

Copyright
by
Wurong Yu
2008

**The Dissertation Committee for Wurong Yu Certifies that this is the approved
version of the following dissertation:**

**Development of a Three-Dimensional Anthropometry System for
Human Body Composition Assessment**

Committee:

Bugao Xu, Supervisor

Alan C. Bovik

Kenneth R. Diller

Stephen D. Hursting

Mia K. Markey

**Development of a Three-Dimensional Anthropometry System for
Human Body Composition Assessment**

by

Wurong Yu, B.S., M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2008

Dedicated to my parents and Shujuan

Acknowledgements

It has been a long journey towards the completion of my Ph.D. study. I could have never reached the destination without the support and assistance of many individuals.

I would like to express my sincerest gratitude to my advisor, Dr. Bugao Xu, for the sound guidance and consistent support he has provided me throughout this research. His trust, encouragement and understanding helped me overcome many hard times.

I am deeply indebted to my committee members, Dr. Alan C. Bovik, Dr. Kenneth R. Diller, Dr. Stephen D. Hursting and Dr. Mia K. Markey. Without their valued advice and constructive criticism, the completion of this dissertation would not have been made possible.

I am very grateful to Dr. Jeanne H. Freeland-Graves and her students for experimental design and implementation. Particularly, Reese Pepper assisted me in writing the IRB proposal, recruiting subjects and collecting data. Brenna Wozniak and Katherine Bontrager also helped conduct the study. Special thanks go to the anonymous volunteers for their contribution to this study.

I am very thankful to my current and former colleagues, Ming Yao, Xun Yao, Qingguan Li, Yan Wan, Yaxiong Huang, Yueqi Zhong and Tong Chen, for discussions and help. My heartfelt thanks also go to friends in Paul and Judith's Sunday classes for their friendship and encouragement.

I would like to thank my parents for bearing, raising, shaping and loving me. Likewise, I am grateful to my parents-in-law for their love and moral support. I miss all my family members in China, and wish I could have spent more time with them. Last but not least, I would like to thank my wife, Shujuan, for her love, patience, support and understanding.

Development of a Three-Dimensional Anthropometry System for Human Body Composition Assessment

Publication No. _____

Wurong Yu, Ph.D.

The University of Texas at Austin, 2008

Supervisor: Bugao Xu

The prevalence of obesity has made it necessary to develop a convenient, reliable and safe tool for timely assessing and monitoring this condition in public health. We suggest that three-dimensional (3D) anthropometry can provide a convenient, accommodating and comprehensive means to body composition assessment.

A 3D anthropometry system based on stereo vision technology is developed. To make it more portable and affordable, the system is reduced to a two-stance design and only uses off-the-shelf components. The system is calibrated in two separate stages: camera calibration and 3D registration. The first stage is relatively complicated, but there is no need to repeat frequently. Therefore, only 3D registration is required when the system is transported. This property contributes to the portability and also reduces cost of maintenance. In this system, image acquisition can be completed in 200 ms, which is important in reducing artifacts caused by slight body movement.

However, the computation in stereo vision is complex and intensive, and is still a challenge. A two-phase stereo matching algorithm is developed. In the first phase, the

foreground is accurately segmented with the help of a predefined virtual interface, and a coarse disparity map is generated with block matching. In the second phase, local least squares matching is performed in combination with global optimization within a regularization framework, so as to ensure both accuracy and reliability.

To make the 3D data more interpretable and manageable, it is essential to convert the raw 3D data to a surface model. For our system, a unique challenge is that there are large gaps in the data caused by occlusions. An effective surface reconstruction algorithm based on subdivision surface representation is developed. It has been verified that the algorithm is reliable in gap closing, accurate in representation, and efficient in data compression.

To make the 3D anthropometry system ready for practical use, a body measurement system dedicated to body composition assessment is developed based on an earlier system that was designed for applications in apparel fitting. The functions of 3D measurement are enhanced by taking advantage of modern graphics hardware. The overall performance of the presented system has been evaluated on mannequins and human subjects. It has been shown that the measurements are highly repeatable. The feasibility of 3D anthropometry in body fat assessment has been demonstrated in comparison to air displacement plethysmography and bioimpedance analysis.

Table of Contents

List of Tables	xi
List of Figures	xii
Chapter 1 Introduction	1
1.1 Goals and Motivation.....	1
1.2 Structures and Contributions of the Dissertation	3
Chapter 2 Background	5
2.1 Introduction.....	5
2.2 Overweight and Obesity	5
2.2.1 Classification and Epidemiology	5
2.2.2 Health Risks	7
2.3 Current Methods for Body Composition Assessment	9
2.3.1 Body Composition Models	9
2.3.2 Underwater Weighing.....	11
2.3.3 Air Displacement Plethysmography	12
2.3.4 Anthropometry	13
2.3.5 Bioimpedance Analysis	13
2.3.6 Dual-Energy X-Ray Absorptiometry.....	14
2.3.7 Computed Tomography and Magnetic Resonance Imaging.....	14
2.3.8 Comparison of Methods to Assess Body Composition	14
2.4 3D Anthropometry for Body Composition Assessment	15
2.5 Conclusion	17
Chapter 3 Framework Design of a Stereo Vision System	18
3.1 Introduction.....	18
3.2 Basic Principles of 3D Surface Imaging.....	22
3.2.1 Laser Scanning.....	22
3.2.2 Structured Light	23
3.2.3 Stereo Vision.....	25

3.2.4 Comparison of Optical Triangulation Techniques.....	28
3.3 System Setup.....	29
3.4 System Calibration.....	32
3.4.1 Camera Calibration	32
3.4.2 3D Registration	35
3.5 Discussion	37
Chapter 4 Stereo Matching	39
4.1 Introduction.....	39
4.2 Review on Stereo Matching Algorithms.....	40
4.2.1 Matching Primitives.....	40
4.2.2 Matching Strategies	41
4.3 Methodology	44
4.3.1 Overview	44
4.3.2 Matching Cost.....	44
4.3.3 Foreground Segmentation.....	45
4.3.3.1 Definition of the Energy Function	46
4.3.3.2 Virtual Interface.....	49
4.3.3.3 Energy Minimization	54
4.3.4 Disparity Refinement.....	56
4.5 Results.....	59
4.6 Discussion	60
Chapter 5 Surface Reconstruction	66
5.1 Introduction.....	66
5.2 Related Work	66
5.3 Methodology	68
5.3.1 Data Resampling.....	69
5.3.2 Initial Mesh Generation	71
5.3.3 Mesh Simplification.....	74
5.3.4 Mesh Subdivision and Optimization.....	76
5.4 Results.....	80

5.5 Discussion	83
Chapter 6 Body Measurement and System Evaluation.....	85
6.1 Introduction.....	85
6.2 Body Measurement	85
6.2.1 Related Work	85
6.2.2 3D Measurement Using Graphics Hardware	86
6.2.2.1 Volume Measurement.....	87
6.2.2.2 Circumference Measurement	88
6.2.2.3 Area Measurement.....	89
6.3 System Evaluation	91
6.3.1 Subjects and Methods	91
6.3.1.1 Mannequins and Measurements.....	91
6.3.1.2 Human Subjects and Measurements	92
6.3.1.3 Statistical Analysis.....	92
6.3.2 Results.....	93
6.3.2.1 Mannequins.....	93
6.3.2.2 Human Subjects	95
6.4 Discussion	102
Chapter 7 Conclusions and Future Work.....	104
7.1 Summary of the Dissertation	104
7.2 Suggestions on Future Work.....	106
Bibliography	107
Vita.....	120

List of Tables

Table 2.1:	Classification of overweight and obesity by BMI, waist circumference and associated disease risk in Caucasian adults.....	8
Table 3.1:	Comparison of three types of optical triangulation technique.	28
Table 4.1:	Planes of the effective imaging volume defined in the world coordinate system.	53
Table 4.2:	Parameters for the test of stereo matching.	60
Table 5.1:	Data sizes and running times for the presented models.	82
Table 6.1:	Repeatability test on three mannequins.	94
Table 6.2:	Longitudinal repeatability test on the size-12 mannequin.	95
Table 6.3:	Dimensions of the size-12 mannequin measured by manual methods and the 3D anthropometry system.	95
Table 6.4:	Characteristics of the human subjects.....	96
Table 6.5:	Repeatability test on 20 human subjects.....	96
Table 6.6:	Comparison of dimensions and volume measured by 3D anthropometry, and tape, anthropometer or ADP in human subjects.....	97
Table 6.7:	Linear regression analysis on dimensions and volume measured by 3D anthropometry, and tape, anthropometer or ADP in human subjects.....	97
Table 6.8:	Bland-Altman analysis on percent body fat.....	102

List of Figures

Figure 2.1: The prevalence of overweight and obesity in the U.S. Reprinted from [18].	6
Figure 2.2: Body composition at the molecular level. Adapted from [30].	10
Figure 2.3: The relationship of 3D anthropometry with BMI, densitometry and traditional anthropometry.	16
Figure 3.1: Schematic illustration of laser scanning.	22
Figure 3.2: Triangulation geometry in laser scanning. Adapted from [87].	23
Figure 3.3: Schematic illustration of structured light.	24
Figure 3.4: Schematic illustration of stereo vision.	25
Figure 3.5: Basic principle of stereo vision.	26
Figure 3.6: Triangulation geometry in parallel-axis stereo vision.	27
Figure 3.7: Schematic illustration of the system setup.	30
Figure 3.8: The prototype stereo vision system.	31
Figure 3.9: A set of images for camera calibration.	34
Figure 3.10: The target for 3D registration.	37
Figure 4.1: Partition of the disparity space by a virtual interface.	47
Figure 4.2: The effective imaging volume of the proposed stereo vision system.	50
Figure 4.3: A 3D plane induces a homography between the image planes in stereo vision.	51
Figure 4.4: Graph construction.	55
Figure 4.5: Virtual interface defined in the disparity space for the (a) front-top, (b) front-bottom, (c) back-top, and (d) back-bottom stereo heads.	61

Figure 4.6: Results on a mannequin for the front-top stereo head. (a) Rectified image pair; (b) foreground segmentation and coarse disparity map; (c) refined disparity map; and (d) the refined disparity map is overlaid onto the reference image.62

Figure 4.7: Results on a mannequin for the front-bottom stereo head. (a) Rectified image pair; (b) foreground segmentation and coarse disparity map; (c) refined disparity map; and (d) the refined disparity map is overlaid onto the reference image.63

Figure 4.8: Results on a human subject for the front-top stereo head. (a) Rectified image pair; (b) foreground segmentation and coarse disparity map; (c) refined disparity map; and (d) the refined disparity map is overlaid onto the reference image.64

Figure 4.9: Results on a human subject for the front-bottom stereo head. (a) Rectified image pair; (b) foreground segmentation and coarse disparity map; (c) refined disparity map; and (d) the refined disparity map is overlaid onto the reference image.65

Figure 5.1: Original scan data of a mannequin in the anterior (a) and lateral (b) views. Data points from the front (back) stereo heads are in red (green).68

Figure 5.2: Smooth merge of the upper and lower data sets in each view.71

Figure 5.3: Triangulation of the leg. (a) Resampled and padded (in red) data points. (b) The front and back meshes are merged by tiling the modified contours (in blue). Frontal projections of the modified front and back contours, corresponding to the region marked with a dashed green box in (a), are shown in (c) and (d), respectively.73

Figure 5.4: Generation of the initial mesh. Shaded models of the mesh before (a) and after (b) closing the gaps.	73
Figure 5.5: Edge collapse. A new vertex \bar{v} is created by collapsing the edge $\overline{v_i v_j}$	74
Figure 5.6: A simplified mesh (a) and its shaded model (b).	76
Figure 5.7: Mesh subdivision by edge split.	77
Figure 5.8: Illustration of Loop's evaluation rule for a vertex point.	78
Figure 5.9: Illustration of the modified Loop's evaluation rule for an edge point.	78
Figure 5.10: An optimized control mesh (a), and its subdivision mesh (b) and the shaded model (c).	80
Figure 5.11: Reconstructed body models of subjects with various shapes and sizes.	81
Figure 5.12: Performance of mesh optimization.	82
Figure 5.13: Close-up views of gap and hole filling: (a) the armpit; and (b) the foot.	83
Figure 6.1: Body segmentation.	87
Figure 6.2: Body volume measurement. (a) A 3D body model rendered in the anterior view; and (b) its thickness map.	88
Figure 6.3: Measurements on a contour. (a) The contour is marked on the body model. (b) Circumference, breadth and depth measurements. (d) Cross-sectional area measurement.	89
Figure 6.4: Illustration of the body measurement system. (a) Results on two subjects. (b) Body segments used for the abdomen-hip volume measurement.	90
Figure 6.5: Automatic measurement on a mannequin model.	94

Figure 6.6: Scatter plot of body volume measured by 3D anthropometry (3D-A) and air displacement plethysmography (ADP).....	98
Figure 6.7: Scatter plot of percent body fat (%BF) by 3D anthropometry (3D-A) and air displacement plethysmography (ADP).....	99
Figure 6.8: Scatter plot of percent body fat (%BF) by bioimpedance analysis (BIA) and air displacement plethysmography (ADP).....	100
Figure 6.9: Bland-Altman plot of percent body fat (%BF) by 3D anthropometry (3D-A) and air displacement plethysmography (ADP).	100
Figure 6.10: Bland-Altman plot of percent body fat (%BF) by bioimpedance analysis (BIA) and air displacement plethysmography (ADP).	101
Figure 6.11: Bland-Altman plot of percent body fat (%BF) by 3D anthropometry (3D-A) and bioimpedance analysis (BIA).	101

Chapter 1

Introduction

1.1 GOALS AND MOTIVATION

The importance of human body composition research has increased due to the prevalence of obesity. Obesity, a chronic disease characterized by an abnormally high proportion of body fat, has emerged rapidly as a global epidemic [1]. The spread of obesity is not restricted to Western industrialized nations, but also occurring across large parts of the developing world. The major concern of obesity is its associated health conditions. It has been well established that overweight and obesity increase the risk of some serious diseases including Type 2 diabetes, hypertension, coronary heart disease, stroke, sleep apnea, osteoarthritis, gallbladder disease, and certain cancers [2, 3]. Obesity is also associated with an increased risk of premature death [4].

The main aim of assessing body composition in obese is to evaluate the proportion of body fat. Obesity can usually be recognized by the overall appearance of the body, which means the size and shape of the body provide rich information on the degree of adiposity and associated health risk. For example, the body mass index (BMI), which is calculated by dividing body weight (in kilograms) by height (in meters) squared, is a widely used indicator of body fatness, and adopted by the World Health Organization (WHO) to classify obesity [1]. However, we need to realize that BMI is a crude measure of total body fat, because the relationship between BMI and fatness varies with muscularity, age, gender and ethnicity [2]. In addition to the total amount of fat, the distribution of fat is also an important factor in accessing health risk. It is believed that excessive abdominal adipose tissue is associated with increased risk of cardiovascular

disease and insulin resistance [5, 6]. Thus, with the same BMI, “apple-shaped” obese individuals have a higher risk of metabolic disorders than “pear-shaped” ones. For abdominal obesity, waist circumference gives a better predictor than BMI [7].

To improve screening for obesity, various techniques have been used for more direct measurements of the amount and distribution of body fat. For instance, Densitometry methods including underwater weighing [8] and air displacement plethysmography [9] are accepted as “gold standards” for body density estimate, but their accuracy in predicting percent body fat is limited by the two-component (fat and fat-free mass) model. Additionally, underwater weighing can be time consuming and uncomfortable for the subject. Bioelectrical impedance analysis [10] is a rapid, safe and inexpensive method to estimate body fat, but its accuracy depends on the hydration status of the subject, and dehydration will result in overestimate of fat mass. Dual-energy X-ray absorptiometry [11], X-ray computed tomography [12] and magnetic resonance imaging [13] are more sophisticated and reliable techniques, but their significant expense and low portability make them almost exclusive to medical setting and clinical research. Furthermore, X-ray is ionizing radiation and thus poses the potential risk of cancer from repeated scans. Details of these techniques will be reviewed in Chapter 2.

Considering the prevalence of obesity, a convenient, reliable, safe, and relatively inexpensive device is necessary for timely assessing and monitoring fatness in public health. Recent studies have shown that three-dimensional (3D) body surface imaging is a potential alternative to assess body fat or predict risk of metabolic syndrome [14]. This kind of imaging device, commonly called a body scanner, captures the surface geometry of the human body by using non-contact optical techniques. Theoretically, body scanning can take the place of densitometry since the body volume and thus the body density can be readily obtained from 3D body surface data. Nevertheless, body scanning is more

convenient and accommodating compared to current densitometry techniques. In addition, the variables measured in body scanning are not limited to the whole body volume, and actually, numerous other anthropometric measures can be extracted automatically from a 3D model, such as waist and hip circumferences, sagittal abdominal diameter, segmental volumes, body surface area, and so on. Therefore, body scanning can be regarded as a comprehensive anthropometry technique. We refer to body measurement performed on 3D data as 3D anthropometry.

Body scanning technologies are maturing, but the application of 3D anthropometry for body composition assessment is still in its infancy. The reason is multifold. First, body scanning has been widely used in other areas such as clothing and animation industries [15, 16], but its potential and value have not been realized by most body composition researchers. Secondly, various body scanners are commercially available, but their high price and bulky size make them inappropriate for field studies. Finally, software systems capable of body composition assessment are rarely available. For this reason, there is a need to promote 3D anthropometry for body composition research. As a step towards this direction, the purpose of this study was to develop a portable, relatively inexpensive 3D anthropometry system that would be readily accessible to body composition researchers.

1.2 STRUCTURES AND CONTRIBUTIONS OF THE DISSERTATION

The remainder of the dissertation is divided into five chapters. Chapter 2 provides background for this research. Obesity and its classification criteria as well as the associated complications are reviewed. Current body composition techniques for fat assessment are explored. Then the potential and advantages of 3D anthropometry for body composition research are discussed.

Chapter 3 describes the design of a 3D body scanning system after reviewing current major techniques of 3D surface imaging. The system is based on stereo vision technology. To reduce the cost of hardware, it is made up of inexpensive off-the-shelf components including digital cameras, projectors, and a personal computer. The design takes portability into account. A two-stage system calibration method is presented, which involves camera calibration and 3D registration.

The major challenge of the system is to compute 3D surface points from acquired images. The problem is called stereo matching and is addressed in Chapter 4. A novel two-step stereo matching algorithm is presented. In the first step, the foreground object is segmented and a rough disparity map is created simultaneously. In the second step, the disparity map is refined within a regularization framework to reach sub-pixel accuracy.

Chapter 5 deals with the problem of surface reconstruction. A subdivision surface-based algorithm is presented to reconstruct a smooth 3D surface model of the human body from incomplete and noisy data.

In Chapter 6, a body measurement system is developed to estimate the size and shape of the reconstructed human body and use the measures to assess body fat. Numerous anthropometric parameters can be extracted automatically. The parameters include lengths, circumferences, whole body volume and surface area, and segmental volumes. A pilot study is carried out to test the accuracy and precision of the system.

Chapter 7 offers a conclusion and discussion of future study.

Chapter 2

Background

2.1 INTRODUCTION

For the first time in human history, the population of overweight is greater than the population of malnutrition in the world [17]. Thus, it is not surprising that body fat assessment has become the primary focus of body composition researchers. Body fat assessment plays an important role in weight management and obesity monitoring in public health. In this chapter, we first give a brief review on overweight and obesity and associated risks of health complications. Then we examine current methods of body fat assessment. Finally, we propose 3D anthropometry as a potential alternative technique for body fat assessment.

2.2 OVERWEIGHT AND OBESITY

2.2.1 Classification and Epidemiology

Since precise measure of body fatness is difficult, body mass index (BMI) has been traditionally used as a surrogate to classify overweight and obesity. According to the criteria recommended by the World Health Organization (WHO) [1], a BMI of at least 25 kg/m² is defined as overweight, and a BMI of at least 30 kg/m² as obesity. BMI has been used as a general guide to monitor trends in the population. The prevalence of obesity is high and continuing to increase in industrialized countries. Take the U.S. as an example. In 2003-2004, among adults age 20–74 years, 67% of Americans are overweight, and 34% are obese, according to data from the National Health and Nutrition Examination Survey (NHANES) [18], as shown in Figure 2.1. The data also indicate that the prevalence of obesity has increased significantly over last three decades considering that

only 15% of adults were obese in the period of 1976 to 1980, while the overweight but not obese population is relatively stable. The proportion of overweight has also increased significantly in children since the 1970s. With economic growth and urbanization, the prevalence of overweight and obesity is soaring rapidly in the developing world as well, and has joined malnutrition and infectious diseases as major public health problems. Today, among the world population of 6.6 billion, more than 1.1 billion adults are overweight and 312 million of them are obese, and 10% of children are overweight or obese [19].

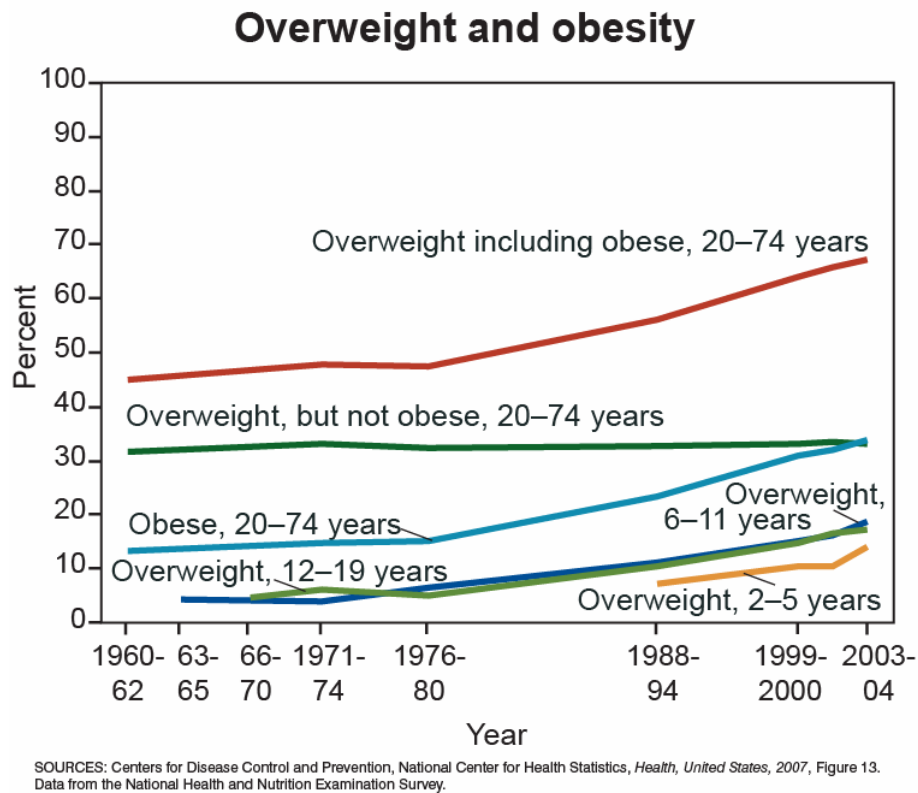


Figure 2.1: The prevalence of overweight and obesity in the U.S. Reprinted from [18].

2.2.2 Health Risks

Essential fat is important for maintaining normal functions of the body, but excess fat caused by undesirable positive energy balance can increase health risks for numerous diseases including hypertension, Type 2 diabetes and coronary heart disease [2, 3, 20-26]. In addition, the distribution of fat is as important as total fat amount in determining the risks associated with obesity. It has been well established that excess abdominal fat is a strong, independent risk factor for metabolic complications [5, 6]. The measurement of waist circumference provides a simple and convenient method for assessing abdominal obesity (also called central obesity) and associated risks [7]. But populations differ in the amount of abdominal fat at a particular waist circumference, so a universal cutoff point for risk prediction does not exist. The sex-specific cutoff points adopted by WHO are 102 cm for men and 88 cm for women in Caucasians [1]. Different criteria should be used for distinct racial groups. The relative disease risks associated with BMI and waist circumference are listed in Table 2.1, which is used in the clinical guidelines released by the National Institutes of Health (NIH) [2]. NIH uses the same classification system of overweight and obesity as WHO does, except it refers to overweight as a BMI between 25.0 and 29.9 kg/m² while WHO's definition of overweight includes obesity.

