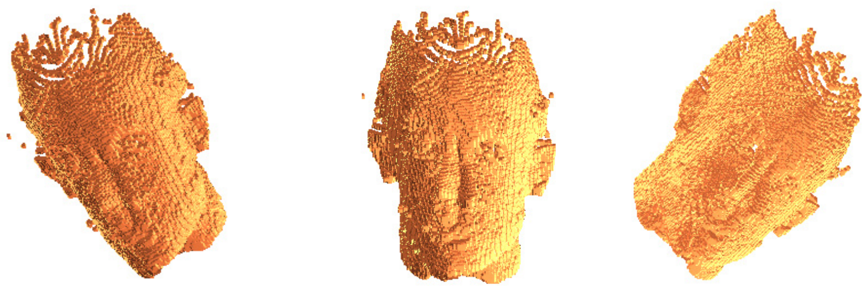

Figure 9.6: The volumetric rendering in Matlab for the point set of Figure 9.5

9.5 Face Super-Resolution

In Chapter 8, the super-resolution of the simulated face is generated by 2.5D MRF when the registration error is small enough. To achieve this, two views with about 3 degree rotations are registered to the target view. Figure 9.7 shows the 2.5D MRF combining result, in which, more details are recovered when k and c are large and more smoothness when k and c are small.

9.6 Summery

Finally, the real manikin face is on its stage here. The depth information of the face is first extracted by a series of preprocessing procedures. Then, the different views are registered by the ICP algorithm with variants. Based on the registration result, we can concatenate views so as to cover a broader part

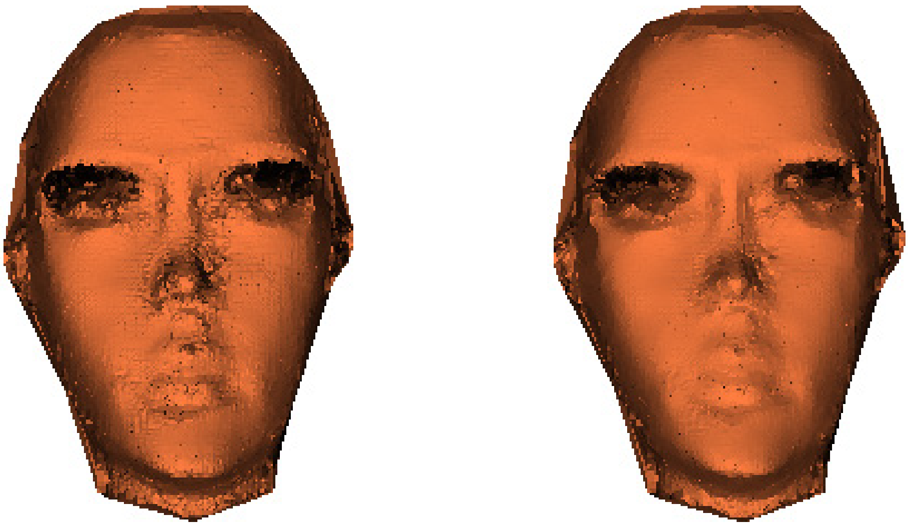

Figure 9.7: The super-resolution from 3 views by 2.5D MRF. The left is generated by large k and c resulting in more details and the right is generated by small k and c showing more smoothness.

of the face, or utilize the repeated measurements in the same part for super-resolution. The result of concatenation shown in this report from 15 views is done by incremental simple averaging, and the super-resolution is achieved by 2.5D MRF from 3 views.

This chapter also expresses how different between the real and ideal cases in computer vision. The good property of ICP tested in simulation is greatly deteriorated here, and the registration error is propagated when connecting multiple views. The main reason is maybe the inconsistency of the measurement in different views. The algorithm tries to match the majority smoothness part of the face like the forehead and cheek but omits the details. However, these smoothness parts may change their shapes a bit during the rotation in the depth image captured by SwissRanger.