We give a brief review of obesity-related health consequences, which range from increased risk of premature death to serious chronic diseases that reduce overall quality of life. Here details of underlying pathophysiologic mechanisms are ignored since they are beyond the scope of this dissertation. The risk of premature death elevates with the increasing level of BMI in obese people. The risk of mortality increases by 30% at a BMI above 30 kg/m², and by at least 100% at a BMI above 40 kg/m², compared to people with a healthy weight [27]. Each year, about 300,000 deaths in the U.S. are related to obesity

Table 2.1: Classification of overweight and obesity by BMI, waist circumference and associated disease risk in Caucasian adults.

	BMI (kg/m ²)	Obesity Class	Disease risk* relative to normal weight and waist circumference	
			Men ≤ 102 cm Women ≤ 88 cm	> 102 cm > 88 cm
Underweight	< 18.5		—	—
Normal†	18.5–24.9		—	—
Overweight	25.0–29.9		Increased	High
Obesity	30.0–34.9	I	High	Very high
	35.0–39.9	II	Very high	Very high
Extreme obesity	≥ 40	III	Extremely high	Extremely high

* Disease risk for Type 2 diabetes, hypertension, and cardiovascular disease.

† Increased waist circumference can also be a marker for increased risk even in persons of normal weight.

Adapted from [2].

[27, 28]. Obesity, especially central or visceral obesity, is strongly associated with increased insulin resistance and glucose intolerance, which are precursors to Type 2 diabetes. It has been estimated that over 80% of people with diabetes are overweight or obese [27]. Overweight individuals are particularly prone to have hypertension and hyperlipidemia, which can lead to coronary heart disease and stroke. It has been estimated that more than 85% of hypertension cases arise in individuals with overweight or obesity [29]. Additionally, obesity is also associated with an increased incidence of respiratory problems (sleep apnea and asthma), reproductive complications, and some types of cancer including endometrial, colon, gall bladder, prostate, kidney, and postmenopausal breast cancer [2, 29].

2.3 CURRENT METHODS FOR BODY COMPOSITION ASSESSMENT

BMI plays an important role in screening and monitoring overweight and obesity in adults at population level, but it should be used with caution when evaluating an individual's adiposity [2]. BMI may not reflect actual body fat, because it cannot distinguish body fat from muscle mass. In addition, the relationship between BMI and body fatness varies between men and women, but the classification based on BMI is not sex-specific. Further, a given BMI may not correspond to the same adiposity among different racial groups. Therefore, to better assess health risks associated with obesity, body composition methods should be used to estimate body fatness more directly.

2.3.1 Body Composition Models

Prior to investigating current techniques for body composition assessment, we first examine body composition models.

The central model in body composition research is the five-level model in which all components of the body mass are organized at each of the five levels—atomic, molecular, cellular, tissue-organ, and whole body [30]. Within the five levels, the molecular level is especially important since it is the basis of various methods for body composition assessment. Body composition at this level is described in Figure 2.2. At the molecular level, major components of the body mass include fat, water, protein, carbohydrates, and minerals. Total body water can be further divided into extracellular and intracellular water, and minerals can be further classified into soft tissue minerals and bone minerals. The sum of all components except fat is called fat-free mass. The remainder is conventionally called lean soft tissue when the component of bone minerals is separated from fat-free mass.

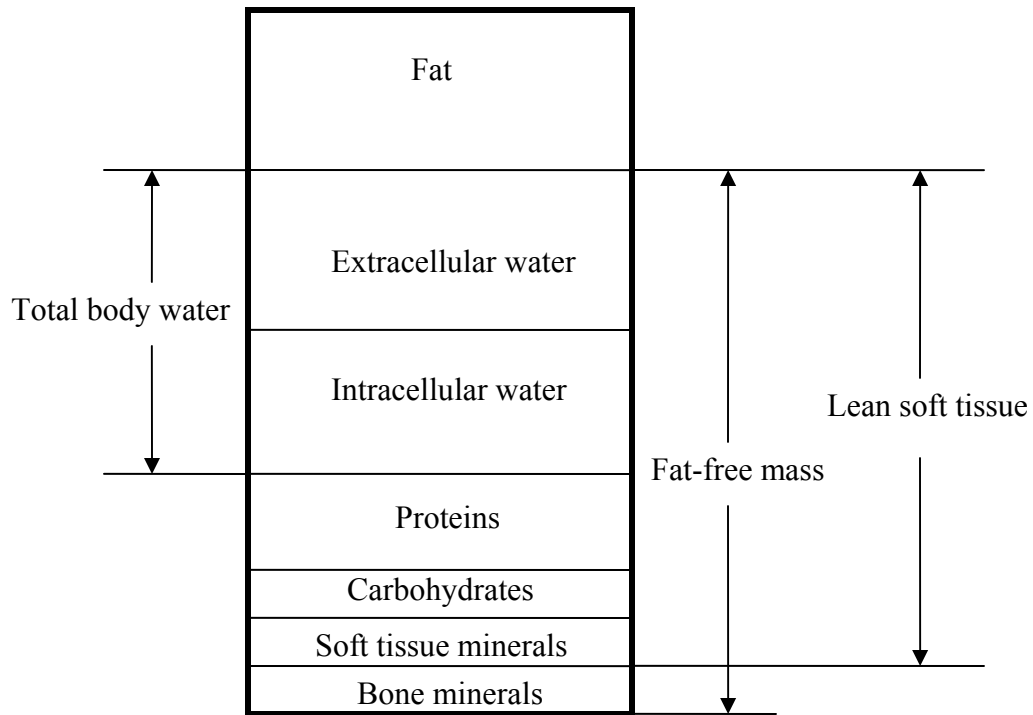


Figure 2.2: Body composition at the molecular level. Adapted from [30].

At this point, the definition of fat should be emphasized. Fat, lipid and adipose tissue are sometimes used interchangeably, but they are distinct [31]. In body composition studies, fat specifically refers a group of molecules called triglycerides. Yet lipids encompass not only triglycerides, but also fatty acids and their derivatives as well as cholesterol. So the definition of lipid is broader than that of fat. Fat is also different from adipose tissue; the latter is a concept at the tissue-organ level. The major component of adipose tissue is fat, but adipose tissue is also made up of other molecules including proteins, minerals, water and nonfat lipids. On the other hand, although fat exists

primarily in adipose tissue, it is also found somewhere else such as liver and skeletal muscle.

2.3.2 Underwater Weighing

Underwater weighing (UWW) [8], also called hydrodensitometry (HD) is a means to estimate body density (D_b) by measuring total body volume from the water displaced by the body when it is fully submerged. According to Archimedes's principle, the weight of displaced water is equal to the loss of body weight in water. Thus, the body volume, which equals the volume of displaced water, can be calculated by

$$BV = (W_a - W_w) / D_w, \quad (3.1)$$

where W_a and W_w are the body weight in air and water, respectively. The effect of lung volume should be accounted. If the underwater weight is measured after a maximal expiration, then the residual lung volume (RV) should be measured. RV can be measured simultaneously with UWW using the oxygen dilution technique [32]. Additionally, the volume of gas in the gastrointestinal tract (V_{GI}) is taken into account. V_{GI} is small and usually approximated with a constant value (100 ml) [33]. Then the body density is calculated as

$$D_b = \frac{W_a}{BV - (RV + V_{GI})}. \quad (2.2)$$

If we assume the body is composed of fat and fat-free mass (FFM) and their densities are constant, then the percentage of body fat (%BF) can be calculated by

$$\%BF = \left(\frac{A}{D_b} - B \right) \times 100, \quad (2.3)$$

where A and B are derived from the assumed fat density (D_F) and FFM density (D_{FFM}). D_F is relatively stable and usually set as 0.9 kg/L. But slightly different values of D_{FFM} appear in the literature. In Siri's equation [34],

$$\%BF = \left(\frac{4.95}{D_b} - 4.50 \right) \times 100, \quad (2.4)$$

where $D_{FFM} = 1.1 \text{ kg/L}$ is used. The other commonly used equation was proposed by Brozek [35],

$$\%BF = \left(\frac{4.570}{D_b} - 4.412 \right) \times 100. \quad (2.5)$$

A limitation of the two-component-based equations is that they assume the D_{FFM} remains a constant, but in fact there is considerable interindividual variation [31]. The principle of UWW is simple, but this technique is not convenient, is stressful to the participant, and is difficult for certain populations such as small children, very old or sick people to perform.

2.3.3 Air Displacement Plethysmography

Recently, air displacement plethysmography (ADP) [9], commercially available as the BodPod (Life Measurements Instrument, Concord, CA), has been introduced as an alternative to UWW for body volume measurement. The BodPod is a closed dual-chamber plethysmography filled with air. While the subject is sitting in the test chamber, a moving diaphragm between the two chambers oscillates to induce pressure perturbations in both chambers, and then the air volume of the test chamber is determined by measuring changes in pressure. The body volume of the subject can be calculated by subtraction of the chamber volume when it is empty. However, additional corrections are required for the body volume since air close to skin and in the thoracic channel (under isothermal conditions) is more compressible than the rest of air in the chamber (under adiabatic conditions). The thoracic gas volume can be measured using a breathing tube attached to the BodPod or estimated by sex-specific prediction equations [36]. The

artifact caused by air close to skin is adjusted by estimating the body surface area from the Dubois formula [37].

Once body volume is obtained, body density can be calculated, and percent body fat can be estimated from the same equations as used in UWW. ADP is more convenient and accommodating than UWW, but the expense limits its use.

2.3.4 Anthropometry

The skinfold (SKF) method [38, 39] assumes there are interrelationships among subcutaneous fat, internal fat, and whole body density. Percentage body fat is estimated by measuring skinfold thickness at several sites.

Anthropometry [40-43] predicts percentage body fat from circumferences, segment lengths and, in some cases, skeletal diameters, in addition to height and weight. SKF measures can be used in combination with anthropometry, but the predictive accuracy of anthropometric equations is not greatly improved by adding SKF measures. Moreover, anthropometric equations using only circumferences estimate body fatness more accurately than SKF in obese adults [44].

2.3.5 Bioimpedance Analysis

Bioimpedance analysis (BIA) [10, 45, 46] applies the principle of contrast of electric impedance of different tissues. Lean soft tissue is a good conductor because it contains a large amount of water and dissolved electrolytes, whereas fat and bone have relatively poor conductance properties. BIA usually works at a single frequency (50 Hz), but also at multiple frequencies [47]. FFM or percentage body fat is estimated using population-specific prediction equations that take into account measured impedance as well as body weight, height and age. BIA is relatively inexpensive, portable and easy to

use, but it is sensitive to hydration status and body geometry (in which water distribution differs), thus limiting its usefulness in severe obesity [48].

2.3.6 Dual-Energy X-Ray Absorptiometry

Dual-energy X-ray absorptiometry (DEXA) [11, 49] is based on the fact that the attenuations of X-ray through bone, lean soft tissue, and fat are different, and the differences in the attenuation properties for tissues decrease with increasing photon energy. The amount of different tissues can be estimated by the transmittance rates of X-ray at dual energies, typically 40 and 70 keV [11]. It is sometimes considered as a three-component method. The image of a whole-body scan is analyzed by a computer to estimate the proportions of bone, fat and lean soft content in two steps [50]. First, the proportions of bone and soft tissue are derived at pixels that contain bone. Secondly, pixels containing bone are excluded to differentiate fat and lean soft mass.

2.3.7 Computed Tomography and Magnetic Resonance Imaging

X-ray computed tomography (CT) [12, 51] and magnetic resonance imaging (MRI) [13, 52] estimate body composition by imaging internal body structures. CT measures the differences in the attenuation of X-ray through different tissues, whereas MRI measures the abundance of hydrogen nuclei in different tissues. CT and MRI are capable of measuring both total body fat and regional fat distribution. One of the important applications of CT and MRI is in abdominal obesity research since they can effectively differentiate visceral adipose tissues from subcutaneous adipose tissues [31].

2.3.8 Comparison of Methods to Assess Body Composition

Depending on whether energy intervention with tissues in the body is involved, the aforementioned techniques can be classified into two types. The first type includes HD, ADP, SKF and anthropometry, where no energy intervention is used. The second

type includes BIA, DEXA, CT and MRI, which apply electromagnetic energy such as electric field, X-ray, or magnetic field. Among Type II techniques, DEXA, CT and MRI provide relatively direct measurements of percentage body fat and fat distribution and are the most reliable methods. However, limited availability and high cost make them impractical for widespread use. On the other hand, BIA devices are portable and inexpensive, but their accuracy has been questioned. Among Type I techniques, HD and ADP are accepted as “gold standards” for measuring D_b , and they give similar accuracy in predicting percentage body fat, where the accuracy is mainly limited by the two-component body model. SKF and anthropometry are widely used in clinic and field studies because they are very simple and inexpensive, but their accuracy depends on the prediction equations.

2.4 3D ANTHROPOMETRY FOR BODY COMPOSITION ASSESSMENT

The 3D anthropometry technique developed in this work is closely related to Type I methods. Based on current development in computer vision technology, the 3D surface model of a person in tight undergarments can be reconstructed from several images captured by digital cameras [53-55]. Then anthropometric variables can be automatically extracted from the 3D model by using dedicated algorithms. The measures include height, circumferences, segment lengths, total volume, segmental volumes and surface areas. Wells et al. [14] offered a review of the potential of 3D anthropometry in obesity research. The technique will provide a more efficient, more objective and more comprehensive means to body measurement than conventional tape anthropometry. Additionally, it is totally non-contact and non-invasive. Further, the 3D model is reusable, so new measures can be extracted if they are required by an anthropometrist later on. It also covers densitometry because body volume can be calculated from the 3D model, but it is more convenient and accommodating than HD and AP techniques [56,

57]. SKF is not considered here because a 3D model cannot give any SKF information, but we note that it has been reported that anthropometric prediction equations are better than SKF equations for estimating body composition of individuals who are obese [44]. 3D measurement can be regarded as a generalized anthropometric technique. Its relationship with BMI, densitometry and conventional anthropometry is shown in Figure 2.3.

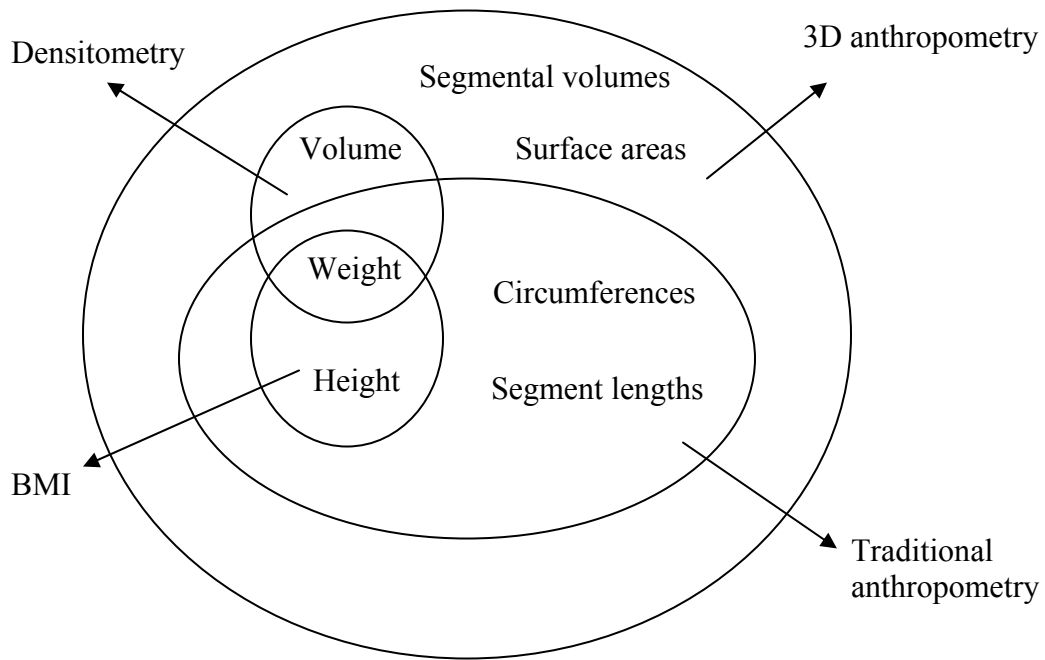


Figure 2.3: The relationship of 3D anthropometry with BMI, densitometry and traditional anthropometry

The proposed technique can also be integrated into some Type II methods and will improve their prediction outcome. For example, in current BIA, it is assumed that the body is composed of several cylindrical components and their lengths are estimated from body height. Obviously, this treatment is oversimplified. But if we take into account the actual body size and shape with 3D measurement, a more accurate circuit model can be established and thus the prediction equations can be improved. In another example, the general procedures of DEXA need measures of body height, weight, and sagittal abdominal diameter (SAD) [58]. SAD is a measure of anteroposterior thickness of the abdomen at the umbilical level. Research indicates that SAD is an excellent indirect measure of visceral fat [59, 60]. With 3D measurement, SAD can be easily obtained. Far beyond that, the thickness over the whole body can be computed from a 3D model. If the thickness map of the body can be registered with its DEXA image, it can be expected that more accurate body fat distribution will be predicted.

2.5 CONCLUSION

The prevalence and associated health risks of obesity have been discussed in this chapter. Some general concepts related to methods for body composition assessment have been reviewed. We have also demonstrated the prospects of 3D anthropometry as a comprehensive technique for obesity assessment.

In the following chapters, a 3D anthropometry system based on stereo vision technology will be presented.

Chapter 3

Framework Design of a Stereo Vision System

3.1 INTRODUCTION

In this dissertation, 3D anthropometry is defined as the study of human body measurement using 3D surface imaging techniques. 3D anthropometry has been explored since the 1950s. Hertzberg et al. [61] first realized the potential of close-range photogrammetry as an anthropometric tool. In the 1970s, researchers worked on the Skylab of NASA [62, 63] developed a photogrammetric system to detect the changes in body size and shape of astronauts after a flight. The method was called biostereometric analysis. Later on, Sheffer and others [64, 65] applied biostereometrics to breast volume measurement. However, these earlier photogrammetric systems used film cameras and analysis was not automatic. After development, the films were manually analyzed on a stereo-plotter to get 3D data points. The manual process was cumbersome and time-consuming.

With the development of semiconductor and computing technologies, fully automated body surface imaging systems have emerged since the 1980s. These systems are commonly called body scanners. One of the earliest body scanners was delivered by Cyberware [66] for head scanning. The scanning unit is composed of a laser line projector and a camera. It is rotated around 360 degrees to capture a 3D image of the subject's head in about 15 seconds. Later, Cyberware also developed a whole body model that consists of four vertically moving scanning units. The LASS scanner is one of the first whole body scanners [67]. It uses white light projection instead of laser projection. Four vertical lines are projected onto the subject simultaneously and the images are

captured by multiple cameras. The system is rotated horizontally to cover the whole body in about 60 seconds. To reach whole body scanning, an alternative means is to make the subject rotate continuously over 360 degrees while keeping the scanning unit still [68].

The body scanner commercially available from TC² is based on structured light projection [69]. The configuration of the original version forms a triangle and is composed of six sensors. The latest model NX16 consists of 16 sensors and every four sensors are stacked at each corner of the cubic scanning booth to cover partial of the body. Each sensor consists of a projector and a camera. The projector casts a sinusoidal stripe pattern onto the body. The stripes of the pattern are shifted and multiple images of the scene are acquired. The phase at each pixel can be calculated from these images and thus depth is determined. A complete scan takes about eight seconds. Similar principles have been applied in other systems, such as the body scanners from InSpeck [70]. Compared to laser scanning, the most significant advantage of structured light is that no moving components are involved in the system since mechanical scanning process has been avoided. The major challenge of structured light is that ambiguity can arise if the phase difference between the adjacent pixels is larger than 2π , which usually occurs at discontinuities of the surface [71]. To resolve the ambiguity, multiple patterns of various levels of stripe density are usually used.

The other category of body scanners is based on the same principle of photogrammetry as we mentioned before, but the computation of 3D data is fully automatic. For example, Siebert and Marshall [72] described a photogrammetric system called C3D with applications in facial, head and breast imaging. 3D data are calculated with a multi-resolution correlation-based image matching algorithm. Human skin is not rich in texture, which would make image matching fail, so artificial speckle texture is projected onto the body with a white light projector to facilitate image matching. A

similar strategy has been utilized in the photogrammetric systems commercialized by the 3dMD company [73]. The strength of this technique is that image acquisition can be as fast as within several milliseconds.

Today, numerous body scanners have appeared on the market, the majority of which are based on laser scanning and structured light technologies. According to a recent report [74], there are over 20 companies around the world that are delivering whole body scanners. Some review articles offer more detailed discussion of body scanning technologies. Daanen and van de Water [55] provided an overview of eight whole body scanning systems. Istook and Hwang [15] reviewed major body scanning systems with focus on applications in apparel industry.

3D body scanning encompasses a wide range of applications, such as virtual try-on and mass customization in the clothing industry [15, 74-76], animation in movies and computer games [16, 77], breast and facial imaging for the planning and evaluation of plastic surgery [78, 79], and body surface area measurement for the determination of medication dosage [80]. Body scanning is convenient, rapid, and non-contact, so it is suitable for population studies. For example, the CAESER (Civilian American and European Surface Anthropometry Resource) project has been undertaken with a goal to design better fitting clothes, more comfortable seats in public transportation and more ergonomic cars [81]. It was an international anthropometric survey joined by the U.S., Canada, Italy and Netherlands, and scanned 8,000 subjects. The SizeUK study scanned approximately 11,000 subjects in the U.K. with an intention to improve the sizing system [82]. A similar study called SizeUSA is being carried out in the U.S. to obtain scan data of at least 12,000 American adults [83]. For a large-scale study, the use of body scanning can greatly reduce cost and save time when compared to conventional manual measurement [84].

The application of body scanning is still unappreciated as a 3D anthropometry method in body composition study. But its success in sizing survey for the clothing industry has inspired researchers to consider its potential applications in body composition—especially, obesity—research [14]. However, to make body scanning accessible to public health, there is a need to develop less expensive and more portable body scanners that are suitable for routine use in clinical settings and health clubs. First, the cost of whole body scanners has dropped dramatically, but is still relatively high. For example, the NX16 model from TC² costs \$40,000 [85], and the WBX model from Cyberware costs \$200,000 [86]. A body scanner is still a big investment for small business. Secondly, portable body scanners are available at present, but they are only capable of scanning part of the body, such as head and torso, so the portability of whole body scanners still needs improvement. Portability means not only a small size, but also the ease of disassembling and assembling as well as the ease of calibration and maintenance. Finally, most of current body scanners only output 3D data points, and do not provide software for body surface reconstruction and measurement. Even though some systems offer built-in functions of body measurement, they are usually designed for clothing industry. For body composition analysis and health risk assessment, some special measures may be required. For example, body volume measure is rarely considered in apparel design, but it is important in body fat assessment.

In this chapter, we describe the framework of a stereo vision system we have developed for 3D anthropometry. Following a brief review of basic principles underlying current body scanning techniques, we will present the setup and calibration of our system.

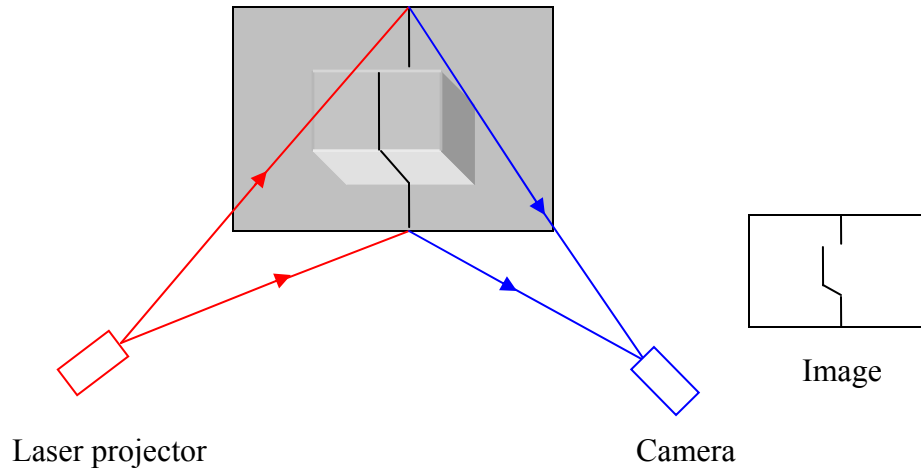


Figure 3.1: Schematic illustration of laser scanning.

3.2 BASIC PRINCIPLES OF 3D SURFACE IMAGING

3.2.1 Laser Scanning

A laser scanning sensor [87, 88] typically includes a laser projector and a camera. As shown in Figure 3.1, the laser projector emits a laser beam (plane of light) onto the object, and a stripe is formed by intersection of the plane of light with the object surface. The image of the laser stripe is captured by a camera. The deformation of the profile determines the depth of surface points. The calculation of 3D coordinates is illustrated in Figure 3.2. The distance between the laser projector and the camera, b , is called the baseline length. If we assume the focal length of the camera is f , and the tilt angle of the projector is θ , then the 3D coordinates of an object point $P(X, Y, Z)$ can be derived from its image coordinates $p(x, y)$ using similarity of triangles [87],

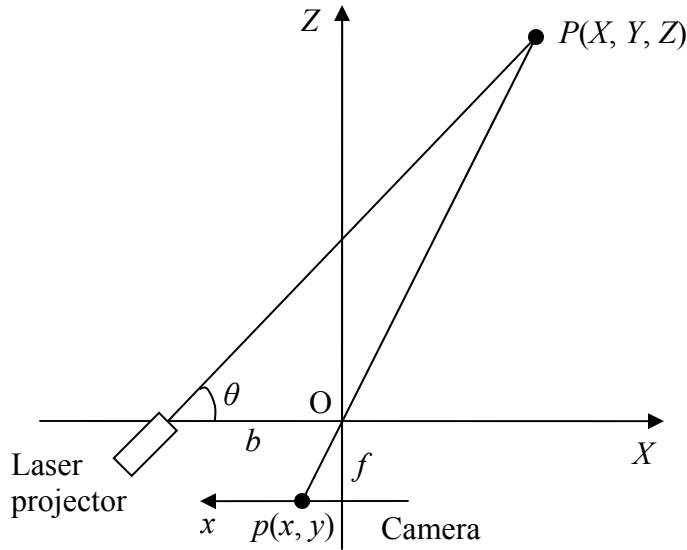


Figure 3.2: Triangulation geometry in laser scanning. Adapted from [87].

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \frac{b}{f \cot \theta - x} \begin{pmatrix} x \\ y \\ f \end{pmatrix}. \quad (3.1)$$

The principle is called optical triangulation.

In laser scanning, in order to cover the full range of an object, a mechanical scanning process is needed to sweep the laser stripe across the object. It can be realized by moving the whole sensor with a step motor or redirect the laser beam with a rotating mirror.

3.2.2 Structured Light

As opposed to laser scanning where only a single profile of the surface can be obtained at each time, we can project multiple stripes simultaneously onto the object to get denser data. This is called structured light [89], as illustrated in Figure 3.3. But

multiple stripes would cause ambiguity, to reduce which a sequence of different stripe patterns should be used. In Figure 3.3, a binary pattern of eight stripes is used. If we use 0 to represent a black stripe, and 1 to a white stripe, then we can represent the pattern shown in Figure 3.3 as 01010101. If we project sequentially two additional stripe patterns with bits 00110011 and 00001111 respectively, then we can find that each stripe can be coded with a unique binary number. In this case, only three patterns are needed to code eight stripes. In general, $\log_2 n$ binary patterns are required to code n stripes. An alternative code is called Gray code that is more robust than binary code for structured light analysis [90]. For the simple example of eight stripes, the three patterns with Gray coding are $\begin{bmatrix} 01100011 \\ 00111001 \\ 00001111 \end{bmatrix}$. We can see that successive numbers of the Gray code differs exactly in one bit, so wrong decoding with one bit misread will only cause a misplacement of one stripe interval.

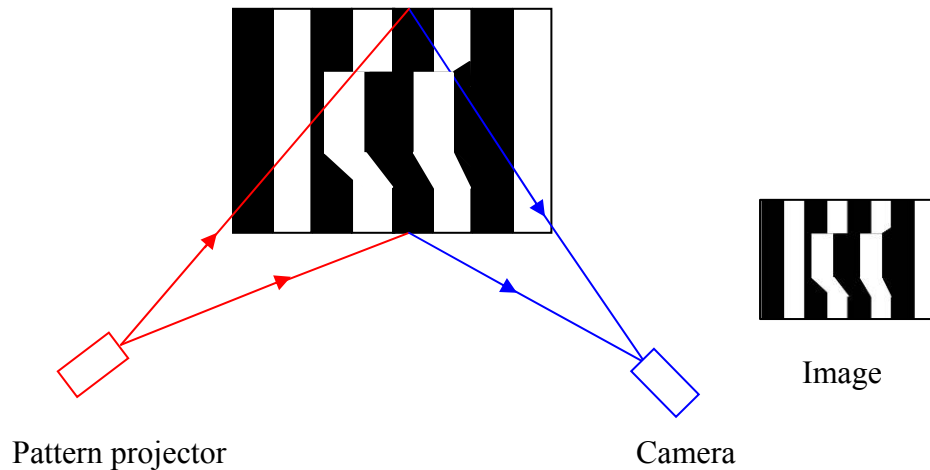


Figure 3.3: Schematic illustration of structured light.

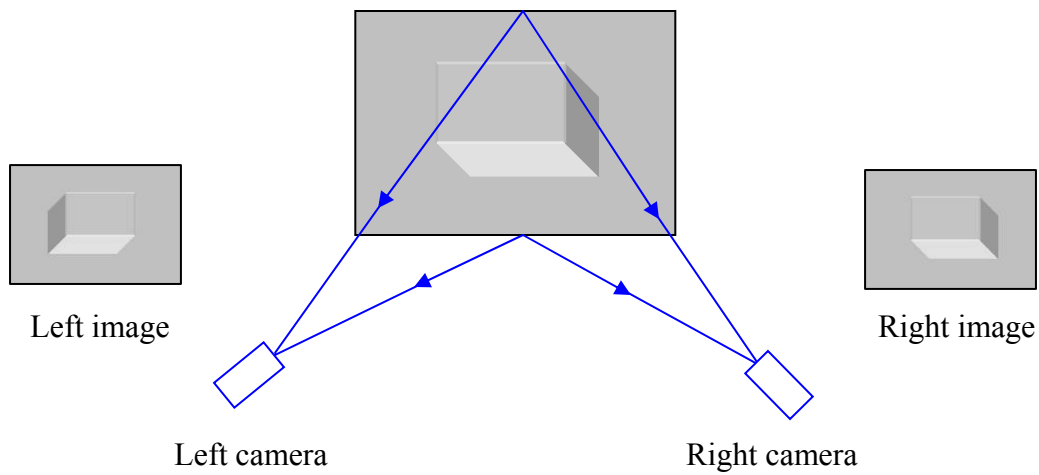


Figure 3.4: Schematic illustration of stereo vision.

For binary or Gray coded pattern projection, the resolution is limited by the number of stripes. To overcome this issue, the patterns can be replaced by or used together with sinusoid patterns of shifting phases [91]. Often, patterns at different levels of resolution are projected. Coarse patterns are used to reduce ambiguity and fine patterns to increase accuracy.

Structured light follows the same triangulation principle as in laser scanning except that the tilt angle θ in Equation (3.1) is obtained from the order of stripes or the phase.

3.2.3 Stereo Vision

Stereo vision [92, 93] works similar in concept to human binocular vision. As shown in Figure 3.4, a second camera is used to replace the projector in structured light. Since the two cameras observe the object from slightly different views, the captured left

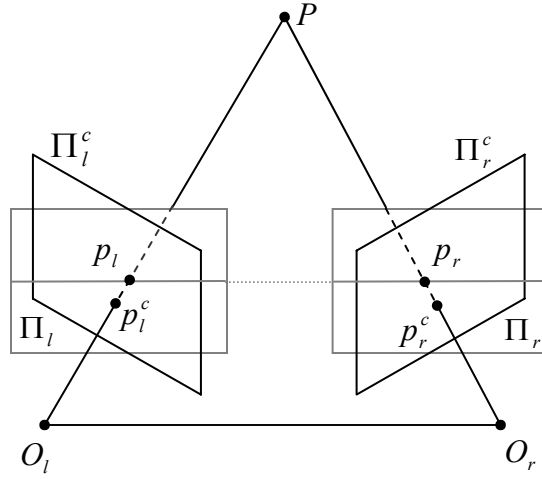


Figure 3.5: Basic principle of stereo vision.

and right images are not exactly the same. The relative displacement of the object in the two images is called the disparity, which is used to calculate the depth. It is not surprising that the underlying principle is still optical triangulation. The basic principle is shown in Figure 3.5. As a convention adopted in this dissertation, we use subscripts “ l ” and “ r ” to denote the left and right camera, respectively. The position of an object point P can be determined by intersecting two rays $\overrightarrow{O_l p_l^c}$ and $\overrightarrow{O_r p_r^c}$, where O_l and O_r are called the principal points of the left and right cameras, and p_l^c and p_r^c are the projections of P in the corresponding image planes Π_l^c and Π_r^c . In practice, to simplify the computation, we can rectify the images by re-projecting to the specific image planes Π_l and Π_r which are parallel and equidistant to the baseline $\overline{O_l O_r}$. In this case, the new projections p_l and p_r of P are located in the same horizontal scanline. If the focal length f and the baseline length (distance between the two cameras) b are known, then the depth of P relative to the cameras can be obtained by triangulation based on the geometry illustrated

in Figure 3.6. If the origin of the 3D coordinate system is set at the left principal point O_l , then it follows

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = N \begin{bmatrix} x_l \\ y_l \\ f \end{bmatrix} = N \begin{bmatrix} x_r \\ y_r \\ f \end{bmatrix} + \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix}, \quad (3.2)$$

with a magnification coefficient

$$N = -\frac{b}{d} = \frac{b}{x_l - x_r}, \quad (3.3)$$

where $d = x_r - x_l$ is defined as disparity, and x_l and x_r are relative to the image centers C_l and C_r , respectively.

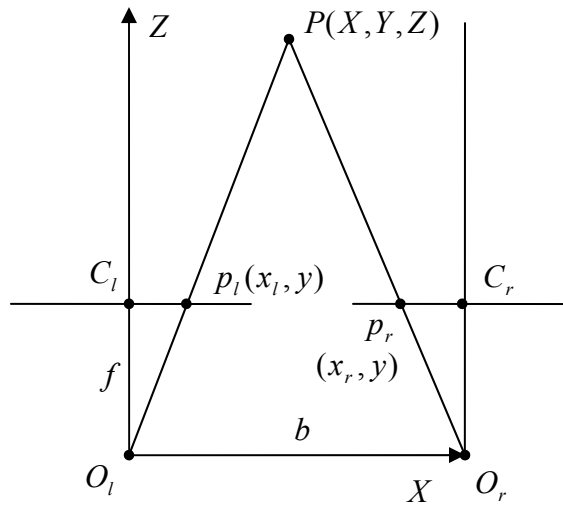


Figure 3.6: Triangulation geometry in parallel-axis stereo vision.

Table 3.1: Comparison of three types of optical triangulation technique.

	Laser scanning	Structured light	Stereo vision
System complexity	Moving part	Dedicated lighting	Relatively simple
Acquisition speed	Slow	Fast	Very fast
Computation	Easy	Intermediate	Intensive
Accuracy	Very good	Good	Good

3.2.4 Comparison of Optical Triangulation Techniques

We have examined three types of optical triangulation: laser scanning, structured light and stereo vision. They share the same fundamental principle: the 3D position of an object point is obtained by intersecting two optical rays. It is necessary to compare their strengths and limitations so as to choose a proper method for our system. The advantage of laser scanning is its high accuracy originated from the sharpness of laser stripes, although the accuracy is limited by speckle noise. However, laser scanning requires moving parts, so the system is more complex and expensive than the other two types. Additionally, a system involving a mechanical moving component is usually less durable and more difficult to maintain. Structured light needs dedicated lighting, which makes the hardware relatively more complex than stereo vision. In principle, stereo vision is a passive method, which means no light source is required. The major challenge in stereo vision is that the computation of disparity is hard for a textureless surface. Unfortunately, human skin is not rich in texture. Therefore, artificial texture projection is usually applied in a practical system. However, the texture projection does not need to be sophisticated patterns as in structured light. Data acquisition in stereo vision is very fast because only a

pair of images needs to be captured. In contrast, a sequence of images needs to be captured in synchronization with pattern projection in structured light. In the case of whole body scanning, the speed of image acquisition in structured light can be further slowed down, because multiple sensing units in the system cannot work simultaneously, otherwise pattern projections from different units may interfere with each other. Rapid data acquisition is critical to reduce artifacts caused by body movement. A slight body movement of a subject may induce unacceptable inaccuracy in some measures such as body volume. Based on the above analysis, we have chosen stereo vision for our 3D anthropometry system. A major disadvantage of stereo vision is the computation is complex and intensive. A summary of the comparison of the three types of optical triangulation technique is offered in Table 3.1.

3.3 SYSTEM SETUP

A prototype stereo vision system has been developed in this study. In this section, we describe its hardware design. The engineering factors of our major concern are cost, portability, and accuracy. To reduce the cost and shorten the duration of development, we have used off-the-shelf components including cameras and projectors. The basic unit of the system is a stereo head that consists of a pair of cameras and a projector. The projector is used to shed artificial texture onto the body. Multiple stereo heads are needed for full body imaging. The previous work on a rotary laser scanner indicates that full body reconstruction can be made from two scanning units that are placed in front and back of the subject, respectively [94]. A similar construction has been used in the study. However, two stereo heads are needed to cover each side of the body, due to the limited field of view of the cameras and projectors. Therefore, there are a total of four stereo heads in the system. The configuration is illustrated in Figure 3.7. The four stereo heads are mounted on two steady stands. Compared to some existing whole body scanners, our

system is more compact and portable. For instance, the NX16 body scanner from TC² consists of 16 scanning units [85].

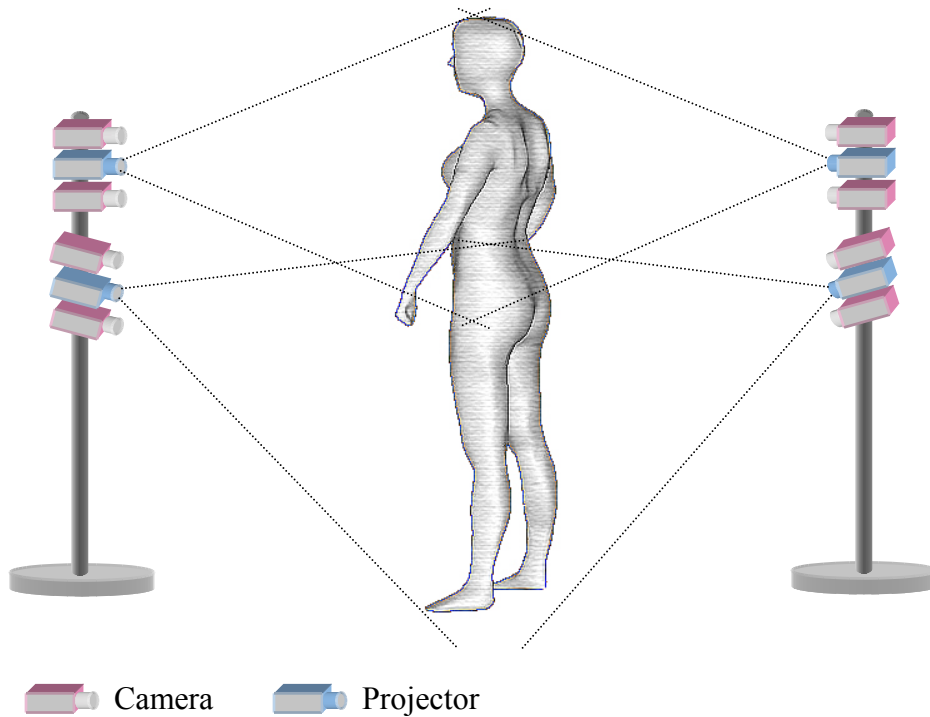


Figure 3.7: Schematic illustration of the system setup.

A more specific description of the system is given here. We used four pairs of monochromatic CMOS cameras (Videre Design, Menlo Park, CA) with a resolution of 1280×960 . The focal length of the cameras is 12 mm. The baseline length is set as 9 mm. The throw ratio was the most important specification to be considered when choosing the projectors. The throw ratio of a projector is defined as the projection distance divided by the image size. A shorter projection distance for a specific image size

(a smaller throw ratio) means the two stands can be put closer so as to reduce the size of the system. We used NEC 575VT LCD projectors (NEC Corp., Tokyo, Japan) since they were one of the few types of portable ultra-short throw projector on the market at the initiation of this project. At a projection distance of 2.3 m, the image size is 1.5 m \times 1.15 m. Hence, when two such projectors are used together with a slight overlap, the field of view can be as large as 1.5 m \times 2.0 m, which is large enough for the majority of population.



Figure 3.8: The prototype stereo vision system.

A personal computer with an AMD Athlon™ 2.0 GHz dual-core CPU and 1.0 G RAM is used to control the cameras and projectors. The cameras communicate with computer via IEEE 1394 Firewire. An NVIDIA GeForce 6500 dual-port graphics card is used to send a texture pattern to the projectors through a VGA hub (Gefen Inc., Woodland Hills, CA). All components are off-the-shelf and readily available.

3.4 SYSTEM CALIBRATION

System calibration includes two phases: camera calibration and 3D registration. In camera calibration, we need to estimate the internal parameters of the cameras and determine the relative position and orientation between the two cameras in each stereo head. In 3D registration, each camera coordinate system is registered to a common world coordinate system.

3.4.1 Camera Calibration

Camera calibration is a procedure of determining the intrinsic and extrinsic camera parameters. The intrinsic parameters correct the distortion induced in each individual camera by imperfect lens and lens displacement. The extrinsic parameters describe the position and orientation of the each individual camera in a reference coordinate system. Based on the extrinsic parameters of the two cameras of a stereo head, their relative position and orientation can be determined. The Small Vision System (SVS) shipped with the cameras provides standard procedures for camera calibration. Details of the implementation are not available, but the principles of camera calibration are described in the manual [95]. The intrinsic camera model is closely related to Tsai's and Heikkila and Silven's models [96, 97]. The intrinsic parameters include the effective horizontal and vertical focal lengths (f_x , f_y) of the lens, the principal point (c_x , c_y) which describes the decentering of the lens, the radial lens distortion coefficients (κ_1 , κ_2 ,

κ_3), and the tangential lens distortion coefficients (τ_1, τ_2). The description of the parameters is similar to that in Bouguet's Calibration Toolbox [98]. The extrinsic parameters can be described by a rotation matrix \mathbf{R} and a translation vector \mathbf{t} . It involves four steps of transformation from the 3D world coordinate to the 2D computer image coordinate.

1. Rigid body transformation from the 3D world coordinates (X_w, Y_w, Z_w) to the 3D camera coordinates (X, Y, Z) ,

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \mathbf{t}. \quad (3.4)$$

2. Transformation from the 3D camera coordinates (X, Y, Z) to the 2D undistorted, normalized image coordinates (x_u, y_u) using perspective projection with pinhole camera geometry,

$$\begin{bmatrix} x_u \\ y_u \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} X \\ Y \end{bmatrix}. \quad (3.5)$$

3. Transformation from the 2D undistorted image coordinates (x_u, y_u) to the 2D distorted image coordinates (x_d, y_d) ,

$$\begin{bmatrix} x_d \\ y_d \end{bmatrix} = \begin{bmatrix} x_u \\ y_u \end{bmatrix} + \begin{bmatrix} D_x^{(r)} + D_x^{(t)} \\ D_y^{(r)} + D_y^{(t)} \end{bmatrix}, \quad (3.6)$$

where $(D_x^{(r)}, D_y^{(r)})$ describes the radial lens distortion,

$$\begin{bmatrix} D_x^{(r)} \\ D_y^{(r)} \end{bmatrix} = (\kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6) \begin{bmatrix} x_u \\ y_u \end{bmatrix}, \quad (3.7)$$

and $(D_x^{(t)}, D_y^{(t)})$ describes the tangential lens distortion,

$$\begin{bmatrix} D_x^{(t)} \\ D_y^{(t)} \end{bmatrix} = \begin{bmatrix} 2\tau_1 x_u y_u + \tau_2 (r^2 + 2x_u^2) \\ 2\tau_2 x_u y_u + \tau_1 (r^2 + 2y_u^2) \end{bmatrix}, \quad (3.8)$$

where $r = \sqrt{x_u^2 + y_u^2}$.

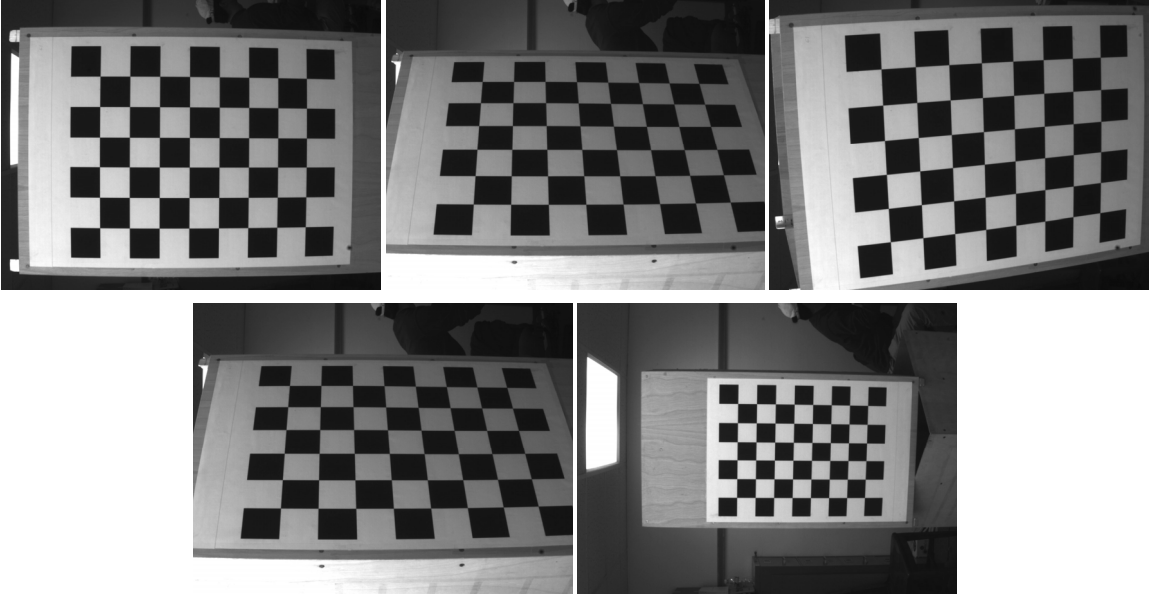


Figure 3.9: A set of images for camera calibration.

4. Transformation from the 2D distorted image coordinates (x_d, y_d) to the 2D digital image coordinates (x_f, y_f) ,

$$\begin{bmatrix} x_f \\ y_f \end{bmatrix} = \begin{bmatrix} f_x & 0 \\ 0 & f_y \end{bmatrix} \begin{bmatrix} x_d \\ y_d \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \quad (3.9)$$

where the effective focal lengths (f_x, f_y) and the principal point (c_x, c_y) are expressed in pixels.

To make a target for calibration, a checkerboard pattern was plotted and impressed to a planar board. The manufacture error was controlled under 0.5 mm. The checkerboard pattern includes 9×7 blocks, and the size of each block is 61 mm \times 61 mm. The 48 internal corners are used as feature points. Based on our experience, at least 5 images need to be taken in order to get good results. The target should be placed at

different positions and orientations. A set of images captured by one of the cameras are shown in Figure 3.9. The SVS system is capable of automatic feature extraction and calibration. The calibration error is around 0.15 pixels. The plane-based calibration technique was originally proposed by Zhang [99].

Once the cameras are calibrated, the image planes can be rectified to follow the parallel-axis stereo geometry as shown in Figure 3.6. The focal length f and the baseline length b in Equation (3.2) are determined as well.

3.4.2 3D Registration

The above camera calibration procedure is performed separately on each individual stereo head, and each stereo head has its own camera coordinate system. The goal of 3D registration is to transform each camera coordinate system to a common world coordinate system so that 3D data from each view can be merged. This transformation follows the rigid body model since it does not change the Euclidean distance between any points. To determine a rigid body transformation, three non-collinear points are sufficient. Let $\{\mathbf{X}_{c,i} \mid i = 1, 2, 3\}$ and $\{\mathbf{X}_{w,i} \mid i = 1, 2, 3\}$ be the coordinates of three non-collinear points in the camera and world coordinate systems, respectively. We are looking for a transformation of the form

$$\mathbf{X}_{w,i} = \mathbf{R} \mathbf{X}_{c,i} + \mathbf{t}, \quad (3.10)$$

from the camera to the world coordinate system, where \mathbf{R} is a rotation matrix, and \mathbf{t} is a translation vector. It should be clarified that we have abused the notations, and the \mathbf{R} and \mathbf{t} are different from that in Equation (3.4). Because there are measurement errors, we can only find an optimum solution that will minimize

$$\sum_{i=1}^3 \|\mathbf{X}_{w,i} - (\mathbf{R} \mathbf{X}_{c,i} + \mathbf{t})\|^2. \quad (3.11)$$

Horn [100] presented a closed-form solution to this problem, which is outlined as follows.