Part VII

End Remarks

CHAPTER 10

Future Work and Conclusions

This report shows a trial to modelling and super-resolution a manikin face from range images. These images are captured by SwissRanger SR-3000, a newly invented range detection camera. The trial mainly includes an analysis of the camera, restoration of the range images by a recent proposed MRF scheme, utilization of the widely accepted ICP algorithm with some variants to registration and originating a surface super-resolution method – 2.5D MRF. Due to the time and knowledge limitation of the author, the final result is not fully satisfactory. This also makes more spaces to the future research, so that the future work as far as I can see and conclusions from my previous work may help the succession researches in IMM. The following statement will be arranged by camera test, range image restoration, registration method, surface reconstruction and system consideration.

10.1 Camera Test

SwissRanger is a newly invented range detection camera based on TOF technology. It has the advantages of portable size, low energy consumption, fast frame acquisition rate and the measurements can be easily converted to Cartesian coordinate. However, it also suffers to the low resolution, inaccurate and inconsistent measurement. The characteristics and applications of this camera are still uncertain to the researchers. Simple experiments reveal that it has a sharp spatial psf having 0.4 pixel deviation if fitted to a Gaussian and the integration time should be carefully selected to ensure low noise non-overflow data.

The most annoying thing is that the measurement may be affected by the reflection angle, the material and the colors of the object. This is obvious by looking at the manikin's eyes, which are black and concave and result in two deep holes on the measured face. These physical problems may cause the major limitations of the applications and more or less exist on other laser equipments.

To understand and full utilize SwissRanger, a serial of analytic tests need to be designed to evaluate the performance of this camera. For example, the tests should partly answer the following questions

- How does the illumination affect the measurement?
- How do the reflection angle, material and color affect the measurement? Could it be corrected, maybe by using intensity information or additional instruments like a colorful photo camera?
- Does the psf change with the depth of object? If it does, how?
- What is the deviation of the measurement in different range?

10.2 Range Image Restoration

The depth image produced by SwissRanger is spatially low resolution and noisy, but the byproduct intensity image has high contrast and low noise level. The initial goal is to use the intensity information to restore depth information. This is best achieved by MRF, with designing the intensity information in the prior. Paper [9] introduces a MRF structure for this. By using this structure, the range images are well restored with obtaining the good property of the intensity information.

Although using intensity image is the best low cost solution, it can not be beyond the physical limitations of SwissRanger. One can use an additional photo camera as a new source of prior. This is what the author did in paper [9]. A high quality colorful photo can help to create a super-resolution depth image. Furthermore, it can provide natural color texture.

10.3 Registration Method

The ICP algorithm has dominated the 3D automatic registration applications. This algorithm is relatively fast, simple and accurate if the source is good initialized. Depending on specific problems, some variants of ICP can be introduced to improve the registration performance. In my experiments, the variants are randomly picking part of the points, removing correspondents on edge, deleting outliers with large distances and selecting correspondents having compatible normals. The algorithm is good enough for simulation, but greatly deteriorated when applying on real data.

The problem is maybe caused by the inconsistency of the measurement due to the physical limitations of the SwissRanger. Besides doing corrections on the camera part, an convenient way is to design more variants to ICP. Currently the matches is mainly decided on the smooth part of the surface, so as to omit the details. Then one can first extract the details and align them along. The source initialization is always a big problem. Bad initialization may cause the matching easily trapped into a local minimum. The solution can be matching from low to high resolution, so called *multi-resolution registration*, or trying different initials then selecting the best one. Most directly and efficiently, let the user place the source to an approximate initial. Every successful 3D modelling system nowadays has this machine-user interaction interface. No systems can be purely automatic under considerable errors whatever how good the data they can acquire.

10.4 Surface Reconstruction

Reconstruct a surface from point set is not a new problem, but attracts more and more attentions with the development of scanning techniques. For the different application, new problems may arise. For the data from surface scanning, an robust point consolation algorithm both to noise and registration error is required. Up to now, no algorithms have fully satisfactorily done it.