1. Refer the points to the centroids,

$$\mathbf{X}'_{c,i} = \mathbf{X}_{c,i} - \bar{\mathbf{X}}_c, \quad (3.12)$$

$$\mathbf{X}'_{w,i} = \mathbf{X}_{w,i} - \bar{\mathbf{X}}_w, \quad (3.13)$$

where $\bar{\mathbf{X}}_c$ and $\bar{\mathbf{X}}_w$ are the centroids of the points in the camera and world coordinate systems, respectively. Now the new centroids of the points are $\mathbf{0}$ in both coordinate systems.

2. Rotate the plane containing the points in the camera coordinate system to make it coincidence with the plane containing the points in the world coordinate system.

$$\mathbf{X}''_{c,i} = \mathbf{R}_1 \mathbf{X}'_{c,i}, \quad (3.14)$$

where \mathbf{R}_1 is the rotation matrix that can be determined from the normals of the two plane.

3. Find an in-plane rotation \mathbf{R}_2 that we wish to minimize

$$\sum_{i=1}^3 \|\mathbf{X}'_{w,i} - \mathbf{R}_2 \mathbf{X}''_{c,i}\|^2. \quad (3.15)$$

4. Then, it follows

$$\mathbf{R} = \mathbf{R}_2 \mathbf{R}_1, \quad (3.16)$$

and

$$\mathbf{t} = \bar{\mathbf{X}}_w - \mathbf{R} \bar{\mathbf{X}}_c. \quad (3.17)$$

We have designed a target for 3D registration. There are six dots on each side of the rig as show in Figure 3.10, and each stereo head needs only three of them. The world coordinates of the centers of the dots were manually measured with accuracy higher than

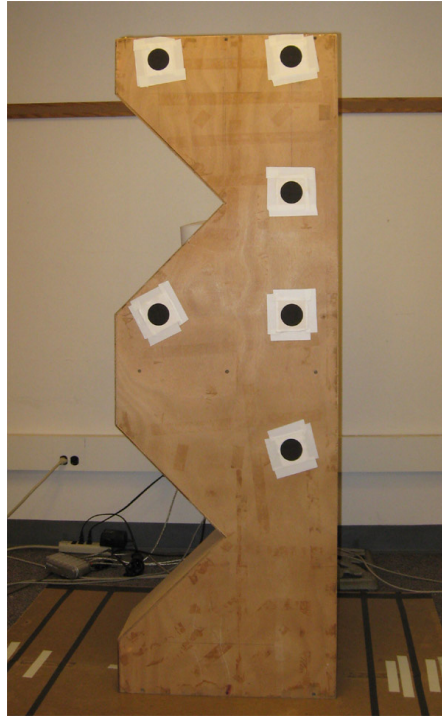


Figure 3.10: The target for 3D registration.

0.5 mm. The images of the target are first rectified, and then the dots are identified and sorted. Next, the centers of the dots are estimated. The camera coordinates of each point can be calculated by Equation (3.2).

3.5 DISCUSSION

The technology of 3D body scanning is maturing after over 20 years' development. However, there is still a need to make the technology more portable and affordable. In this study, a stereo vision-based body scanner has been developed in an effort to achieve this goal. The framework of the proposed body scanner has been described in this chapter. We have set up a prototype using off-the-shelf components. Compared to most commercial body scanners, the construction of our system is relatively

simple, since it doesn't need moving parts and dedicated lighting system. The system can be easily disassembled, transported and reassembled. A two-step system calibration method has been accomplished. The parameters of each stereo head can remain stable if locking lenses are used, so there is no need to repeat camera calibration frequently. Hence, we only need to redo the procedure of 3D registration when the system is relocated. The 3D registration only requires a very simple target and is easy to operate. This property improves the portability of the system and reduces the cost of maintenance. It is worthy noting that the size of the prototype is constrained mainly by the throw ratio of the projectors. However, we can expect that the system will take less room when projectors with shorter throw distance are available.

Chapter 4

Stereo Matching

4.1 INTRODUCTION

Stereo matching solves the correspondence problem in stereo vision [92, 93]. As the most challenging process in a stereo vision system, it has been intensively investigated over the last three decades. Lots of progress has been made in this area, but the research is still far from complete yet. On the one hand, there still doesn't exist a stereo matching algorithm that is general enough to be applicable under most circumstances. On the other hand, stereo vision is an application-oriented problem, and objectives and requirements may vary across different systems. Thus, for a specific application, we need to choose or develop algorithms that best suit its characteristics and requirements, which include the smoothness of the surface, the desired level of geometric details, the texture properties of the scene, the required density of data, and the time efficiency.

For the stereo vision system developed in this study, high accuracy is required for the purpose of 3D anthropometry. Based on the system setup described in the last chapter, a stereo matching algorithm with sub-pixel accuracy is needed to reach the quality of 3D data demanded by body measurement. Additionally, because the system has been designed to capture the front and back views of the body only, some portions of the body are invisible to the cameras. To deal with this issue, we have developed a surface reconstruction algorithm (to be presented in the next chapter) that is capable of filling in the gaps in 3D data caused by occlusions. However, if the boundaries of the body in each view cannot be accurately located, it will be difficult to recover the surface from

incomplete data. Therefore, in addition to high accuracy in matching, the algorithm should be able to accurately segment the body from the background. In this chapter, we present a sub-pixel stereo matching algorithm that can accurately recover the boundaries of foreground objects.

The remainder of the chapter is structured as follows. Section 4.2 provides a brief literature review on stereo matching algorithm. Section 4.3 describes details of the proposed algorithm. Experimental results are given in Section 4.4. We conclude in Section 4.5 with some discussions.

4.2 REVIEW ON STEREO MATCHING ALGORITHMS

A large number of stereo matching algorithms have been developed during the last three decades, but we only intend to conduct a brief survey of some representative methods to demonstrate some basic principles. Interested readers are referred to an early review [92] and more recent ones [93, 101].

Traditionally, stereo matching algorithms are classified into two categories, i.e., area-based and feature-based. But this classification has been insufficient to subsume many algorithms developed in recent years. Actually, area-based algorithms only include a small class of techniques which measure correlation between patches in the images. Instead, we will distinguish an algorithm from two aspects: matching primitives and matching strategy.

4.2.1 Matching Primitives

We divide matching primitives into two types: dense primitives and sparse primitives. Dense primitives are located at every pixel, and include such as image intensities [102], intensity gradients [103], phases or filter-bank responses [104-106], and so on. Sparse primitives are symbolic features extracted from intensity images, and

include points of interest [107], edge points [108], edge segments [109], and contours [110], to name a few. For dense primitives, local matching costs can be measured by squared differences, absolute differences or correlation. For sparse primitives, local matching costs are measured by simple comparisons between feature descriptors. The algorithms based on sparse primitives only produce sparse disparity maps, and thus interpolation is usually required as a post-process to get dense data fields. It is interesting to notice that most algorithms in the early review [92] are associated with sparse primitives, while most of the modern algorithms covered in the recent surveys [93, 101] are based on dense primitives. Such a trend can be explained by the fact that the early applications were focused on robot navigation which usually only needs depth information from surface contours, while nowadays applications in image-based rendering and scene synthesis are receiving more attentions.

4.2.2 Matching Strategies

Like many other inverse problems, stereo matching is under-constrained and ill-posed [111], which means that the solution from local matching between primitives is not unique. To obtain a physically plausible solution, a set of a priori assumptions have been proposed to impose constraints on the solution space. Some of the commonly used assumptions are summarized as follows.

1. Uniqueness constraint: Correspondence should be unique, which means that each primitive should have at most one match. But this is not valid for transparent objects.
2. Continuity constraint: The disparity map is assumed to be piecewise smooth except at surface boundaries.
3. Epipolar constraint: Matching primitives must lie on the corresponding epipolar lines that are coincident with scanlines for rectified images.

4. Ordering constraint: A left-right relationship of primitives in the left image should hold for their correspondences in the right image. But for narrow occlusions, this assumption may be violated.

It is essential to a matching algorithm to effectively utilize these constraints to achieve a globally consistent solution, which is called a matching strategy in this dissertation. We shall briefly introduce some typical matching strategies.

Perhaps correlation is the simplest matching strategy which has been used in what are traditionally called area-based methods. Such methods assume constant disparities within a window centered on a pixel in each image and measure similarity by correlation or cross-correlation of image intensities between windows. The disparity with the highest similarity will be selected for each pixel. It is critical to choose a suitable window size. A large window size is needed to make the solution robust to noise, but a small window size is preferred to produce a detailed disparity map. Thus, some methods [112, 113] employ a variable window size to adapt to local variations of intensity and disparity. It is obvious that disparities are measured locally in this strategy.

Cooperative algorithms are among the earliest methods motivated by Marr and Poggio's computational theory of human stereopsis [114]. In such algorithms, computations are performed by iteratively diffusing support among neighboring match values and inhibiting values along similar lines of sight under the continuity and uniqueness assumptions. A more globally consistent solution can be obtained by diffusing within a 3D local support area defined in the disparity space [115].

Stereo matching can also be formulated by minimizing global energy functions defined in the disparity space [116]. These energy functions usually include two terms, i.e., a data term and a smoothness term. The data term measures how well the disparity map agrees with the input images, while the smoothness term imposes the smoothness

constraint on the solution. But typically, these energy functions have many local minima, and to search for the global minimum is a combinatorial explosion problem which cannot be solved in practice. For example, simulated annealing [117] can be used to achieve the global minimum in theory, but its practical implementations are very slow. Many optimization techniques have been published to approximate the global minimum.

Although global optimization over the whole image is not practical for current computers, its implementation along a single scanline is practically feasible. Dynamic programming [118] is such a technique in which matching is performed by searching for a path with minimum cost in a matching space subject to ordering and uniqueness constraints, where the matching space is defined as a two-dimensional (2D) space with axes representing the corresponding left and right scanlines. But the major drawback of dynamic programming is that inter-scanline consistency cannot be guaranteed. Several techniques have been presented to alleviate this problem. For example, vertical discontinuities were taken into account by iterations [102]; a two-stage dynamic programming technique was reported in [119].

In contrast to 1D-optimization of DP, graph-cuts/maximum-flow [120-122] and belief propagation [123, 124] are two state-of-the-art approaches that perform optimization in 2D. In graph-cuts, the matching costs and smoothness constraints are associated to edges in a flow graphs as capacities, and stereo matching is converted into a maximum-flow/minimum-cut problem. In belief propagation (BP), stereo matching is formulated as a Bayesian inference problem where the maximum a posteriori (MAP) estimate is obtained by iterative message propagation in a network. For loopy networks such as Markov random fields, BP can only give approximate solutions [125], and Pearl's algorithm [126] is widely used. The results given by graph-cuts and BP are comparable,

and are more globally consistent than those by dynamic programming, but they have a much more extensive computation.

The genetic algorithm (GA) [127] is also a popular global optimization technique which simulates the mechanisms of natural evolution. In GA-based stereo matching algorithms [128, 129], disparity maps are coded as 2D chromosomes with fitness values inversely proportional to their energy. Then the set of chromosomes evolves with genetic operators including mating, chromosome crossover, gene mutation and natural selection. When the process converges, the best fit chromosome will be selected as the desired disparity map.

4.3 METHODOLOGY

4.3.1 Overview

In this study, the developed stereo matching algorithm involves two major phases. In the first phase, foreground objects are accurately segmented from the background of the scene, and meanwhile, a disparity map with integer-pixel accuracy is computed. In the second phase, the disparity map is iteratively refined to reach sub-pixel accuracy.

4.3.2 Matching Cost

In developing a matching algorithm, it is critical to choose a proper matching metric (similarity or dissimilarity measure between two corresponding pixels). Let $I_l(x, y)$ and $I_r(x, y)$ be the left and right intensity images, respectively, and the left image is taken as the reference image. At this point, we suppose the images are perfectly rectified, i.e., the epipolar lines coincide with the scanlines and disparities only exist in the horizontal direction. If we assume the surface is Lambertian and the images are corrupted with uniform white Gaussian noise, then the cost function (dissimilarity measure) associated with a match can be defined as

$$c(x, y, d) = (I_l(x, y) - I_r(x + d(x, y), y))^2, \quad (4.1)$$

which would give a maximum likelihood solution to the disparity function d .

But in practice, the above cost function may fail due to unbalanced exposure, gain and contrast of the camera pair as observed in our experiments. We thus will use a more robust method, normalized cross-correlation (NCC), as the similarity measure, which is defined as

$$\rho(x, y, d) = \frac{\sum_{(u,v) \in W(x,y)} (I_l(u, v) - \bar{I}_l(x, y))(I_r(u + d, v) - \bar{I}_r(x + d, y))}{N\sigma_l(x, y)\sigma_r(x + d, y)}, \quad (4.2)$$

where $W(x, y)$ is a correlation window around (x, y) with a total pixel number N , and $\bar{I}_l(x, y)$ ($\bar{I}_r(x, y)$) and $\sigma_l(x, y)$ ($\sigma_r(x, y)$) are the local mean and standard deviation of intensity for the left (right) image. The normalization in the local mean and standard deviation makes NCC insensitive to photometric distortions [119, 101]. Based on $\rho(x, y, d)$, the cost function can be defined by

$$C(x, y, d) = 1 - \rho(x, y, d). \quad (4.3)$$

Since $-1 \leq \rho(x, y, d) \leq 1$, we have $0 \leq C(x, y, d) \leq 2$. $C(x, y, d)$ is defined in the whole image space and at each possible disparity; this trivariate function is usually called the disparity space image (DSI) [93]. For the sake of conciseness, we will also denote the cost function as $C_p(d)$ with p being a pixel.

4.3.3 Foreground Segmentation

In this study, foreground segmentation is related to a class of matching algorithms called layered stereo [130-132], which has received attention lately because it is more effective in dealing with occlusions and discontinuities in the scene. Nevertheless, these existing methods almost exclusively rely on color segmentation. For our application, the natural appearance of the scene is eclipsed by the projection of artificial texture, which

makes it difficult to perform segmentation from color, contrast, or texture. However, we can take advantage of enhanced stereo cues, since artificial texture would reduce ambiguity in stereo matching.

4.3.3.1 Definition of the Energy Function

The problem of foreground segmentation can be formalized in the framework of energy minimization. Let P denote the pixel set of the reference image. We define $L = \{F, B\}$ as a label set with F and B representing the foreground and background, respectively. Then the goal is to find a segmentation (or labeling) $f(P) \mapsto L$ that minimizes an energy function $E(f)$ defined on a given stereo image pair I_l and I_r .

The energy function $E(f)$ usually consists of two terms [120],

$$E(f) = \sum_{i \in P} D_p(f_p) + \sum_{(p,q) \in N} V_{p,q}(f_p, f_q), \quad (4.4)$$

where $N \subset P \times P$ is the set of all neighboring pixel pairs. $D_p(f_p)$ is derived from the input images that measures the cost of assigning the f_p to the pixel p . $V_{p,q}(f_p, f_q)$ imposes the spatial coherence of the labeling between the neighboring pixels p and q .

Here we derive $D_p(f_p)$ from the disparity space image $C_p(d)$. First, we assume the disparity space can be divided into two subspaces: the foreground space and the background space that contain the object and the background, respectively, as shown in Figure 4.1. We assume there exists a virtual interface between the two subspaces, which is denoted by $d^*(P)$. Now we define $C_p^F = \min_{d_{\min} \leq d \leq d_p^*} C_p(d)$, $C_p^B = \min_{d_p^* < d \leq d_{\max}} C_p(d)$, and thus $C^F(P)$ and $C^B(P)$ represent the minimum surfaces in the foreground and background spaces, respectively. If $C_p^F < C_p^B$, then we can expect that there is a good chance that the pixel p belongs to the foreground. The same applies to $C_p^B < C_p^F$ and the background. Therefore, we can define $D_p(f_p)$ by

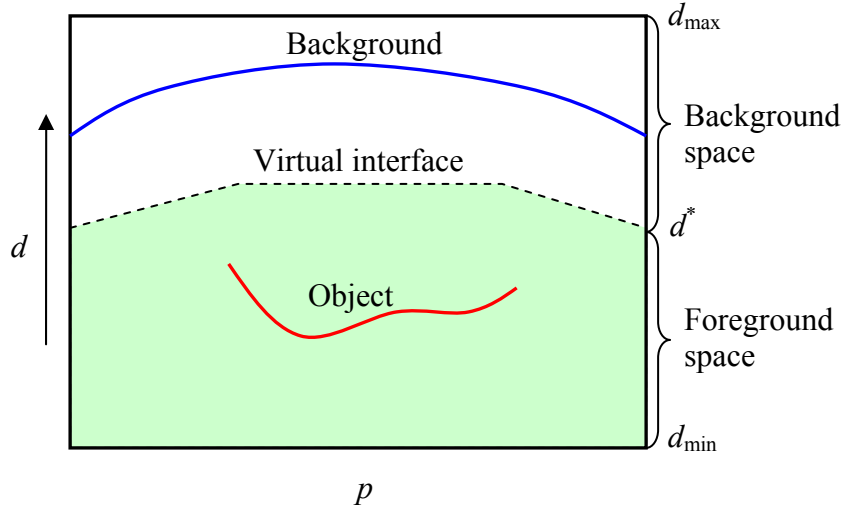


Figure 4.1: Partition of the disparity space by a virtual interface.

$$D_p(f_p) = \begin{cases} C_p^F, & f_p = F \\ C_p^B, & f_p = B \end{cases} \quad (4.5)$$

However, the above definition is invalid for pixels that cannot be matched. It usually occurs at occlusions, but can also happen in textureless regions that are usually caused by shadows in our system. For the unmatched pixels, we assign constants to the $D_p(f_p)$,

$$D_p(f_p) = \begin{cases} C_O^F, & f_p = F \\ C_O^B, & f_p = B \end{cases} \quad (4.6)$$

Here we set $C_O^B < C_O^F$ to favor the background, since we assume that occlusions and shadows exist in the background.

Now the problem becomes to compute the disparity space image $C_p(d)$, to determine the virtual interface $d^*(P)$, and to detect unmatched pixels. The computation

of the $C_p(d)$ is straightforward and can be expedited by the box filtering [119] or running sum algorithm [133], both of which has a time complexity that is independent of the size of matching window.

In most cases, the $d^*(P)$ is not available, since we usually lack the prior knowledge about the structure of the scene. But fortunately, for our body imaging system, the virtual interface can be well defined based on the system construction, which will be described in the next subsection. For the moment, we assume that the $d^*(P)$ has been determined.

To detect unmatched pixels, we use some conventional methods based on block matching. In block matching, the disparity for each pixel is obtained by searching the minimum in the DSI, i.e.,

$$d_p = \arg \min_d C_p(d) , \quad (4.7)$$

which is equivalent to searching the correlation peak according to Equation 4.3. However, false matches can occur, because disparities are undefined at occlusions, and matching also may fail in other regions due to image noise, geometric distortion, or insufficient texture. We will take the false matches as unmatched. Three criteria are used for deciding a good match. First, the variation of intensity in the matching window should be above a threshold σ_t , otherwise the definition of NCC (and thus the matching cost) is unstable. Secondly, the correlation value should be greater than a threshold ρ_t . Thirdly, the match should pass the left-right check, which means it is also related to a correlation peak if we take the right image as the reference. There is a tradeoff in setting the parameters σ_t and ρ_t : the larger they are, the more confident we are in decided good matches, but the chance of missing good matches will also increase. Ideally, σ_t should be set above the noise level of image, and ρ_t should be determined by such factors as noise level, degree of perspective distortion, size of matching window, and accuracy of

image rectification. But in practice, it is hard to optimize these parameters by incorporating the above-mentioned factors, so in our experiments, they are set empirically.

Now we consider the spatial coherence term in Equation 4.4. Since there are only two states in the label space L , the Potts model [120] can be used, i.e.,

$$V_{p,q}(f_p, f_q) = \begin{cases} \beta_{p,q}, & f_p \neq f_q \\ 0, & f_p = f_q \end{cases}. \quad (4.8)$$

In the 8-neighborhood system, we set $\beta_{p,q} = \beta_0$ if p and q are horizontal or vertical neighbors, and $\beta_{p,q} = \frac{\beta_0}{\sqrt{2}}$ if they are diagonal neighbors.

4.3.3.2 Virtual Interface

The success of the segmentation technique depends on a correct definition of the virtual interface that partitions the disparity space into the foreground and background subspaces. Here we describe how to determine the virtual interface for the developed stereo vision system based on the effective imaging volume.

The effective imaging volume of the stereo vision system is defined as the volume in which the body can be fully captured by the stereo heads. It is located in between the two imaging stands as shown in Figure 3.7. According to the optical geometry of the system and the body sizes of the majority of population, the dimensions of the effective imaging volume is set as 1200 mm \times 2000 mm \times 800 mm (width \times height \times depth), as illustrated in Figure 4.2. The origin of the world coordinate system, O_w , is at the center of the floor plane of the volume, and the positive Z_w -axis points to the frontal stereo heads. The space within the volume should be clear except the subject during imaging, and any external object should be ignored by the matching algorithm. Thus, we can use the virtual walls of the volume to divide the 3D space into the foreground and background. In practice, the two side walls are not required because objects beyond them

are invisible to the cameras. The necessary floor, roof, front and rear walls are indexed from 0 to 3 in Figure 4.2. For each stereo head, three of them are applied to segment the foreground from the background. For example, the floor, roof and rear walls are used for the frontal stereo heads.

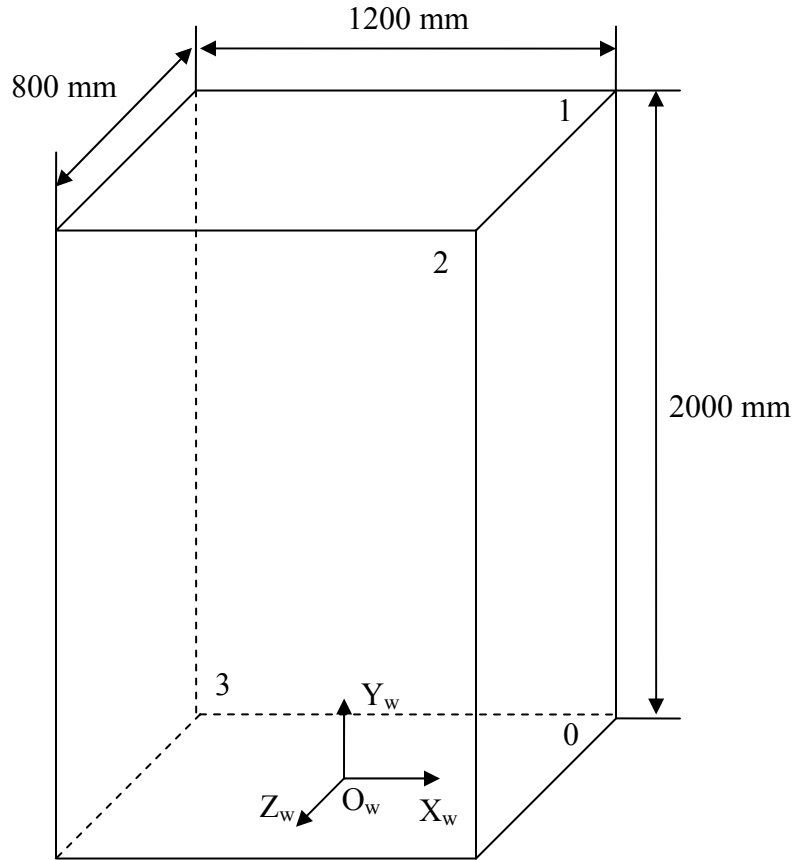


Figure 4.2: The effective imaging volume of the proposed stereo vision system.

Nevertheless, we need to convert the interface in the 3D space to the virtual interface in the disparity space. The problem is essentially how to compute the disparity

map of a 3D plane. We will show that a 3D plane actually induces a homography [134, 135] between the two image planes in stereo vision.

In Figure 4.3, two camera coordinate systems with the parallel-axis stereo geometry are defined. The 3D plane Π is defined in the left camera coordinate system with the normal \mathbf{n} and the perpendicular distance from the origin s . Let \mathbf{X}_l and \mathbf{X}_r be the left and right camera coordinates respectively of an arbitrary point P in Π . We assume the transformation between the two camera coordinate systems are known,

$$\mathbf{X}_r = \mathbf{R} \mathbf{X}_l + \mathbf{t}. \quad (4.9)$$

Since $\mathbf{n}^T \mathbf{X}_l = s$, i.e., $\frac{1}{s} \mathbf{n}^T \mathbf{X}_l = 1$, it yields

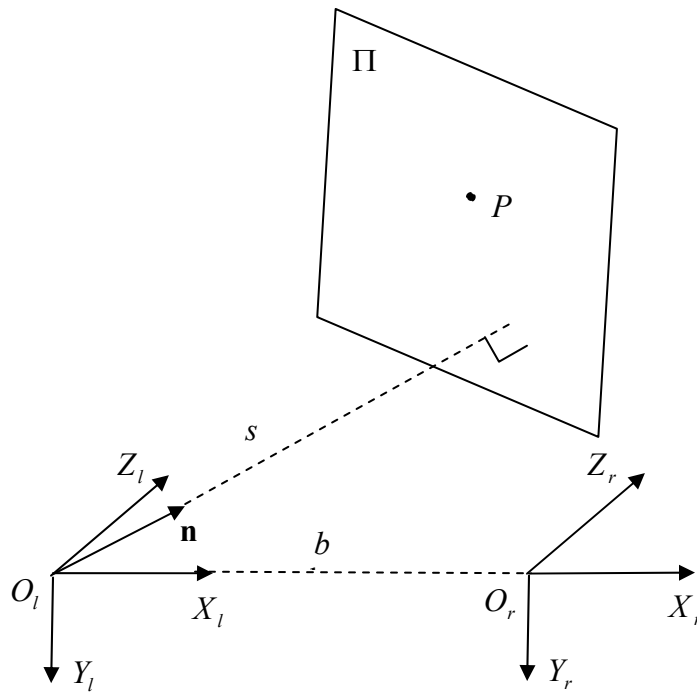


Figure 4.3: A 3D plane induces a homography between the image planes in stereo vision.

$$\mathbf{X}_r = \mathbf{R} \mathbf{X}_l + \mathbf{t} \frac{1}{s} \mathbf{n}^\top \mathbf{X}_l = \left(\mathbf{R} + \frac{1}{s} \mathbf{t} \mathbf{n}^\top \right) \mathbf{X}_l = \mathbf{H} \mathbf{X}_l, \quad (4.10)$$

with

$$\mathbf{H} = \mathbf{R} + \frac{1}{s} \mathbf{t} \mathbf{n}^\top, \quad (4.11)$$

which is the homograph matrix associated with Π . Specifically, for the parallel-axis stereo geometry, $\mathbf{R} = \mathbf{I}_3$, $\mathbf{t} = \begin{bmatrix} -b \\ 0 \\ 0 \end{bmatrix}$, and thus we have

$$\mathbf{H} = \begin{bmatrix} 1 - \frac{b}{d} n_x & -\frac{b}{d} n_y & -\frac{b}{d} n_z \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4.12)$$

Denote $\tilde{\mathbf{x}}_l = \begin{bmatrix} x_l \\ y_l \\ f \end{bmatrix}$ and $\tilde{\mathbf{x}}_r = \begin{bmatrix} x_r \\ y_r \\ f \end{bmatrix}$, which are the homogeneous coordinates of the

projections of the point P in the left and right image planes, respectively. Then according to the perspective projection, we obtain $\lambda_l \tilde{\mathbf{x}}_l = \mathbf{X}_l$, and $\lambda_r \tilde{\mathbf{x}}_r = \mathbf{X}_r$, where λ_l and λ_r are scalar values. In addition, we have $\lambda_l = \lambda_r$ for the parallel-axis stereo geometry. Then by replacing \mathbf{X}_l and \mathbf{X}_r in Equation 4.10, we obtain

$$\tilde{\mathbf{x}}_r = \mathbf{H} \tilde{\mathbf{x}}_l. \quad (4.13)$$

By combining Equations (4.12) and (4.13), we get

$$x_r = x_l - \frac{b}{s} \begin{bmatrix} n_x & n_y & n_z \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ f \end{bmatrix} = x_l - \frac{b}{s} \mathbf{n}^\top \tilde{\mathbf{x}}_l. \quad (4.14)$$

As a result, we can compute the disparity by

$$d = x_r - x_l = -\frac{b}{s} \mathbf{n}^\top \tilde{\mathbf{x}}_l. \quad (4.15)$$

Table 4.1: Planes of the effective imaging volume defined in the world coordinate system.

	$\hat{\mathbf{n}}$	\hat{s} (mm)
Plane 0 (floor)	$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$	5
Plane 1 (roof)	$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$	2000
Plane 2 (front)	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	400
Plane 3 (rear)	$\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$	400

In practice, it is more convenient to define the plane Π in the global world coordinate system, so we need to transform it to each individual camera coordinate system. We assume the plane equation in the world coordinate system is

$$\hat{\mathbf{n}}^T \mathbf{X}_w = \hat{s}, \quad (4.16)$$

and the transformation between the camera and world coordinate systems are

$$\mathbf{X}_w = \hat{\mathbf{R}} \mathbf{X}_c + \hat{\mathbf{t}}, \quad (4.17)$$

where we assume the camera coordinate system is defined on the left camera, i.e., $\mathbf{X}_c = \mathbf{X}_l$. Then by inserting Equation (4.17) to Equation (4.16), we obtain

$$(\hat{\mathbf{n}}^T \hat{\mathbf{R}}) \mathbf{X}_c = \hat{s} - \hat{\mathbf{n}}^T \hat{\mathbf{t}}. \quad (4.18)$$

By comparing to $\mathbf{n}^T \mathbf{X}_c = s$, we obtain the plane parameters in the camera coordinate system,

$$\mathbf{n} = \hat{\mathbf{n}}^T \hat{\mathbf{R}}, \quad (4.19)$$

and

$$s = \hat{s} - \hat{\mathbf{n}}^T \hat{\mathbf{t}}. \quad (4.20)$$

The planes of the effective imaging volume in Figure 4.2 are defined in Table 4.1. The floor plane has been slightly offset by 5 mm so as to separate the body from the floor.

4.3.3.3 Energy Minimization

Belief propagation [123, 124] and graph-cuts [120-122] are among the state-of-the-art methods to solve labeling problems in computer vision. However, belief propagation can only provide approximate solution when there are loops in the graph (such as a 2D image), even if the label space is binary [125]. In contrast, exact minimum of the energy can be obtained by graph-cuts for a binary segmentation problem [136]. Thus, we use graph-cuts to perform the energy minimization of Equation 4.4.

Let $G = \langle V, E \rangle$ be a weighted graph. The set V contains the nodes that correspond to the pixel set P and two additional nodes called terminals (the source s and the sink t). The nodes are connected by the edges in the set E .

In construction of the graph for our application, we let s represent the foreground (F), and t be the background (B). As shown in Figure 4.4, for each node that is associated to a pixel, say p , we connect it to s and t , and denote the edges as e_p^s and e_p^t , respectively. For each pair of neighboring pixels, say $(p, q) \in N$, we connect the corresponding nodes and denote the edge as $e_{p,q}$. The edges are assigned weights (costs) as follows: $c(e_p^s) = D_p(F)$, $c(e_p^t) = D_p(B)$, and $c(e_{p,q}) = \beta_{p,q}$. A cut $S|T$ is defined as a partition of the nodes in V into two disjoint sets S and T , subject to $s \in S$ and $t \in T$. The cost of $S|T$ is the sum of costs of all edges that go from S to T ,

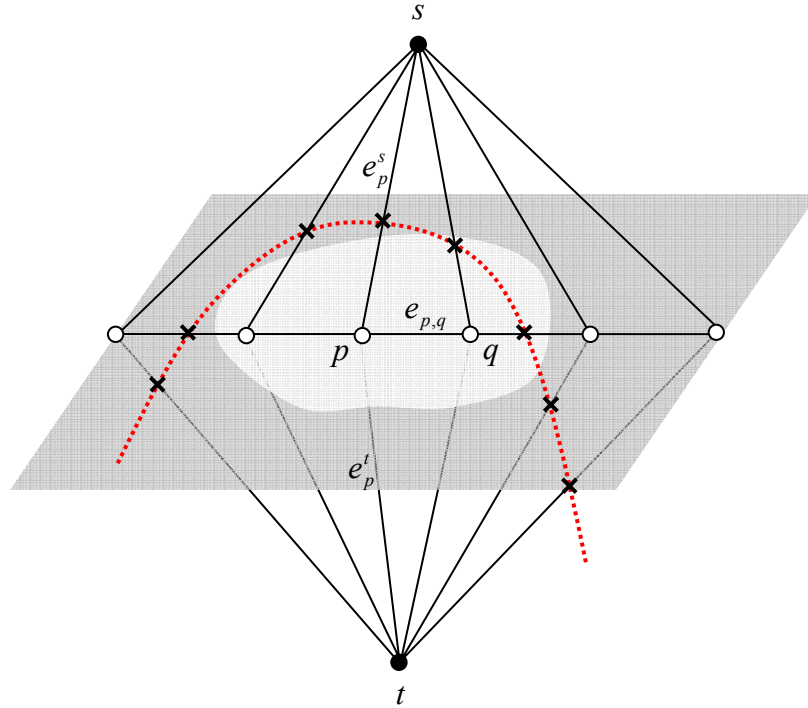


Figure 4.4: Graph construction.

$$c(S|T) = \sum_{p \in T} c(e_p^s) + \sum_{p \in S} c(e_p^t) + \sum_{p \in S, q \in T} c(e_{p,q}). \quad (4.21)$$

It is easy to see that the sum of the first two terms in $c(S|T)$ corresponds to the first term of the energy function in Equation 4.4, and the third term in $c(S|T)$ corresponds to the second term of the energy function. Therefore, the cut $S|T$ is equivalent to a labeling f , and $c(S|T) = E(f)$. As a result, to minimize the energy function is equivalent to searching for a cut with the minimum cost. According to the theorem of Ford and Fulkerson [137], the minimum cut problem can be solved by computing the maximum flow from the source to the sink. Some implementations of the maximum flow algorithms with polynomial complexities are available [120, 121].

Once the foreground is segmented, its pixels are assigned a disparity based on Equation 4.7. However, the obtained disparity map can be noisy. A median filter [138] is used to quench the impulse noise. Furthermore, morphological close and open operators [139] are used to smooth the contour.

4.3.4 Disparity Refinement

So far, the disparity map takes discrete values, which is not sufficient to recover geometric details. A disparity refinement process is needed to achieve sub-pixel accuracy. One of the standard methods is fitting a curve (e.g., parabolic [119] or Gaussian curve [140]) to the matching costs defined at discrete values. However, the curve fitting technique suffers from systematic error called “pixel-locking” effect in which disparity values are pulled towards integers [140]. Some research efforts have been made to address this problem. For example, Shimizu and Okutomi [141] attempted to reduce the bias by performing additional curve fitting on matching costs defined at half-pixel locations. Nehab et al. [142] suggested symmetric refinement by fitting a parametric surface over a 2D neighborhood of the matching cost function. Stein et al. [143] proposed an iterative refinement method that is essentially based on Lucas-Kanade algorithm [144].

It should be noted that the aforementioned improvements are all focused on reducing the “pixel-locking” effect and make disparity refinement on each individual pixel independently. However, in practice, like all other local methods, the refined disparity map is prone to be noisy. Thus, it is necessary to take into account spatial coherence during disparity updating.

Here we have developed a method that iteratively performs disparity refinement at a global level within a regularization framework [111, 116]. There are two steps in each iteration: local estimation and global optimization. For the first step, the amount of

update is estimated locally for each pixel. The estimation can be made by minimizing the matching cost function defined in Equation 4.3,

$$\delta d = \arg \min_{\delta d} C(x, y, d + \delta d) = \arg \max_{\delta d} \rho(x, y, d + \delta d), \quad (4.22)$$

where d is the current disparity value, and δd is the amount to be updated. However, the process is difficult since the correlation function ρ is highly nonlinear. Although it is possible to perform linearization of ρ with first-order approximation, the computation is still extensive. So instead, we will apply the sum of squared differences (SSD) as the matching cost as in Lucas-Kanade algorithm [144]. If the SSD takes into account the gain and bias factors between cameras, it is essentially equivalent to normalized cross-correlation. Now the matching cost is defined as

$$C_{SSD}(x, y, d) = \sum_{(u,v) \in W(x,y)} (I_r(u + d, v) - (aI_l(u, v) + b))^2, \quad (4.23)$$

where a and b are the gain and bias factors, respectively. Here we assume the disparity is constant within the matching window W . But this assumption is generally not true except for frontal-parallel surfaces. To allow the disparity to vary within the window, we first warp the right image based on the current disparity map,

$$\hat{I}_r(x, y) = I_r(x + d(x, y), y). \quad (4.24)$$

To estimate δd , a and b , we define an error function with \hat{I}_r based on the SSD,

$$e^2(\delta d, a, b; x, y) = \sum_{(u,v) \in W(x,y)} (\hat{I}_r(u + \delta d, v) - (aI_l(u, v) + b))^2. \quad (4.25)$$

With a first-order approximation, we get

$$e^2(\delta d, a, b; x, y) = \sum_{(u,v) \in W(x,y)} (\hat{I}_r(u, v) + \hat{I}_{rx}(u, v)\delta d - (aI_l(u, v) + b))^2, \quad (4.26)$$

where $\hat{I}_{rx} = \frac{\partial \hat{I}_r}{\partial x}$ is the intensity gradient of the warped right image.

Let $\mathbf{p} = [\delta d \quad a \quad b]^T$, $\mathbf{a} = [I_{rx} \quad -I_l \quad -1]^T$, then a concise form of Equation 4.26

is

$$e^2(\mathbf{p}) = \sum (\mathbf{a}^T \mathbf{p} + I_r)^2. \quad (4.27)$$

This is a classic least squares problem. To minimize $e^2(\mathbf{p})$ is equivalent to solve the normal equations,

$$\mathbf{A}\mathbf{p} = \mathbf{b}, \quad (4.28)$$

where $\mathbf{A} = \sum \mathbf{a}^T \mathbf{a}$, and $\mathbf{b} = -\sum I_r \mathbf{a}$.

We have described how to estimate δd at each pixel. Now we show how to update the disparity map at a global level. First, an energy function is defined by

$$E(d) = \iint (d(x, y) - \tilde{d}(x, y))^2 dx dy + \lambda \iint (d_x^2 + d_y^2) dx dy, \quad (4.29)$$

where \tilde{d} is the local estimate of the disparity, and d_x , d_y are the disparity gradients.

The first term in the equation measures the consistency with the local estimation, and the second term imposes smoothness constraints on the solution. λ is called the regularization parameter that weighs the smoothness term.

For the n -th iteration, we set $\tilde{d}^n = d^{n-1} + \delta d^n$. Then the discrete form of $E(d)$ can be expressed as

$$E(d) = \sum_{(i,j) \in I} \left(d^n(i, j) - (d^{n-1}(i, j) + \delta d^n(i, j)) \right)^2 + \lambda \left(d^n(i+1, j) - d^n(i, j) \right)^2 + \left(d^n(i, j+1) - d^n(i, j) \right)^2, \quad (4.30)$$

where (i, j) is the discrete coordinates of a pixel in the image plane I , and the discrete gradients are computed using the forward difference. Minimizing the energy function yields

$$(1 + k_p \lambda) d_p^n - \lambda \sum_{q \in N(p)} d_q^n = d_p^{n-1} + \delta d_p^n \quad (4.31)$$

for each pixel p whose number of neighboring pixels is $k_p = |N(p)|$. Then we can establish a linear system

$$\mathbf{P} \mathbf{d} = \mathbf{h}, \quad (4.32)$$

where $[\mathbf{P}]_{p,p} = 1 + k_p \lambda$, $[\mathbf{P}]_{p,q} = \begin{cases} -\lambda, & q \in N(p) \\ 0, & \text{otherwise} \end{cases}$, $[\mathbf{d}]_p = d_p^n$, and $[\mathbf{h}]_p = d_p^{n-1} + \delta d_p^n$.

Since \mathbf{P} is a sparse, positive, symmetric matrix, the solution can be searched efficiently using the conjugate gradient method [145].

4.5 RESULTS

The stereo matching algorithm has been tested on images of static objects and human subjects captured by our system. The cameras were carefully calibrated using the techniques described in the last chapter, and the images were rectified prior to performing matching. The same set of parameters was used throughout the test, as listed in Table 4.2. The virtual interface in the disparity space was created for each stereo head according to its calibration parameters, and the results are shown in Figure 4.5. The disparities are coded with the standard cold-to-hot color mapping that corresponds to “far-to-close” to the cameras. The results on a mannequin and a human subject are shown in Figures 4.6–4.9. For each figure, the image pair is shown in (a); the coarse and refined disparity maps are shown in (b) and (c), respectively; and to better evaluate the performance of object segmentation, the refine disparity map has been overlaid onto the reference (left) image as shown in (d). The results show that the algorithm is effective in both foreground segmentation and sub-pixel matching, and is promising for our application.

The test was carried out on a personal computer with an AMD Athlon™ 2.0 GHz dual-core CPU and 1.0 G RAM. Because, in principle, the computation for the four stereo heads can be undertaken in parallel, we can take advantage of the multithreading function of a multi-core CPU to improve the time efficiency. In our system, two pairs of images can be matched simultaneously, and it takes about 80 s to complete the computation for a full set (four pairs) of images.

Table 4.2: Parameters for the test of stereo matching.

Parameter	Description
$W_{NCC} = 5 \times 5$	Window size of NCC
$\sigma_t = 1.0$	Threshold of the variation of intensity for detecting unmatched pixels
$\rho_t = 0.6$	Threshold of NCC for detecting unmatched pixels
$C_O^F = 1.0$	Cost of assigning an unmatched pixel to the foreground
$C_O^B = 0.2$	Cost of assigning an unmatched pixel to the background
$\beta_0 = 1.0$	Parameter in the Potts model
$W_{Median} = 21 \times 21$	Window size of the median filter for reducing noise in the coarse disparity map
$r_{STE} = 3.3$	Radius of the circular structural element in the morphological close operator for smoothing the contours of foreground objects
$N_{Iter} = 15$	Number of iterations in disparity refinement
$W_{SSD} = 11 \times 11$	Window size of SSD in disparity refinement
$\lambda = 10.0$	Regularization parameter in disparity refinement

4.6 DISCUSSION

For our application, the projection of artificial texture makes it difficult to use color or contrast for segmentation. Fortunately, we can make use of the stereo cues and prior knowledge on the structure of the scene by defining a virtual interface between the foreground and background. Based on the method, an initial disparity map can be obtained with accurate foreground segmentation. Then disparity refinement is performed

on the foreground pixels. The latter process involves local least-squares matching and global optimization. High accuracy can be reached by the least-squares matching, but stability is maintained and potential errors are corrected by the global optimization in a regularization framework.

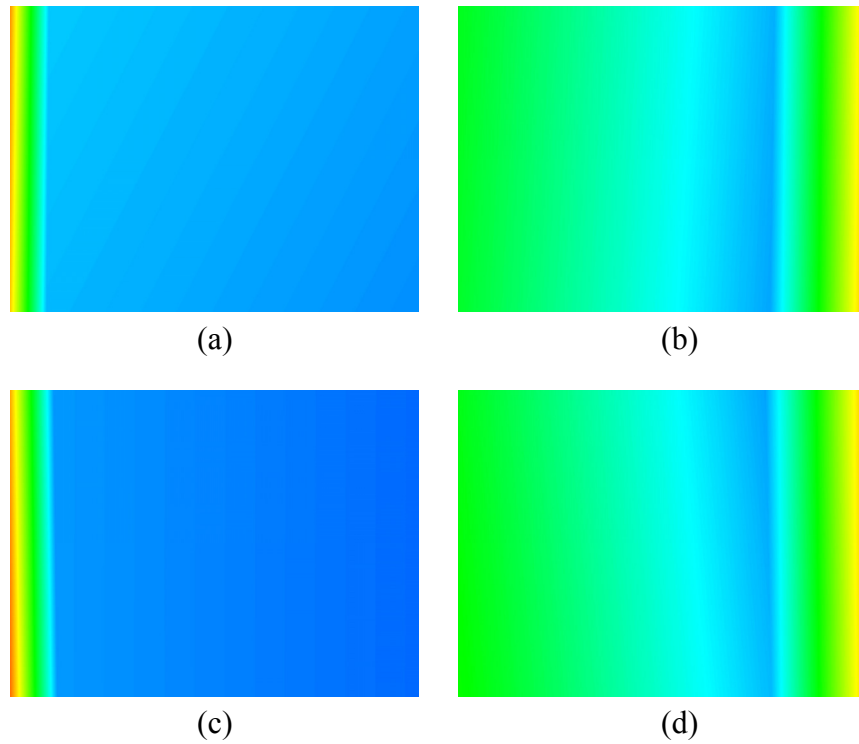


Figure 4.5: Virtual interface defined in the disparity space for the (a) front-top, (b) front-bottom, (c) back-top, and (d) back-bottom stereo heads.

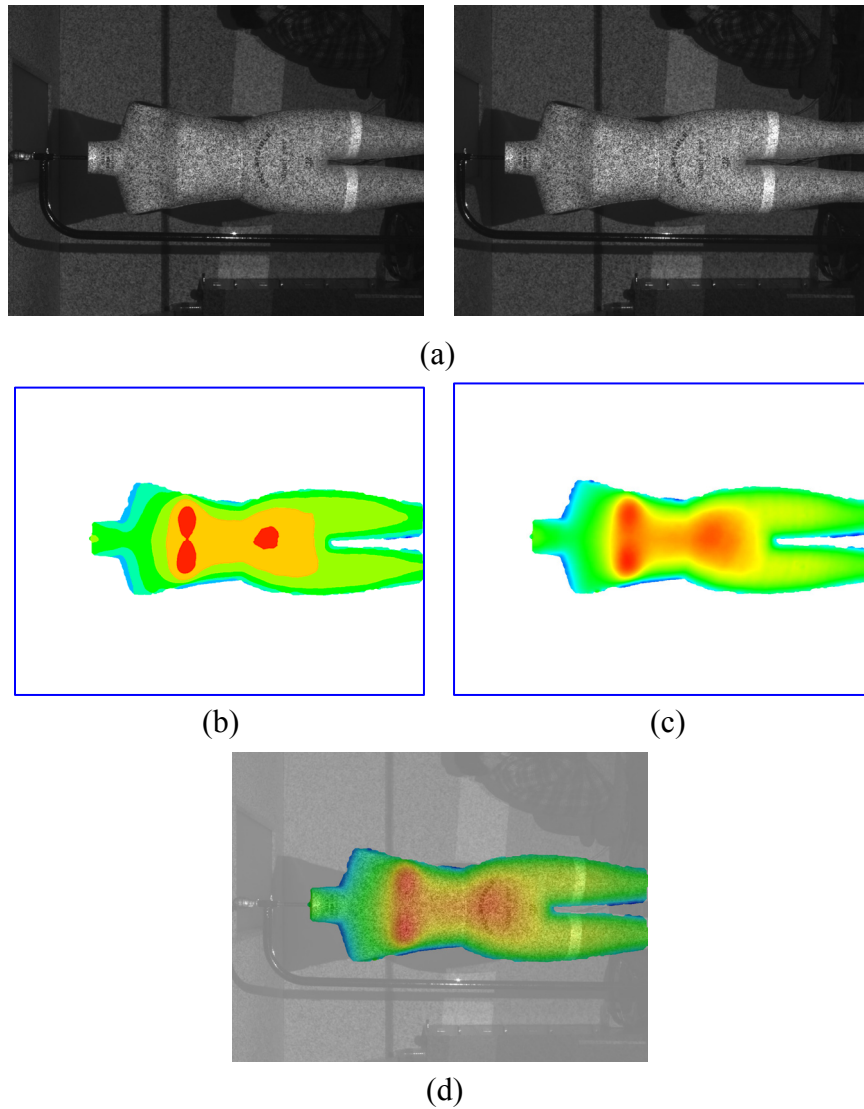


Figure 4.6: Results on a mannequin for the front-top stereo head. (a) Rectified image pair; (b) foreground segmentation and coarse disparity map; (c) refined disparity map; and (d) the refined disparity map is overlaid onto the reference image.