Some of the widely accepted algorithms include averaging, moving least squares and volumetric method. I proposed a 2.5D MRF to combining surfaces, which is inspired by the one used in 2D range image restoration. In this high dimensional MRF, the neighborhood of a point is divided into alien neighbors which are the vertices of the closest triangle in another mesh and self neighbors which are the vertices in the same triangles of the point. The combination policy is that the smooth part should be moved as close as possible and the details kept the original value. The local smoothness of a point is evaluated by the maximum angle difference between the point and its self neighbors'. This 2.5D MRF can produce super-resolution surface. The simulation also proves the noise and registration robustness of this algorithm. However, it is computationally heavy when handle more than two surfaces and the considerable registration error in the real case is detrimental to the final result.

Except for simple averaging, there is no comparison between 2.5D MRF and other surface reconstruction algorithms. This is necessary to evaluate the advantage and disadvantage for each one. Somebody can create a new assumption to change the form and working procedure of this MRF. Some small revisions could be using gradient to evaluate local smoothness, incorporating different neighbors or optimizing the model other than 2.5D ICM.

10.5 System Consideration

The error and time control is always the major obstacle to the applications of 3D modelling systems. The experiment indicates the dominating error of the system is the registration error. This error will affect the final surface reconstruction process and can not be efficiently reduced after registration step. Several suggestions to reduce pairwise alignment have been stated in the previous section. The accumulation error can be minimized by evenly diffusing the error when all the multi-views are acquired[22].

An 3D modelling systems should be real-time or near real-time, so that the user can simultaneously found the error and correct it by the machine-user interaction interface. The ICP has been proved to be suitable in such systems[25]. However the MRF is usually not. The MRFs implemented in this report are Matlab programmed. They should be converted and optimized. And the suitable rendering method for the point set is also needed. The volumetric rendering result in Matlab is too rough.

Appendix A

Compute 3D Rigid Transform from Correspondents

Given target point set C and source point set P, where any point p_i in P has one correspondent c_i in C, in order to finding a rotation matrix R and translation matrix t applying on P, so that they minimize

$$\frac{1}{N_p}\sum_{i=1}^{N_p}\|c_i - Rp_i - t\|^2 \tag{A.1}$$

where N_p is the number of points in P, one could use unit quaternion or SVD method.

A.1 Unit Quaternion

The unit quaternion method is proposed in reference [13]. The following description of its detailed usage in ICP algorithm is based on the reference [21].

First finding the mean of two point sets:

$$m_c = \frac{1}{N_p}\sum_{i=1}^{N_p} c_i \quad m_c = \frac{1}{N_p}\sum_{i=1}^{N_p} c_i \tag{A.2}$$

Make the two point set zero mean:

$$c_i' = c_i - m_c \quad p_i' = p_i - m_p \tag{A.3}$$

Compute a 3×3 matrix:

$$\begin{aligned} M &= \sum_{i=1}^{N_p} p_i' c_i'^T \\ &= \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix} \end{aligned} \tag{A.4}$$

Then construct a 4×4 matrix:

$$N = \begin{bmatrix} S_{11}+S_{22}+S_{33} & S_{23}-S_{32} & S_{31}-S_{13} & S_{12}-S_{21} \\ S_{23}-S_{32} & S_{11}-S_{22}-S_{33} & S_{12}+S_{21} & S_{31}+S_{13} \\ S_{31}-S_{13} & S_{12}+S_{21} & -S_{11}+S_{22}-S_{33} & S_{23}+S_{32} \\ S_{12}-S_{21} & S_{31}+S_{13} & S_{23}+S_{32} & -S_{11}-S_{22}+S_{33} \end{bmatrix}$$