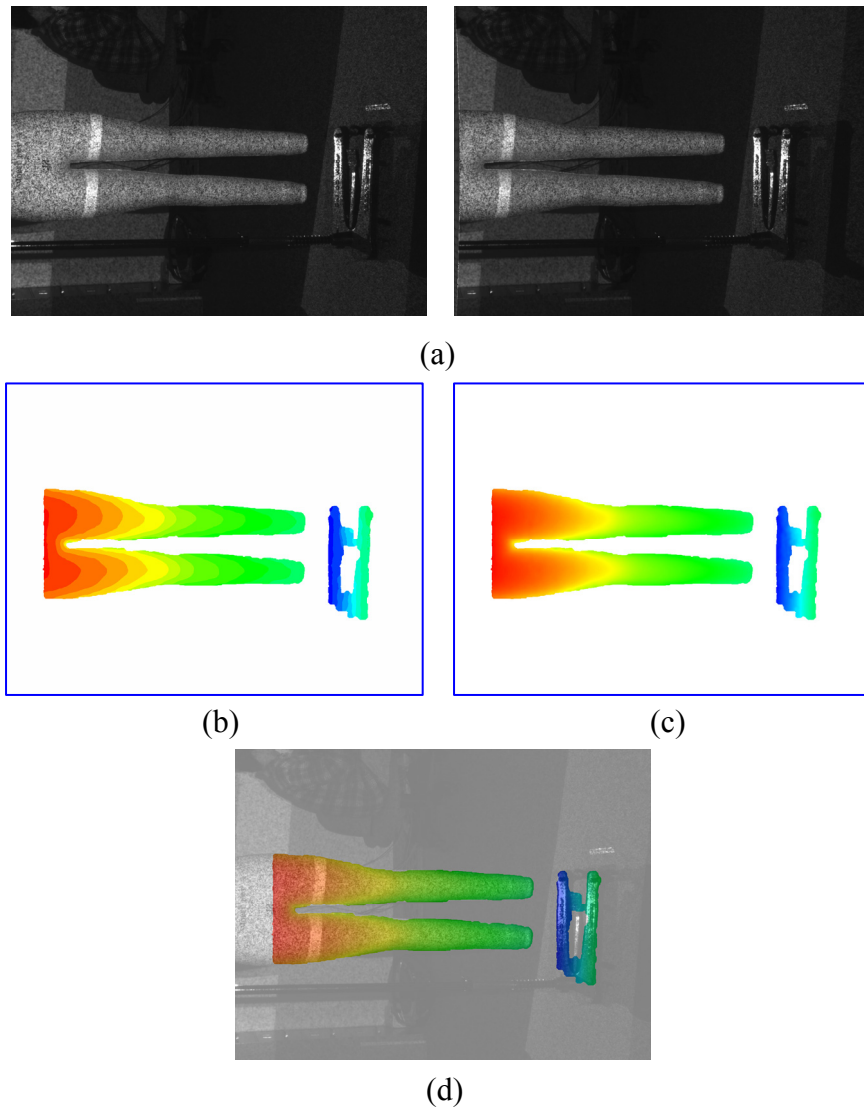


Figure 4.7: Results on a mannequin for the front-bottom stereo head. (a) Rectified image pair; (b) foreground segmentation and coarse disparity map; (c) refined disparity map; and (d) the refined disparity map is overlaid onto the reference image.

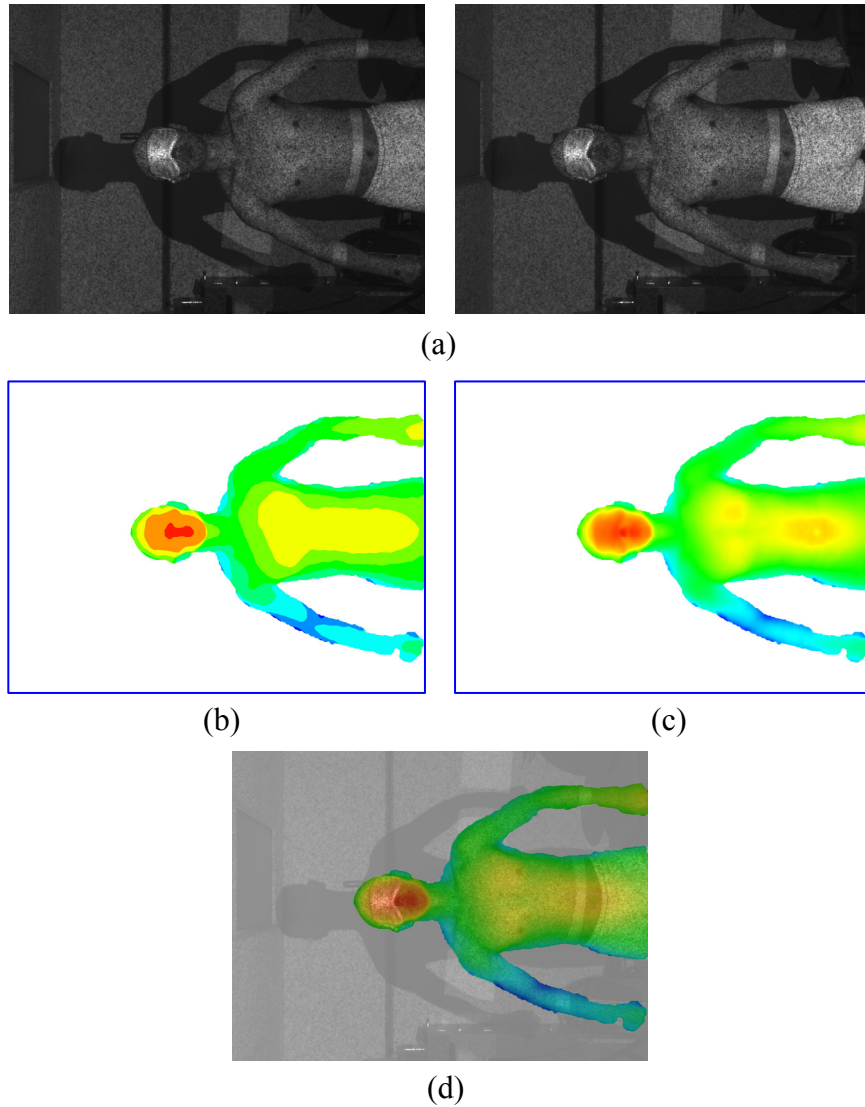


Figure 4.8: Results on a human subject for the front-top stereo head. (a) Rectified image pair; (b) foreground segmentation and coarse disparity map; (c) refined disparity map; and (d) the refined disparity map is overlaid onto the reference image.

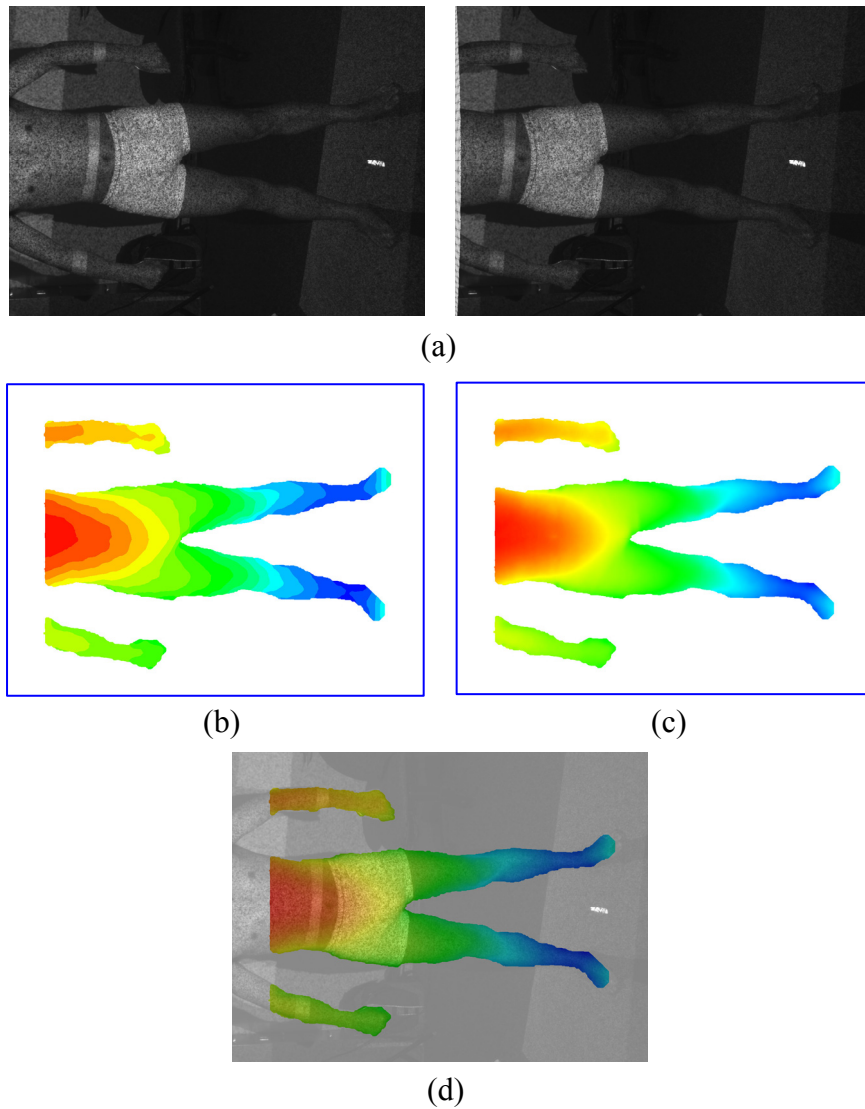


Figure 4.9: Results on a human subject for the front-bottom stereo head. (a) Rectified image pair; (b) foreground segmentation and coarse disparity map; (c) refined disparity map; and (d) the refined disparity map is overlaid onto the reference image.

Chapter 5

Surface Reconstruction

5.1 INTRODUCTION

The raw data acquired with a typical body scanner are usually comprised of hundred thousands of scattered 3D points, from which it is hard to read and handle the desired information directly. Body modeling, the focus of this chapter, is a process that accurately fits the scan data with a more manageable representation so that the data can be manipulated and interpreted more easily for some specific applications. Since, in general, such a representation is in the form of 3D surface, the process is also called body surface reconstruction.

In developing a surface reconstruction algorithm, first we need to choose a proper surface representation. One of the most common representations is the B-spline surface representation due to its attractive properties such as piecewise smoothness, local support, and the same differentiability as with the basis functions [146]. But B-spline patches require cylindrical or quadrilateral topology, and intricate boundary conditions are needed to zipper patches together to represent a more complex surface. In contrast, a piecewise smooth subdivision surface resulted from iteratively refining a control mesh of arbitrary topology gives a more flexible representation [147]. In this chapter, we present an effective body surface reconstruction algorithm, which is based on subdivision surface representation, for the developed stereo vision system.

5.2 RELATED WORK

With the advances of body scanning technology, a number of research efforts have also been made on body surface reconstruction. Since the human body has a specific

branching structure, the topological information of the overall shape can be used to facilitate the modeling process. For example, the algorithm developed by Douros et al. [147] segments the body into several parts so that each of which has a cylindrical topology and can be modeled through B-spline surface interpolation, but it needs to devise boundary conditions at branches to join the surfaces smoothly. The algorithm was later improved by using a number of small B-spline patches to perform local interpolation on the data, so as to get better segment joining [148]. In our previous study [149], we first resampled the data on a regular grid and then applied B-spline surface approximation instead of interpolation to fit each body segment, and made the density of control mesh adaptable to a given error bound. Allen et al. [150] described a template-based scheme which creates morphable body models by fitting template meshes to body scans with sparse 3D marks. It is noticeable that the main drawback of the shape-specific methods is that body segmentation or mark detection is required.

Some general-purpose surface reconstruction algorithms may also be considered for body modeling. One of the most remarkable algorithms was proposed by Hoppe et al. [151, 152]. The algorithm determines the topology of the surface and produces an initial estimate of the geometry by estimating the signed distance function of the surface, and then fits a piecewise smooth subdivision surface to the data. Although the method can deal with arbitrary topologies, it requires a uniform sampling and is not robust to noise. Since the scan data are usually subject to gaps and holes due to occlusions, and are not noise-free, Hoppe's algorithm is not applicable to this study. However, we have borrowed its idea of subdivision surface fitting when developing our new algorithm. One of the nice characteristics of subdivision surface modeling is that it makes possible the representation of the whole human body with a single surface, so body segmentation is no longer necessary.

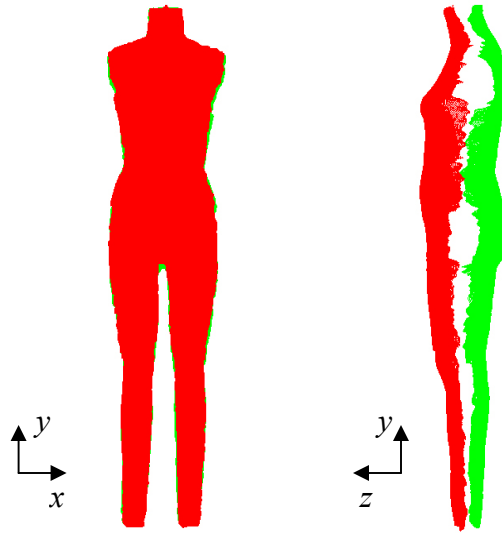


Figure 5.1: Original scan data of a mannequin in the anterior (a) and lateral (b) views. Data points from the front (back) stereo heads are in red (green).

5.3 METHODOLOGY

As described in Chapter 3, the developed body scanner is made up of four stereo heads mounted on two stands that are placed in front and back of the subject, and thus the scan data can be grouped into two sets that correspond to the front and back views, respectively. The advantage of this kind of construction lies in its portability and low cost. However, it leaves large gaps between the two views due to occlusions. As an example, the raw scan data of a mannequin is shown in Figure 5.1. The mannequin data will be used throughout the description of the algorithm. The scan data comprises of around 910,000 scattered 3D points. The data are noisy, incomplete, non-uniformly distributed, and have outliers. The objective of surface reconstruction is to create an accurate, smooth, complete and compact 3D surface model which will be used in applications such as 3D body measurement. A desirable reconstruction technique should

not only be able to produce a surface that is a good approximation to the original data, but also be capable of filling the holes and gaps and smoothing out noise. We will show that the goals can be reached by the proposed subdivision surface reconstruction algorithm. The basic idea of the method is described here. First, we resample the original 3D data points on a regular grid and use the explicit neighborhood information of the resampled data to create an initial dense mesh. Secondly, the initial dense mesh is simplified to produce an estimate of the control mesh. Finally, the control mesh is optimized by fitting its subdivision surface to the original data, and accordingly, the body model is reconstructed.

In summary, the algorithm consists of four steps: (1) data resampling; (2) initial mesh generation; (3) mesh simplification; and (4) mesh subdivision and optimization. Details are described in the following subsections.

5.3.1 Data Resampling

The first challenge of this algorithm is to generate an initial mesh from the raw data, but this is not a trivial process, especially when the data are noisy, incomplete and irregularly distributed. To facilitate the generation of an initial mesh, the data are resampled on a regular grid and thus explicit neighborhood information can be easily extracted.

The data set of both the front and back views can be regarded as samples on a bivariate surface defined in the frontal projection plane, i.e., the x - y plane in the (x, y, z) Cartesian coordinates as shown in Figure 5.1. The projections of the data in the x - y plane have irregular distributions across the body. Data resampling is a procedure which reorganizes the data on a regular grid defined in the x - y plane. First, we need to select the grid density, i.e., the sampling intervals in the x and y directions. A good sampling interval should be considerably larger than the system resolution but not too large to

separate different body segments. We select the sampling intervals as $\Delta x = \Delta y = 3 \text{ mm}$, considering the system resolution is about 1 mm in the x and y directions. For each grid, we collect all points falling in it, and then take the mean or median z value of the points as the z value of the grid point. This method is computationally effective. Some geometric details may be lost after resampling, but fortunately, geometric accuracy is not important in initial mesh generation.

Another important issue raised up to this point is how to merge data from different stereo heads in each view. Since in either the front or the back view, the data are from two stacked stereo heads, a deviation from a perfect alignment may occur due to imperfect system calibration. In each view, the upper and lower point clouds are partially overlapped. If direct averaging is applied in this region, the resampled data will appear rough due to the misalignment. To tackle the problem, data are weighted prior to resampling to reach a seamless blend. As illustrated in Figure 5.2, let y_a and y_b be the minimum y value of the upper data set and the maximum y value of the lower data set, respectively, and the overlap means we have $y_b > y_a$. By denoting $t = \frac{y - y_a}{y_b - y_a}$, the

weighting function for the upper data set is defined by

$$w_U(y) = \begin{cases} 3t^2 - 2t^3, & 0 \leq t \leq 1 \\ 1, & t > 1 \end{cases}, \quad (5.1)$$

where $t \geq 0$. Similarly, the weighting function for the lower data set is

$$w_L(y) = \begin{cases} 1, & t < 0 \\ 1 - (3t^2 - 2t^3), & 0 \leq t \leq 1 \end{cases}, \quad (5.2)$$

where $t \leq 1$. Because the S-shaped function $3t^2 - 2t^3$ is of the first-order continuity, smooth merge of the data sets can be obtained by the proposed weighted averaging. Additionally, by using these weighting functions, the weight of a point decreases with the distance from the center of the data set, which is favorable since the distortion caused by

imperfect camera calibration is usually more obvious in the peripheral regions than at the center of an image and thus we should put less confidence in peripheral data.

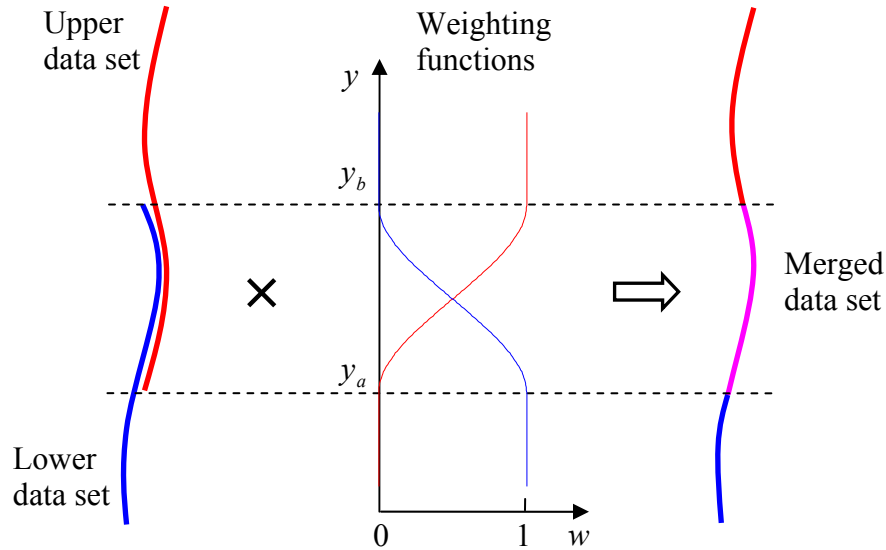


Figure 5.2: Smooth merge of the upper and lower data sets in each view.

5.3.2 Initial Mesh Generation

Prior to mesh generation, we need preprocess the resampled data to fill holes and remove outliers. First, we define a mask for each view. The mask is actually a binary version of the range image; 1s are assigned to the occupied grid points, and 0s to the empty points. Then the front and back masks are combined by a logical OR operation. Small blobs in the combined mask are removed as outliers, and only the largest connected region is kept as the body. If there are holes in the remaining region, they are filled by flooding. The combined mask will serve as a reference in the following triangulation process.

The neighborhood information is explicit in a range image, and a triangle mesh can be produced by triangulation between adjacent rows of data. But a problem arises

when we try to merge the front and back meshes to form a complete one, because the mesh contours are so irregular that some conventional tiling algorithms (e.g., [153]) would fail. To tackle this problem, a technique called Add-Delete is developed in this study. First, some points are padded to both range images according to the combined mask so that their contours have the same projection in the x - y plane. The z values of the added points are set by interpolating their nearest neighbors. With this procedure, holes are also filled. Then triangulation is performed separately on both modified range images. To merge the two triangle meshes, the modified contours are tiled by simply connecting corresponding points. When the meshes are merged, a vertex collapse algorithm can be used to delete these added points. In vertex collapse, a vertex is removed, and its adjacent vertices are re-triangulated to form a new mesh. However, an improved method is used in this work. We assign a small weight to the added points, and consequently, these points will tend to be deleted first in the following mesh simplification phase. This “soft deletion” strategy can avoid sharp edges that may occur in direct vertex collapse and re-triangulation.

Another issue raised up to this point is that the webbing effect can occur when different body segments are in close proximity or even in touching. It typically occurs in the crotch area and the armpits, especially for overweight subjects. To deal with this issue, we have developed an interactive edit tool with which the user can manually mark an area where the webbing affect is likely to occur. Original data points in the marked area will be temporarily discarded for initial mesh generation stage, but will be used in the mesh optimization stage for surface fitting.

An example of mesh triangulation is shown in Figure 5.3. We can see the frontal projections of the front and back meshes completely overlap after padding some points.

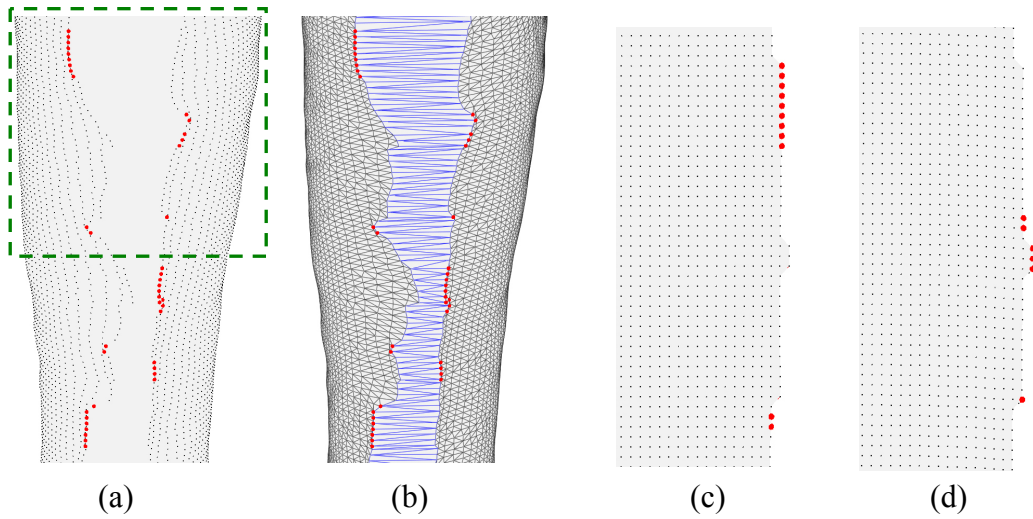


Figure 5.3: Triangulation of the leg. (a) Resampled and padded (in red) data points. (b) The front and back meshes are merged by tiling the modified contours (in blue). Frontal projections of the modified front and back contours, corresponding to the region marked with a dashed green box in (a), are shown in (c) and (d), respectively.

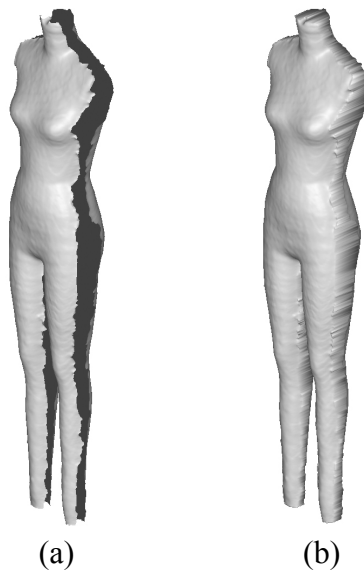


Figure 5.4: Generation of the initial mesh. Shaded models of the mesh before (a) and after (b) closing the gaps.

The generated initial mesh of the whole body is shown in Figure 5.4. Gaps are closed after merging the contours.

5.3.3 Mesh Simplification

The purpose of mesh simplification is to reduce the number of triangles and create a control mesh for the model. The simplification algorithm devised by Garland [154] is employed in this work. This algorithm can produce high-quality approximations by using quadric error metrics. It is realized by edge collapse as shown in Figure 5.5, where the new vertex \bar{v} is evaluated by minimizing the weighted sum of its squared distances (the cost of contraction) to all triangles around vertices v_i and v_j . An edge with the smallest cost of contraction will be collapsed first.

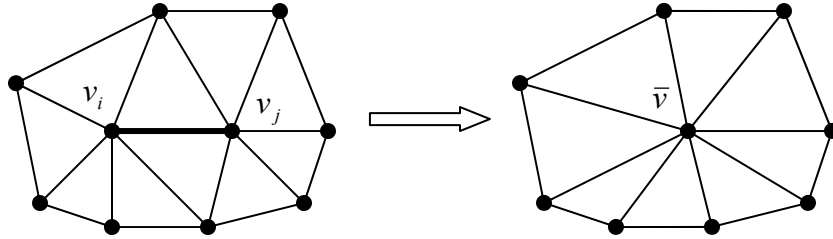


Figure 5.5: Edge collapse. A new vertex \bar{v} is created by collapsing the edge $\overline{v_i v_j}$.