If the largest eigenvalue of N corresponding to the normalized eigenvector $\mathbf{e} = [e_0, e_1, e_2, e_3]'$, the rotation matrix R will be

$$\hat{R} = \begin{bmatrix} e_0^2+e_1^2-e_2^2-e_3^2 & 2(e_1e_2-e_0e_3) & 2(e_1e_3-e_0e_2) \\ 2(e_1e_2+e_0e_3) & e_0^2-e_1^2+e_2^2-e_3^2 & 2(e_2e_3-e_0e_1) \\ 2(e_1e_3-e_0e_3) & 2(e_2e_3+e_0e_1) & e_0^2-e_1^2-e_2^2+e_3^2 \end{bmatrix}$$

and the translation matrix

$$\hat{t} = m_c - Rm_p \tag{A.5}$$

A.2 SVD

It could be also possible to apply SVD to matrix M in A.4 and compute the R and t.

$$X = USV^T$$

$$\hat{R} = VU^T$$

The translation is the same as in A.5.

Appendix B

Interpolating Points in Triangles

The most common interpolation approach is *inverse distance weighted* (IDW) interpolation. In IDW an estimated point value is only decided by the weighted average of a subset of known points. The points having large relative distances will have small weights while small distances have large weights.

Since three 3D points can determine a 2D plane, the following illustration will be in 2D. Now suppose we want to estimate the value of a point (x, y) in a triangle with vertices i, j and k. The estimation is only based on the value of these vertices Q_i, Q_j, Q_k and the relative position of this point to the vertices. There will be several things to be done when applying IDW.

B.1 Is This a Triangle?

This problem, in another word, is that the three points are not in a line or the area enclosed in this three points are not zero. In area measure, it will be

$$A = \frac{x_i y_j + x_j y_k + x_k y_i - y_i x_j - y_j x_k - y_k x_i}{2} > 0$$

where $x_w, y_w, w \in \{i, j, k\}$ are 2D (x, y) Cartesian coordinate of w, and A is the area of the triangle.

B.2 Is the Point in the Triangle?

We only care about interpolating a point within the triangle or on the edges of the triangle. If given three vertices and a point position, how can we know whether this point satisfies the condition? This condition can be tested by the convex property of the triangle, in which all the points in triangle should be at the same side referring the line containing any of the edges of the triangle (See Figure B.1).

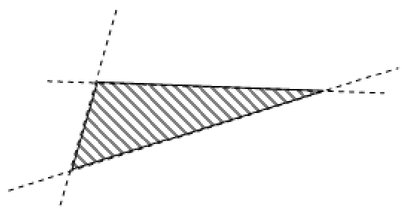

Figure B.1: All the points in the triangle belong to the same side of any of the elongated edges.

Suppose e_i, e_j and e_k are the three edges opposite to the vertices i, j and k respectively. The side, indicated by three sign values S_i, S_j and S_k of a point to these edges are

$$\begin{aligned} S_i &= sign\left((x_k - x_j)y - (y_k - y_j)x - (x_k y_j - x_j y_k)\right) \\ S_j &= sign\left((x_k - x_i)y - (y_k - y_i)x - (x_k y_i - x_i y_k)\right) \\ S_k &= sign\left((x_i - x_j)y - (y_i - y_j)x - (x_i y_j - x_j y_i)\right) \end{aligned}$$

The known point inside the triangle can be selected as the middle point, so that its coordinate is

$$\begin{aligned} \tilde{x} &= \frac{x_i + x_j + x_k}{3} \\ \tilde{y} &= \frac{y_i + y_j + y_k}{3} \end{aligned}$$

and the side of information of the middle point is

$$\begin{aligned}
\tilde{S}_i &= sign\,((x_k - x_j)\tilde{y} - (y_k - y_j)\tilde{x} - (x_k y_j - x_j y_k)) \\
\tilde{S}_j &= sign\,((x_k - x_i)\tilde{y} - (y_k - y_i)\tilde{x} - (x_k y_i - x_i y_k)) \\
\tilde{S}_k &= sign\,((x_i - x_j)\tilde{y} - (y_i - y_j)\tilde{x} - (x_i y_j - x_j y_i))
\end{aligned}$$

Now we have the criterion of a point within a triangle or on the edges of it, which is

$$S_i\tilde{S}_i \geq 0 \;\; and \;\; S_j\tilde{S}_j \geq 0 \;\; and \;\; S_k\tilde{S}_k \geq 0$$

B.3 What Is the Weight Should Be?

If a point is in a triangle, this value can be interpolated by the vertices of triangle according to the relative position of this point. The weights are determined by the Barycentric coordinate of this point. The *Barycentric coordinate* b_i, b_j and b_k of a point is computed by

$$b_i = \frac{a_i}{A} \quad b_j = \frac{a_j}{A} \quad b_k = \frac{a_k}{A}$$

where a_i, a_j and a_k are area of three triangles divided by the estimated point (See Figure B.2). And

$$\begin{aligned}
a_i &= \frac{xy_j + x_j y_k + x_k y - y x_j - y_j x_k - y_k x}{2} \\
a_j &= \frac{x_i y + x y_k + x_k y_i - y_i x - y x_k - y_k x_i}{2} \\
a_k &= \frac{x_i y_j + x_j y + x y_i - y_i x_j - y_j x - y x_i}{2}
\end{aligned}$$

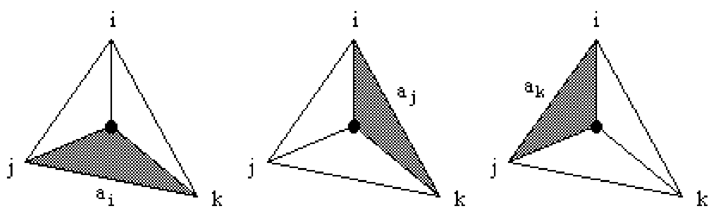

Figure B.2: A point divides the triangle into three sub-triangles which have the area a_i, a_j and a_k

If choose cubic S-shape function as the weight distribution function, the weights are computed as

$$
\begin{aligned}
w_i &= b_i^2(3-2b_i) + 3\frac{b_i^2 b_j b_k}{b_i b_j + b_i b_k + b_j b_k} \\
&\quad \left[b_j \left(\frac{l_i^2 + l_k^2 - l_j^2}{l_k^2} \right) + b_k \left(\frac{l_i^2 + l_j^2 - l_k^2}{l_j^2} \right) \right] \\
w_j &= b_j^2(3-2b_j) + 3\frac{b_j^2 b_i b_k}{b_i b_j + b_i b_k + b_j b_k} \\
&\quad \left[b_i \left(\frac{l_j^2 + l_k^2 - l_i^2}{l_k^2} \right) + b_k \left(\frac{l_i^2 + l_j^2 - l_k^2}{l_i^2} \right) \right] \\
w_k &= b_k^2(3-2b_k) + 3\frac{b_k^2 b_j b_i}{b_i b_j + b_i b_k + b_j b_k} \\
&\quad \left[b_i \left(\frac{l_k^2 + l_j^2 - l_i^2}{l_j^2} \right) + b_j \left(\frac{l_i^2 + l_k^2 - l_j^2}{l_i^2} \right) \right]
\end{aligned}
$$

in which, $l_w \; w \in \{i, j, k\}$ are the length of the edge opposite to vertex w.