The quadric measure of the distance of a point v to a plane determined by a point p and a unit normal \mathbf{n} is given by

$$D^2(\mathbf{v}) = \left((\mathbf{v} - \mathbf{p})^T \mathbf{n} \right)^2 = \mathbf{v}^T \mathbf{n} \mathbf{n}^T \mathbf{v} - 2(\mathbf{n} \mathbf{n}^T \mathbf{p})^T \mathbf{v} + \mathbf{p}^T \mathbf{n} \mathbf{n}^T \mathbf{p}. \quad (5.3)$$

Here we have adopted the convention that the 3D coordinates of a point is expressed as the same letter but in bold typeface.

If we define $\mathbf{A} = \mathbf{n} \mathbf{n}^T$, $\mathbf{b} = \mathbf{n} \mathbf{n}^T \mathbf{p}$, and $c = \mathbf{p}^T \mathbf{n} \mathbf{n}^T \mathbf{p}$, then the quadric error metric can be expressed in this form

$$Q(\mathbf{v}) = \mathbf{v}^T \mathbf{A} \mathbf{v} - 2\mathbf{b}^T \mathbf{v} + c, \quad (5.4)$$

where the quadric Q is defined as a triple

$$Q = (\mathbf{A}, \mathbf{b}, c). \quad (5.5)$$

The quadric for a given vertex v in the original mesh can be expressed as a weighted sum of the fundamental quadrics of its adjacent faces,

$$Q = \sum_k w_k Q_k, \quad (5.6)$$

where Q_k is the quadric of the k -th adjacent face, and the associated weight w_k is set as the face area.

For an edge to be collapsed, such as the edge (v_i, v_j) in Figure 5.5, the quadric is $Q = Q_i + Q_j$, and the cost of contraction is $Q(\bar{v}) = Q_i(\bar{v}) + Q_j(\bar{v})$, where \bar{v} is the new vertex after collapse. By minimizing the function, we can get the optimal position of \bar{v} ,

$$\bar{v} = \mathbf{A}^{-1} \mathbf{b}, \quad (5.7)$$

and the cost

$$Q(\bar{v}) = -\mathbf{b}^T \mathbf{A}^{-1} \mathbf{b} + c. \quad (5.8)$$

In implementation, all candidate edges are sorted in a heap based on costs. At each step, the edge with minimum cost is removed from the heap and collapsed, and then the heap is updated. This procedure is iteratively repeated until enough simplification is achieved. It is worth noting that, to realize the aforementioned “soft deletion” strategy, the cost for an artificially added vertex should be scaled down, so that its associated edges will move upwards in the heap and gain a higher priority for collapse.

An example for mesh simplification is shown in Figure 5.6, where the mesh was obtained by collapsing 96% edges of the model in Figure 5.4(b). The mesh will serve as a control mesh for the surface subdivision described below.

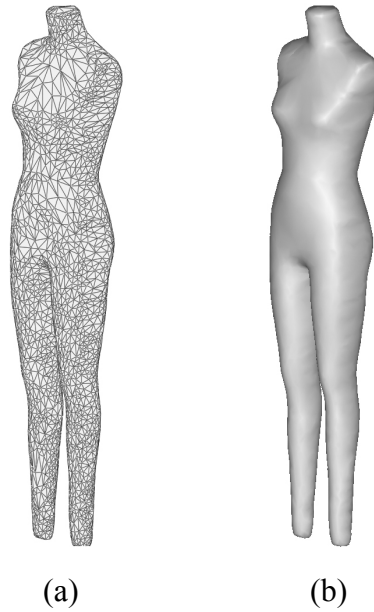


Figure 5.6: A simplified mesh (a) and its shaded model (b).

5.3.4 Mesh Subdivision and Optimization

So far, we have obtained a raw model which can be taken as the control mesh for further surface refinement. A piecewise smooth mesh can be produced by surface subdivision techniques. We use Loop's subdivision algorithm [155] by which a C^2 smooth surface can be obtained in the limit of infinite numbers of subdivisions. This algorithm is based on edge split, as shown in Figure 5.7. At each level of subdivision, each edge is split into two, and thus each face is split into four. The surface obtained by an infinite refinement process is called the limit surface of the control mesh. For our applications, a single level of subdivision is enough to obtain a sufficiently dense mesh that will be used to approximate the limit surface. The refined mesh includes two types of points. The first type is called vertex points which are the displacements of the control vertices. The other type is called edge points which are the inserted points on edges. One

of the advantages of Loop's subdivision scheme is that the limit positions of these points can be explicitly evaluated by some rules.

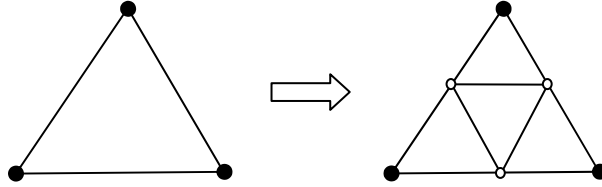


Figure 5.7: Mesh subdivision by edge split.

The evaluation rule for a vertex point p with a valence of n as shown in Figure 5.8 is given by

$$\mathbf{p} = \frac{c_0 \mathbf{v}_0 + c_1 \mathbf{v}_1 + \cdots + c_n \mathbf{v}_n}{c_0 + c_1 + \cdots + c_n}, \quad (5.9)$$

where v_0, \dots, v_n are control vertices, $c_0 = \frac{3}{8}n/a(n)$ with $a(n) = \frac{5}{8} - \frac{(3 + 2 \cos(2\pi/n))^2}{64}$, and $c_1 = \cdots = c_n = 1$.

According to the Loop's subdivision rules, the evaluation of an edge point will involve all 2-neighborhood vertices. For an edge point in an ordinary mesh with a valence of six for each vertex, the number of 2-neighborhood vertices is 14, as shown in Figure 5.9. In our work, to simplify computation, especially in mesh traversal, we only consider 1-neighborhood vertices, which are highlighted in red in Figure 5.9. Our results show that this simplification still provides a good approximation. The approximate evaluation rule for an edge point p is given by

$$\mathbf{p} = \frac{c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + c_3 \mathbf{v}_3 + c_4 \mathbf{v}_4}{c_1 + c_2 + c_3 + c_4}, \quad (5.10)$$

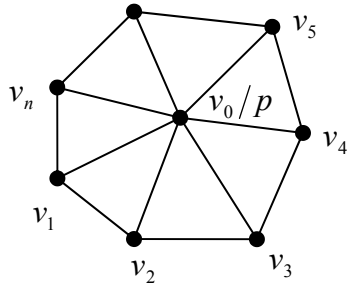


Figure 5.8: Illustration of Loop's evaluation rule for a vertex point.

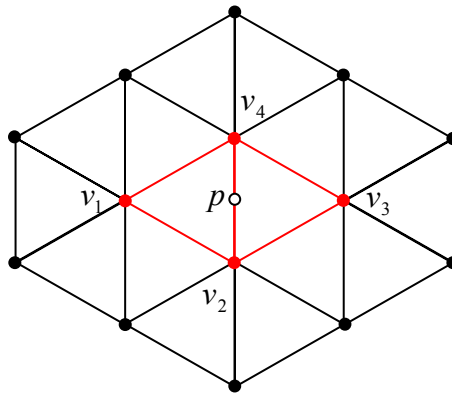


Figure 5.9: Illustration of the modified Loop's evaluation rule for an edge point.

where $c_1 = c_3 = 15$, and $c_2 = c_4 = 32$.

Loop's subdivision algorithm is an approximating scheme, which indicates the limit surface does not pass through the control vertices. In our case, it means the resulting model will pull away from the original scanner data. Therefore, the control vertices should be optimized so that the final model can more accurately approximate the original

data. This optimization process can be realized by minimizing the distance between the limit surface and the original data. The idea is similar to that in Hoppe et al.'s work [152].

The distance can be represented by an energy function as defined by

$$E = \sum_i \|w_i (\mathbf{q}_i - \mathbf{p}_i)\|^2, \quad (5.11)$$

where q_i is an original data point with w_i as its weight calculated from Equation (5.1) or (5.2), and p_i is its closest point on the limit surface. We can derive that p_i is a weighted combination of the control vertices, i.e., $\mathbf{p}_i = \sum_j l_{i,j} \mathbf{v}_j$, where the coefficients $l_{i,j}$ are determined by the projection of q_i on the limit surface and the said evaluation rules.

Thus, the energy function can be rewritten as

$$E = \sum_i \left\| w_i \left(\mathbf{q}_i - \sum_j l_{i,j} \mathbf{v}_j \right) \right\|^2. \quad (5.12)$$

It can also be expressed in matrix form with the Frobenius norm,

$$E = \|\mathbf{W}(\mathbf{Q} - \mathbf{L}\mathbf{V})\|_F^2, \quad (5.13)$$

where $\mathbf{Q} = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \dots \quad \mathbf{q}_N]^T$ is the data vector with N data points assumed, \mathbf{W} is the $N \times N$ diagonal weighting matrix with w_i as the i -th diagonal element, $\mathbf{V} = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_K]^T$ is the vertex vector with K control vertices assumed, and \mathbf{L} is the $N \times K$ evaluation matrix with $l_{i,j}$ as the (i, j) -th element. Then the minimization of E is a least squares problem in nature, and equivalent to solving the linear equations $(\mathbf{W}\mathbf{L})^T(\mathbf{W}\mathbf{L})\mathbf{V} = (\mathbf{W}\mathbf{L})^T\mathbf{W}\mathbf{Q}$. Since $(\mathbf{W}\mathbf{L})^T(\mathbf{W}\mathbf{L})$ is symmetric, positive-definite and sparse, the equations can be solved efficiently by the conjugate gradient method [145]. As a result, the optimization of \mathbf{V} can be achieved by iteratively minimizing E . It should be pointed out that least squares minimization is optimal on Gaussian noise, but is sensitive to outliers, so we should reject those data points which are far away from the limit surface. It should also be noted that, in practice, a regularization term (such as a

spring model) is usually required in the energy function to guarantee a stable solution, and we refer the interested reader to [152] for details.

As an example, an optimized subdivision mesh is shown in Figure 5.10. It was created by optimizing the control mesh in Figure 5.6 with the original data in Figure 5.1.

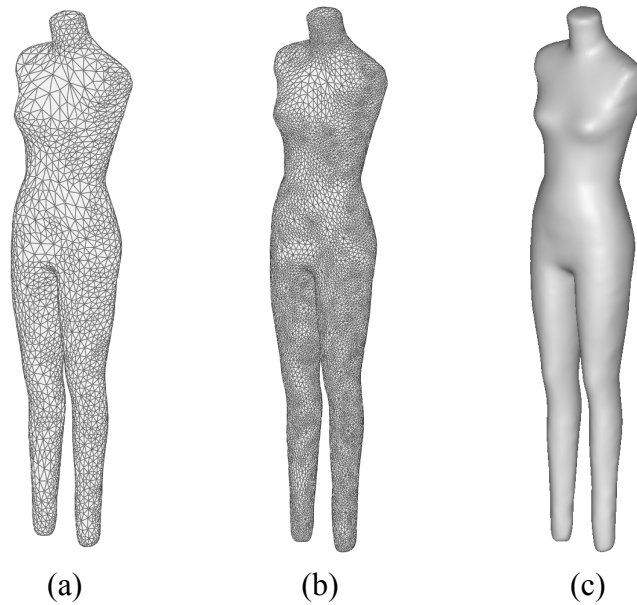


Figure 5.10: An optimized control mesh (a), and its subdivision mesh (b) and the shaded model (c).

5.4 RESULTS

We have demonstrated that a body model has been successfully reconstructed through the steps described above. To evaluate its performance, we have tested the algorithm on numerous scan data of subjects with various body shapes and sizes. In addition to the model shown in Figure 5.10, more reconstructed body models are presented in Figure 5.11. All of the models were obtained by iteratively performing mesh optimization 10 times. The data size and running time on a 2.0 GHz PC for each model

are listed in Table 5.1, where the model in Figure 5.10 is labeled as 1, and the models in Figure 5.11 are labeled as 2–6. Since we only need to store the control mesh, and the final model can be recovered by using the evaluation rules of subdivision surfaces, the data compression ratio (which is defined as the ratio of the size of original data to the size of control mesh) is up to 260. The running times for steps 1 & 2, step 3, and step 4 are shown separately. The algorithm is computationally efficient with an approximately linear complexity. The total running time is about 30 s per 1 million original data points, and the last step takes most of it. The convergence of mesh optimization is demonstrated in Figure 5.12. The modeling accuracy is measured by fitting error which equals $\sqrt{\frac{E}{\sum_i w_i^2}}$, where E is the energy defined in Equation 5.11, and $\sum_i w_i^2$ is the sum of squared weights of original data. After 10 iterations, the optimization process has already been very close to convergence, and the fitting error is less than 1.2 mm.

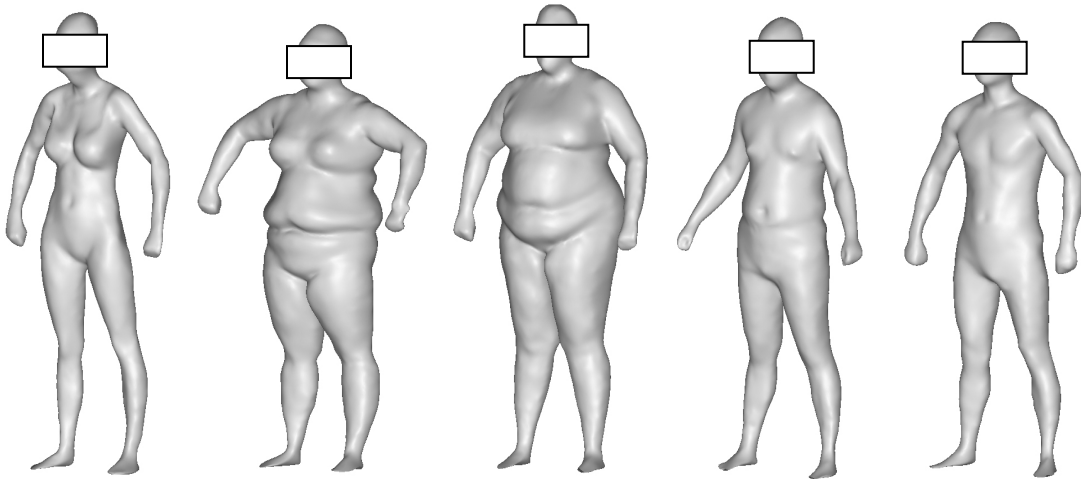


Figure 5.11: Reconstructed body models of subjects with various shapes and sizes.

Table 5.1: Data sizes and running times for the presented models.

Model	Number of points				Time (s)			
	Original	Resampled	Control	Final	1&2	3	4	Total
1	909,182	82,246	3,291	13,158	0.89	7.47	18.47	26.83
2	1,205,410	111,544	4,463	17,846	1.25	10.06	25.97	37.28
3	1,409,440	129,770	5,192	20,762	1.34	11.89	30.45	43.68
4	1,817,015	164,848	6,595	26,374	1.70	15.67	39.36	56.73
5	1,328,199	124,516	4,982	19,922	1.33	11.64	29.64	42.61
6	1,157,952	110,162	4,408	17,626	1.12	9.88	25.12	36.12

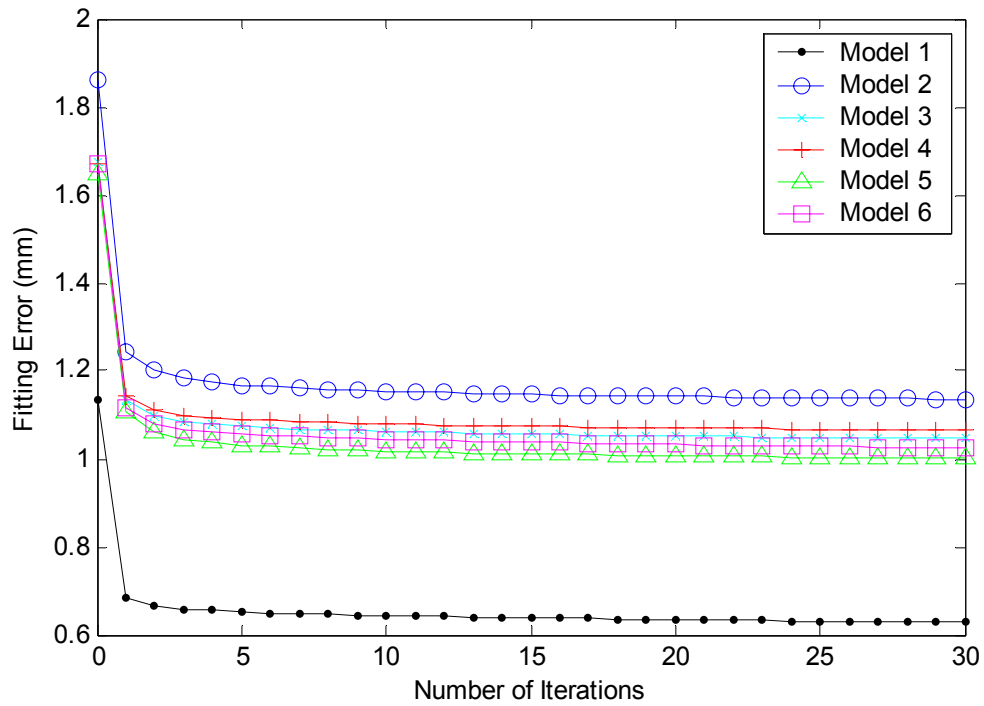


Figure 5.12: Performance of mesh optimization.

To demonstrate that the algorithm is capable of hole and gap filling, some close-up views of the model of a subject are shown in Figure 5.13. We can see the gap under the armpit has been completed and the holes at the sole have been filled. It can also be observed that the original data are noisy, but the reconstructed surface is smooth. The foot is one of the most difficult areas to be reconstructed, due to the missing data and high noise, but our result is acceptable.

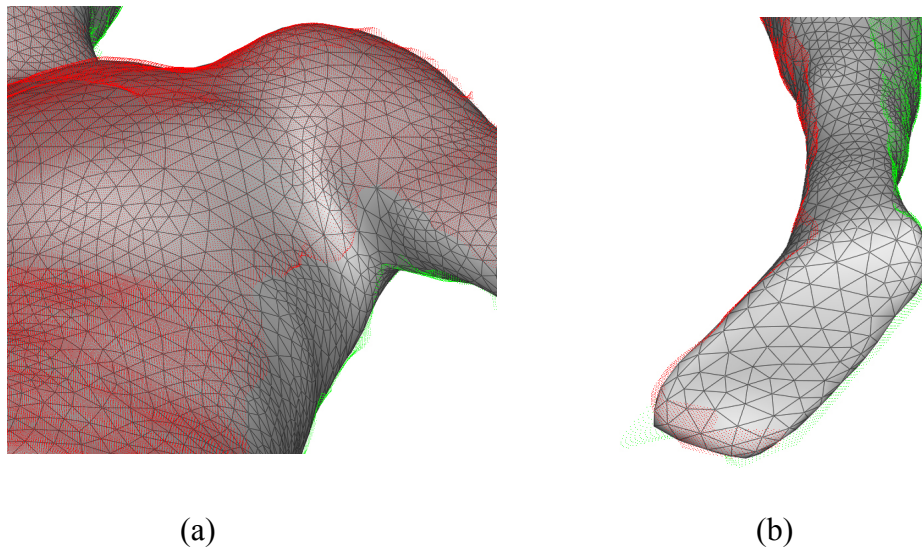


Figure 5.13: Close-up views of gap and hole filling: (a) the armpit; and (b) the foot.

5.5 DISCUSSION

We have presented a surface reconstruction algorithm for human body modeling from 3D scanner data. It represents the body with a single subdivision surface, so body segmentation can be avoided during the modeling process. In addition to the human body, it is also applicable to other objects as long as the scan data satisfy the basic assumption: the data can be separated into two views, and each view can be regarded as samples on a bivariate surface defined in a common projection plane.

For some general-purpose surface reconstruction algorithms, a plausible solution often depends on accurate estimation of the surface normals, which is not only time-consuming, but also sensitive to defects in the data. In contrast, the proposed method does not need normal estimation, so it is more robust and computationally more efficient.

For a two-view scanner, large gaps are inevitable due to the limited field of view, but our experiments have shown that the proposed method is capable of repairing the gaps automatically. For occlusion areas where the surface viewed on the sagittal plane is almost flat and featureless, the method can realistically restore smooth and natural body shapes. Nevertheless, it cannot recover detailed features in regions where the data are largely incomplete. Therefore, the two-view system is not proper for applications where high resolution of details is needed, whereas it might be highly suitable for an anthropometric survey to collect basic body dimensions. The model with a relatively small number of data points makes it suitable for effective storage and online transmission.

In spite of the promise of the proposed method, much further work should be done to improve the current system. For example, the mannequin-like appearance of the surface in some areas (such as the hand and foot) was caused by not only the defects of the original data, but also the smoothness property of subdivision surfaces, and thus the algorithm needs to be improved so as to construct more accurate and feature-preserving surfaces from sparse and noisy data.

Chapter 6

Body Measurement and System Evaluation

6.1 INTRODUCTION

This chapter serves two purposes. First, we describe how to perform body measurement on the reconstructed body model. Second, we present methods and results on the evaluation and validation of the developed 3D anthropometry system. For the latter, the system was tested on mannequins and human subjects to evaluate its accuracy and repeatability and validate its feasibility in body fat assessment.

6.2 BODY MEASUREMENT

6.2.1 Related Work

With the development of body scanning technologies, segmentation and measurement on scanned data has also received attention from researchers. Pargars et al. [157] developed software tools that allow the user to take manual measurements from sliced scan data. Dekker et al. [158] described a model-based approach in which the slice data in each body segment are binned into sectors about the centroid and analyzed to automatically detect landmarks. They also measured the body volume by integrating over the slices and compared the results to that from densitometry. Ju et al. [159] proposed a method that is also based on the information of slices. The body is first segmented into the head and torso, arms, and legs according to slice settings, and then the girth profiles of individual body parts are used to locate the neck, shoulders, waist, elbows, wrists, knees and ankles. Xiao et al. [160] used geodesic distance to segment the body into primary parts. The advantage of this method is that geodesic distance is independent of body postures. Leong et al. [161] proposed an algorithm in which the torso data is

transformed to cylindrical coordinates and then converted into a 2D depth map so that image processing techniques can be used to extract features. Interested readers are referred to a recent review by Werghi [162] on segmentation and modeling of human body from 3D scan data.

6.2.2 3D Measurement Using Graphics Hardware

In the previous work of our lab, Zhong and Xu [163] developed a body segmentation and measurement system that works on triangular meshes with a primary aim for virtual apparel fitting. In this method, key landmarks are searched in some target zones that are predefined based on the proportions relative to the stature. The armpits and neck are searched with the criterion of minimum inclination angle between neighboring triangles. The crotch is detected by observing the transition of cusps along successive horizontal contours. Once the key landmarks have been located, the body is segmented into the torso, head, arms and legs, as shown in Figure 6.1, where the models were captured by the developed stereo vision system. Then various measures including circumferences and lengths are extracted. However, this system is not sufficient for body composition research. For example, it provides limited capability in body volume measurement. To accurately estimate whole body volume, it needs to section the body parts into dense slices and divide each slice into dense line segments. The procedure involves extensive computation of plane-plane and line-line intersections.

Here we present an extension to the previous system. New functions are provided to measure circumferences, areas and volumes. Most of the measurements can be realized by taking advantage of the computational power of modern graphics hardware.

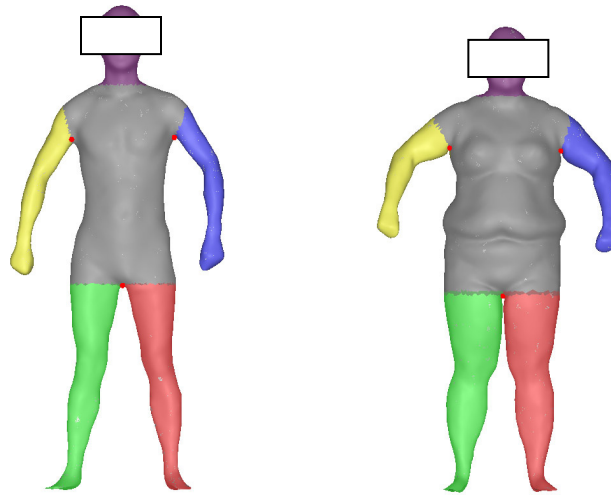


Figure 6.1: Body segmentation.