Finally, the interpolation value $\hat{Q}$ is computed as

$$\hat{Q} = w_i Q_i + w_j Q_j + w_k Q_k$$

Appendix C

Sample Depth and Intensity Images Used in Experiment

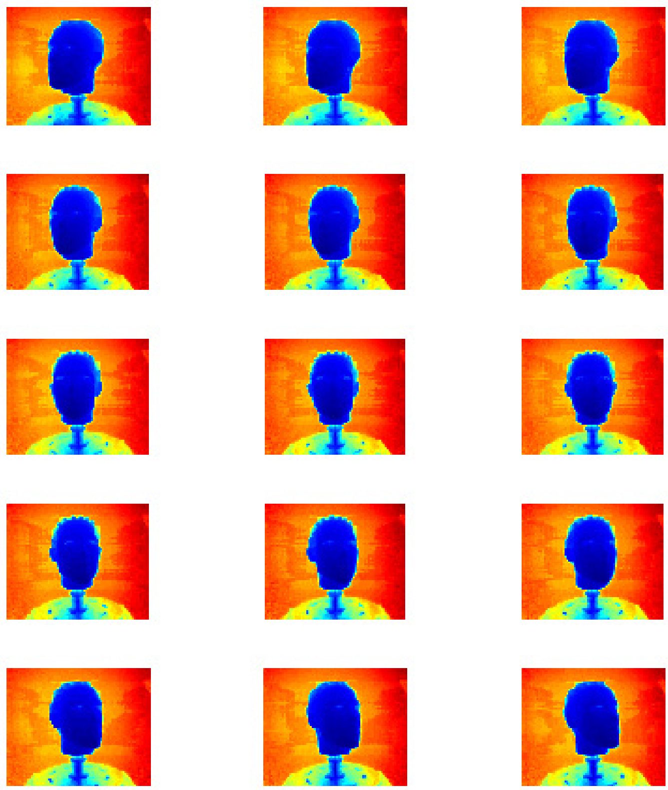

Figure C.1: Some sample depth images with about 5° rotation difference used in experiment.

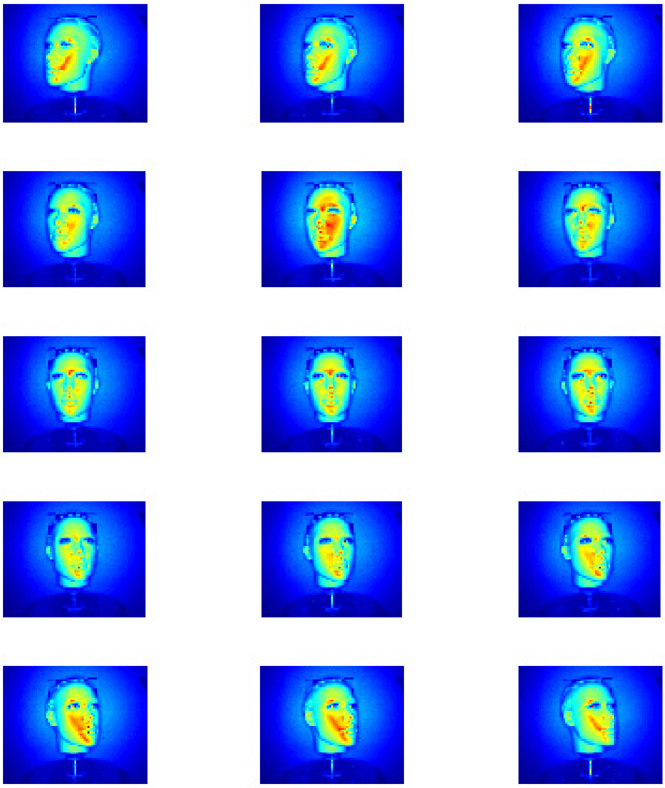

Figure C.2: Some sample intensity images corresponding to the depth images in Figure C.1 used in experiment

Bibliography

[1] *SwissRanger SR-3000 Manual*, 2005.

[2] Mutiple authors. *Handbook of Medical Image Processing*, volume 2. SPIE, 2000.

[3] J. Besag. On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society*, 48(3):259–302, 1986.

[4] P. Besl and N. McKay. A Method for Registration of 3-D Shapes. *Trans. PAMI*, 14(2), Feb. 1992.

[5] A. K. Jain C. Dorai, G. Wang and C. Mercer. Registration and Integration of Multiple Object Views for 3D Model Construction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):83–89, January 1998.

[6] D. P. Capel. Image Mosaicing and Super-resolution. Technical report, 2001.

[7] Y. Chen and G. Medioni. Object Modeling by Registration of Multiple Range Images. In *Proc. IEEE Conf. On Robotics and Automation*, 1991.

[8] B. Curless and M. Levoy. A Volumetric Method for Building Complex Models from Range Images. In *ACM Processings of Siggraph*, pages 303–312, 1996.

[9] J. Diebel and S. Thrun. An Application of Markov Random Field to Range Sensing. In *Proceedings of Conference on Neural Information Processing Systems (NIPS), MIT Press, Cambridge, MA*, 2005.

[10] M. Levoy etc. The Digital Michelangelo Project: 3D Scanning of Large Statues. In *Proc. SIGGRAPH'94, ACM*, 2000.

[11] R. Fabio. From Point Cloud to Surface: The Modeling and Visuralization Problem. In *International Archives of the Photogrammery, Remote Sensing and Spatial Information Sciences*, volume XXXIV-5/W10, pages 24–28, Tarasp-Vulpera, Switzerland, February 2003. Workshop on Visualization and Animation of Reality-based 3D Models.

[12] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, second edition, 2002.

[13] B. K. P. Horn. Closed-form Solution of Absolute Orientation Using Unit Quaternions. *Optical Society of America A*, 4(4):629–642, April 1987.

[14] J. L. Bentley J. H. Friedman and R. A. Finkel. An Algorithm for Finding Best Matches in Logarithmic Expected Time. *ACM Trans. Math. Software*, 3:209–226, 1977.

[15] V. Krishnamurthy and M. Levoy. Fitting Smooth Surfaces to Dense Polygon Meshes. In *Proc. SIGGRAPH'96*, pages 313–324, August 1996.

[16] S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, Tokyo, 1995.

[17] K. H. Møller. 3D Object Modelling via Registration of Stereo Range Data. Master's thesis, Technical University of Denmark, Kongens Lyngby, 2006.

[18] D. Cohen-Or S. Fleishman D. Levin M. Alexa, J. Behr and C. T. Silva. Computing and Rendering Point Set Surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 9(1):3–15, January-March 2003.

[19] M. K. Ng and N. K. Bose. Mathematical Analysis of Super-Resolution Methodology. *IEEE Signal Processing Magazine*, pages 62–74, May 2003.

[20] T. Oggier and F. Lustenberger. CSEM Technical report: Artts. Technical report, CSEM, Febuary 2006.

[21] J. Park and G. N. DeSouza. 3D Modeling of Real-World Objects Using Range and Intensity Images. *Innovations in Machine Intelligence and Robot Perception, ©Springer-Verlag*, 2005.

[22] K. Pulli. Multiview Registration for Large Data Sets. In *Second International Conference on 3D Digital Imaging and Modeling (3DIM'99)*, pages 160–168, Ottawa, Canada, October 1999.

[23] J. Cryer R. Zhang, P. Tsai and M. Shah. Shape from Shading: A Survey. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 28(8), Auguest 1999.

[24] R. Ramamoorthi and J. Arvo. Creating Generative Models from Range Images. In *Proc. SIGGRAPH'99*, pages 195–204, August 1999.

[25] S. M. Rusinkiewicz. *Real-time Acquisition and Rendering of Large 3D Models*. PhD thesis, Stanford University, August 2001.

[26] J. R. Shewchuk. An Introduction to the Conjugate Gradient Method Without the Agonizing Pain, Augest 1994.

[27] M. Lehmann T. Oggier and R. Kaufmann etc. An All-solid-state Optical Range Camera for 3D Real-time Imaging with Sub-centimeter Depth Resolution (SwissRanger). Technical report.