6.2.2.1 Volume Measurement

For our application, volume measurement can be efficiently performed using the depth buffer of the graphics hardware. In modern computer graphics, the depth buffer, also called the z-buffer, records a depth value for each rendered pixel. With 3D APIs such as OpenGL [164], we can switch the z-buffer to keep track of the minimum or maximum depth (distance to the viewport) for each pixel on the screen. To measure the body volume, the 3D body model is rendered twice in the anterior view. During the two renderings, we choose the z-buffer to record the minimum and maximum depth of each pixel, respectively. Then the two depth maps read from the z-buffer correspond to the front and back surfaces of the body, respectively. As a result, we can get a thickness map of the body from the difference between the two depth maps. An example is shown in Figure 6.2. Finally, the body volume is calculated by integrating over the thickness map based on the known pixel scale. It is worthy noting that orthographic projection should be

used to reflect the actual size of the body. In principle, the z-buffering method is equivalent to resampling the surface data on a regular grid, so the size of the viewport that determines the sampling interval may affect the measure accuracy. However, we found that a moderate size of the viewport such as 500×500 is sufficient to reach high accuracy. In our system, it takes about 50 ms to render a typical model (a triangular mesh of about 15,000 vertices), so this technique is extremely efficient in time cost compared to the slice-based methods.

For segmental volume measurement, we only need to render each individual segment and employ the same z-buffering method.

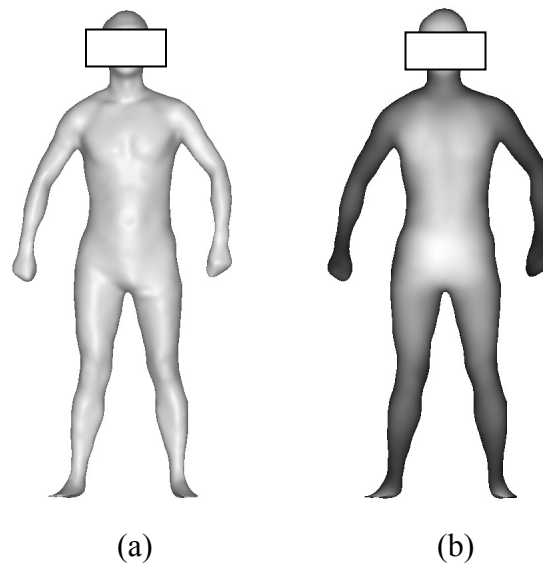


Figure 6.2: Body volume measurement. (a) A 3D body model rendered in the anterior view; and (b) its thickness map.

6.2.2.2 Circumference Measurement

The z-buffering method can also be applied for circumference measurement. It is especially convenient for manual measurement. When the user marks a contour by drawing a line in the rendered image, the 3D data for the contour can be obtained

instantaneously from the depth maps. Then, the circumference as well as the breadth and depth of the contour can be calculated. An example is demonstrated in Figure 6.3. A contour is marked on the body model as shown in Figure 6.3(a), and then its circumference, breadth and depth are calculated as shown in Figure 6.3(b).

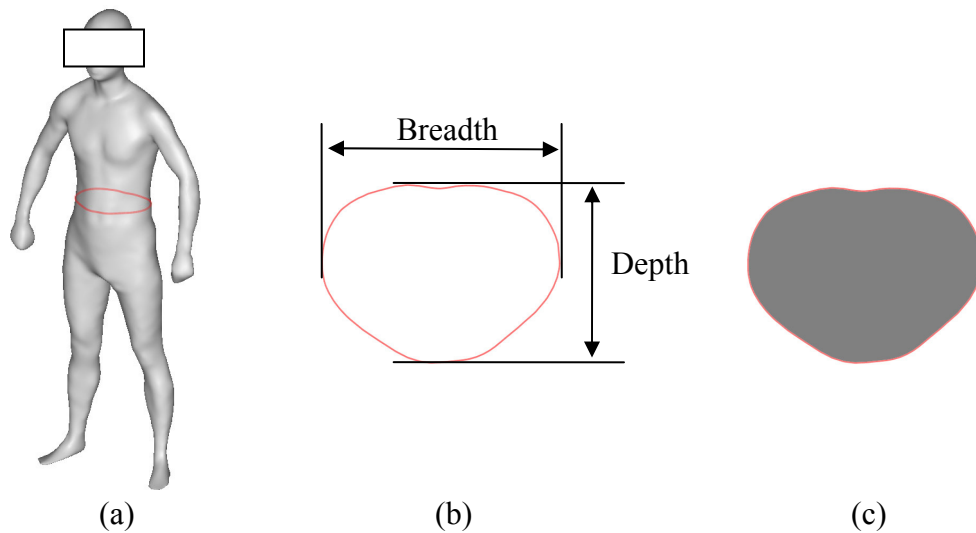


Figure 6.3: Measurements on a contour. (a) The contour is marked on the body model. (b) Circumference, breadth and depth measurements. (d) Cross-sectional area measurement.

6.2.2.3 Area Measurement

When a contour has been extracted in circumference measurement, its cross-sectional area can be estimated by redrawing the contour with its plane coincident with the screen plane. The image of the contour can be read from the depth buffer or the color buffer. The area is calculated by counting the pixels inside the contour. An example is shown in Figure 6.3(c), where the shaded pixels are counted to get the cross-sectional area.

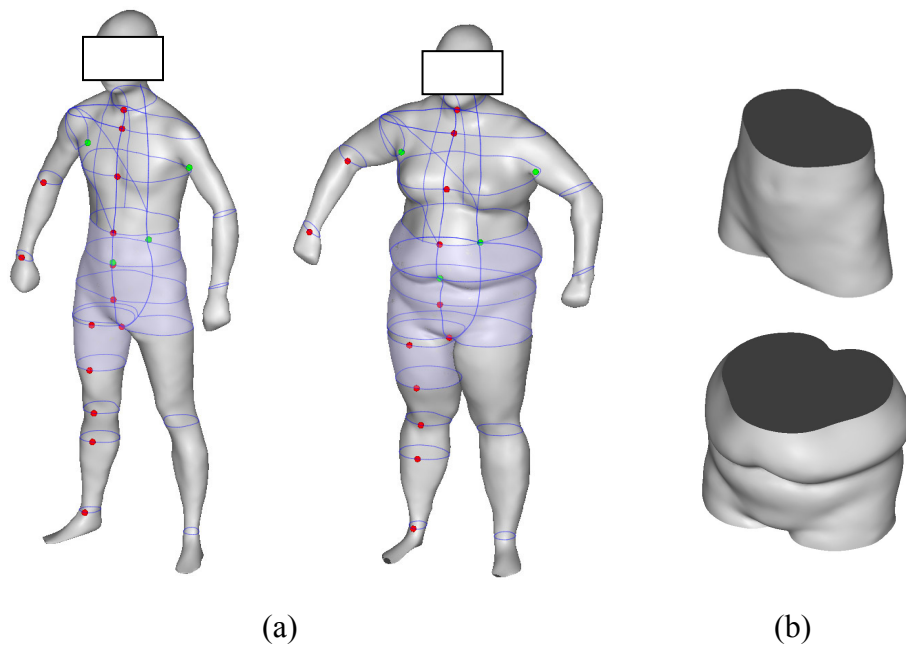


Figure 6.4: Illustration of the body measurement system. (a) Results on two subjects. (b) Body segments used for the abdomen-hip volume measurement.

The body surface area can be estimated by summing up the areas of all triangles in the mesh. In this case, the graphics hardware is not used.

To illustrate the output of the body measurement system, results on two subjects are shown in Figure 6.4. The measured parameters include circumferences and cross-sectional areas of a number of locations (such as the chest, waist, abdomen, hip, upper thigh and so on), whole body volume, segmental volumes (such as the abdomen-hip volume and the upper thigh volume), and body surface area.

6.3 SYSTEM EVALUATION

To evaluate the accuracy and repeatability of the prototype 3D anthropometry system, we have tested it on some mannequins whose dimensions can be measured manually. The system was also tested on human subjects for the measurement of body volume and dimensions. To validate its feasibility in body fat assessment, the system was compared to other methods including air displacement plethysmography (ADP) and bioimpedance analysis (BIA).

6.3.1 Subjects and Methods

6.3.1.1 Mannequins and Measurements

Three mannequins with different sizes (size 8, 10, and 12) were used to evaluate the reliability and accuracy of the system. A MyoTape body tape measure (AccuFitness, LLC, Greenwood Village, CO) was used to measure circumferences, and an anthropometer (Lafayette Instrument Company, Lafayette, IN) was used to measure the depth and breadth of the waist. Each mannequin was imaged 10 times with repositioning in a given hour period. Waist and hip circumferences, waist breadth and depth, and total volume were measured on 3D data automatically. The coefficient of variance (CV) was computed to estimate repeatability. To evaluate the accuracy on circumference, breadth and depth measurements, the results were compared to those obtained with anthropometric methods.

To estimate the longitudinal day-to-day repeatability of the system, the size-12 mannequin was imaged in 10 trials with no more than 3 trials on a single day. For each trial, the measurements were repeated 5 times. The between-trial variance or standard deviation (SD) was estimated.

6.3.1.2 Human Subjects and Measurements

Twenty adult subjects (10 males and 10 females) were recruited in this research. The subjects were aged 24-51 yrs, with weights 47.9-169.5 kg, heights 156.0-193.0 cm, and BMI 18.9-47.8 kg/m². The study was approved by the Institutional Review Board of The University of Texas at Austin. An informed written consent was obtained from each subject at the visit.

The subjects were wearing tight-fitting underwear and a swim cap during the test. First, height, weight, waist and hip circumferences, and waist breadth and depth were measured with conventional anthropometric methods. The same tape and anthropometer were used as in mannequin measurement.

The subjects were imaged with normal breathing by the 3D anthropometry system. During imaging, the subjects were asked to stand still in a specific posture with the legs slightly spread, the arms abducted from the torso, and the hands made into fists. The imaging was repeated 10 times for each subject. The subjects were repositioned between scans. The subjects were also assessed for body fat by ADP (BodPod; Life Measurement Inc, Concord, CA) and leg-to-leg BIA (TBF-300A; Tanita, Tokyo, Japan). The body volume obtained from the 3D anthropometry system should be corrected for thoracic gas volume (TGV) that equals functional residual capacity plus half of tidal volume. In this study, TGV was measured or predicted by the BodPod.

The subjects were instructed to fast at least three hours, stay hydrated, and avoid excessive sweating, heavy exercise, and caffeine or alcohol use before all procedures were performed.

6.3.1.3 Statistical Analysis

Repeatability was determined by computing the intra-class correlation coefficient (ICC) and the coefficient of variance (CV) from the table of one-way random effects

ANOVA. The comparisons of measurement by different methods were performed using *t* tests and linear regression analysis.

Percent body fat was calculated from whole body volume measured by 3D anthropometry and ADP using Siri's Equation (Equation 2.4). Percent body fat estimates determined by ADP and BIA were compared to that obtained by the 3D anthropometry system using paired-sample *t* tests and linear regression. In addition, Bland and Altman analysis [165] was used to assess agreement of percent body fat across methods; a 95% agreement was estimated by the mean difference ± 1.96 SD. For all analyses, statistical significance was $P < 0.05$. The statistical calculations were performed using SPSS 16.0 (SPSS Inc., Chicago, IL).

6.3.2 Results

6.3.2.1 Mannequins

An example of the measurement on one of the mannequins is demonstrated in Figure 6.5. The results of repeatability test on the three mannequins are shown in Table 6.1. The CVs were $< 0.2\%$ for volume and waist and hip circumferences. The CVs increased to around 0.5% for waist breadth and depth due to the relatively small values of these two measures. ICCs were not calculated due to the limited number of subjects. The results of longitudinal repeatability test are given in Table 6.2. The between-trial variance was comparable to the within-trial variance.

The comparison between 3D anthropometry and manual methods on measuring the size-12 mannequins is shown in Table 6.3. Although there were significant differences in three of the four measures, the differences were small. The difference in hip circumference was relatively large because it was difficult to determine the location consistently. The reference value of volume was unknown, so volume was not compared.

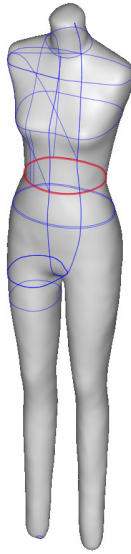


Figure 6.5: Automatic measurement on a mannequin model.

Table 6.1: Repeatability test on three mannequins.

	Mean	MS_w	MS_b	SD_w	SD_b	CV
WC (mm)	675.7	1.3	8806.3	1.2	29.7	0.17
HC (mm)	939.4	1.7	4199.9	1.3	20.5	0.14
Breadth (mm)	231.8	1.7	886.2	1.3	9.4	0.57
Depth (mm)	188.5	0.8	1423.6	0.9	11.9	0.48
Volume (L)	50.752	0.008	47.422	0.092	2.178	0.18

Note: WC, waist circumference; HC, hip circumference. The torso breadth and depth are measured at the waist level. MS_w , within-subject mean square error (MSE); MS_b , between-subject MSE; SD_w , within-subject SD; SD_b , between-subject SD.

Table 6.2: Longitudinal repeatability test on the size-12 mannequin.

	Mean	MS _w	MS _b	SD _w	SD _b
Volume (L)	52.304	0.009	0.051	0.095	0.091

Note: MS_w, within-trial MSE; MS_b, between-trial MSE; SD_w, within-trial SD; SD_b, between-trial SD.

Table 6.3: Dimensions of the size-12 mannequin measured by manual methods and the 3D anthropometry system.

	Tape or anthropometer	3D-A	Difference	<i>P</i>
WC (mm)	704.6 ± 0.8	705.8 ± 1.3	1.2 ± 1.6	0.027
HC (mm)	966.5 ± 1.2	958.9 ± 1.7	-6.6 ± 2.1	< 0.001
Breadth (mm)	238.9 ± 0.1	239.9 ± 1.6	1.0 ± 1.6	0.086
Depth (mm)	201.1 ± 0.0	202.2 ± 0.8	1.1 ± 0.8	0.002

Note: The *P*-values were from *t* tests.

6.3.2.2 Human Subjects

The overall age and anthropometric characteristics of the 20 human subjects are listed in Table 6.4. Eight subjects were of BMI ≥ 25.0 kg/m², and 4 were of BMI ≥ 30.0 kg/m².

The repeatability of the measurements is given in Table 6.5. All ICCs were > 0.99, and all CVs were < 1.0% except the measurement of waist depth. The highest precision was reached in body volume partially due to the fact there was no ambiguity to

calculate whole body volume for a 3D model. However, it was difficult to locate precisely the waist and hip, especially waist in overweight subjects.

The accuracy of 3D anthropometry with reference to tape, anthropometer and ADP measurements is shown in Table 6.6. 3D anthropometry was significantly different from tape (anthropometer) measure in hip circumference and waist depth. The

Table 6.4: Characteristics of the human subjects

	Mean	SD	Range
Age (y)	32.2	6.2	24-51
Height (cm)	171.7	8.4	156.0-193.0
Weight (kg)	79.5	31.3	47.9-169.5
BMI (kg/m ²)	26.6	8.5	18.9-47.8

Table 6.5: Repeatability test on 20 human subjects.

	Mean	MS _w	MS _b	SD _w	SD _b	CV	ICC
WC (mm)	880.3	45.2	495886.7	6.7	222.7	0.76	0.9991
HC (mm)	1065.4	32.8	313843.0	5.7	177.1	0.54	0.9990
Breadth (mm)	305.4	9.0	40913.8	3.0	64.0	0.98	0.9978
Depth (mm)	237.4	19.9	53657.9	4.4	73.2	1.88	0.9963
Volume (L)	80.122	0.156	10523.455	0.394	32.440	0.49	0.9999

Note: WC, waist circumference; HC, hip circumference. The torso breadth and depth are measured at the waist level. MS_w, within-subject MSE; MS_b, between-subject MSE; SD_w, within-subject SD; SD_b, between-subject SD.

Table 6.6: Comparison of dimensions and volume measured by 3D anthropometry, and tape, anthropometer or ADP in human subjects.

	Tape, anthropometer or ADP	3D-A	Difference	<i>P</i>
WC (mm)	884.3 ± 217.6	880.2 ± 222.5	-4.1 ± 29.4	0.543
HC (mm)	1051.2 ± 180.4	1065.1 ± 176.8	13.9 ± 29.2	0.046
Breadth (mm)	314.8 ± 79.2	306.3 ± 65.7	-8.5 ± 24.9	0.152
Depth (mm)	227.0 ± 83.3	240.1 ± 74.5	13.1 ± 25.5	0.038
Volume (L)	76.834 ± 32.445	76.669 ± 32.284	-0.165 ± 0.692	0.300

Note: The *P*-values were from paired-sample *t* tests.

Table 6.7: Linear regression analysis on dimensions and volume measured by 3D anthropometry, and tape, anthropometer or ADP in human subjects.

	<i>a</i>	<i>b</i>	<i>r</i> ²	SEE
WC (mm)	1.014	16.351	0.9827	30.0
HC (mm)	0.967	48.210	0.9739	29.3
Breadth (mm)	0.795	55.954	0.9179	19.4
Depth (mm)	0.854	46.332	0.9097	23.0
Volume (L)	0.995	0.233	0.9996	0.690

Note: The prediction equations are expressed as $y = ax + b$. SEE, standard error of the estimate.

differences were not significant in body volume, and waist circumference and breadth. The degrees of agreement were also characterized by linear regression analysis as shown in Table 6.7. A relatively high correlation was observed between 3D anthropometry and tape (anthropometer) measure in body dimensions with $r^2 > 0.90$, but SEEs were relatively high with the order of 20-30 mm. A very good agreement was reached in body volume as comparing 3D anthropometry to ADP ($r^2 = 0.9996$, $SEE = 0.690$ L). The body volumes are plotted with the regression line in Figure 6.6. The regression line was not significantly different from the line of identity.

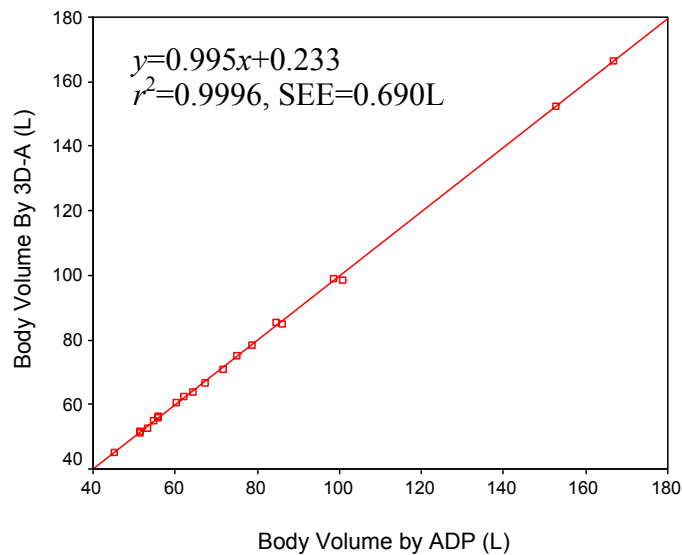


Figure 6.6: Scatter plot of body volume measured by 3D anthropometry (3D-A) and air displacement plethysmography (ADP).

Siri's equation (Equation 2.4) was used to predict body fat for both 3D anthropometry and ADP. The percent body fat (%BF) estimated by these two techniques is shown in Figure 6.7. The prediction equation was obtained from linear regression with ADP as the reference method: $y = 0.891x + 1.917$, $r^2 = 0.9093$, $SEE = 4.002$. As a

comparison, the %BF estimated by BIA and ADP is shown in Figure 6.8, where the prediction line is $y = 0.838x + 2.984$, $r^2 = 0.8174$, $SEE = 5.631$. It was observed that 3D anthropometry and ADP were more closely correlated than BIA and ADP. The %BF estimated by 3D anthropometry and BIA is given in Figure 6.9. The three techniques were also compared using Bland-Altman analysis, as shown in Figures 6.10-12. The bias and SD of difference between each pair of the methods are given in Table 6.8. Paired-sample t tests were also performed and showed that these methods were not significantly different from each other in %BF measurement. However, the limits of agreement demonstrated again that the agreement between 3D anthropometry and ADP was higher than that between BIA and ADP.

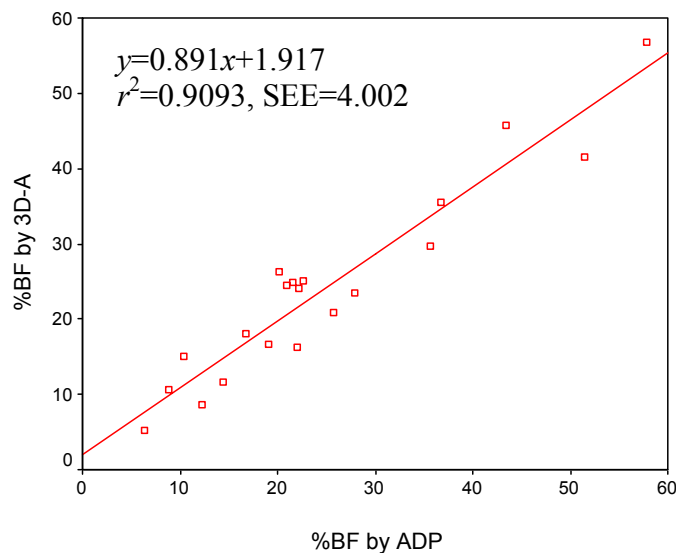


Figure 6.7: Scatter plot of percent body fat (%BF) by 3D anthropometry (3D-A) and air displacement plethysmography (ADP).

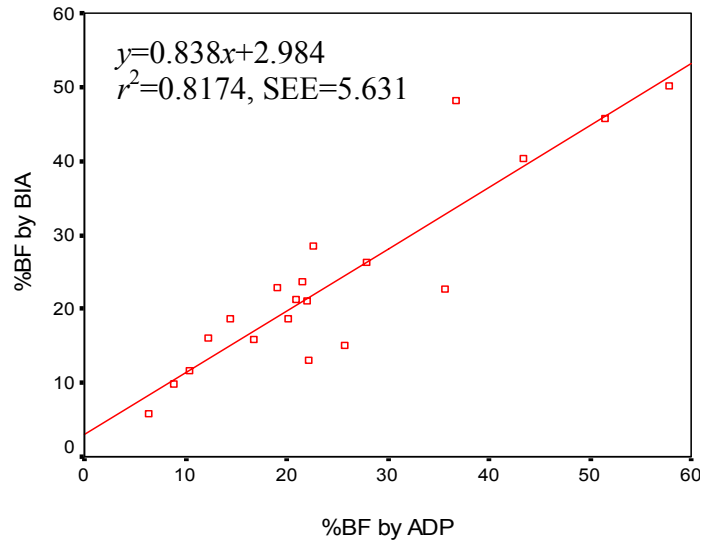


Figure 6.8: Scatter plot of percent body fat (%BF) by bioimpedance analysis (BIA) and air displacement plethysmography (ADP).

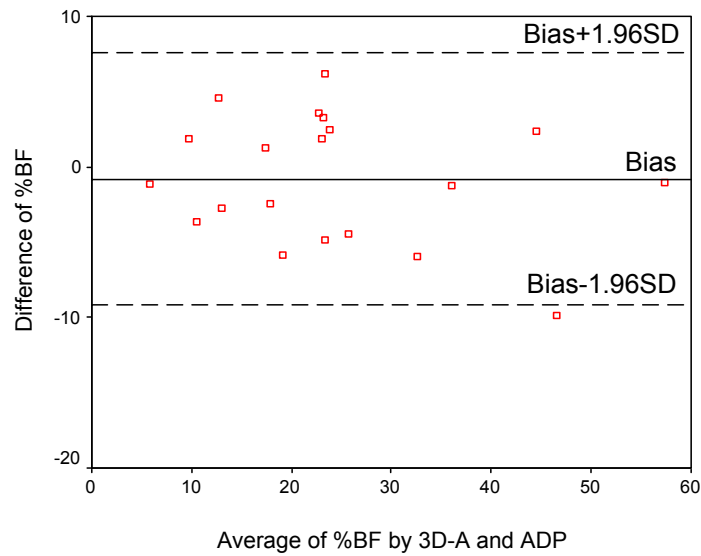


Figure 6.9: Bland-Altman plot of percent body fat (%BF) by 3D anthropometry (3D-A) and air displacement plethysmography (ADP).

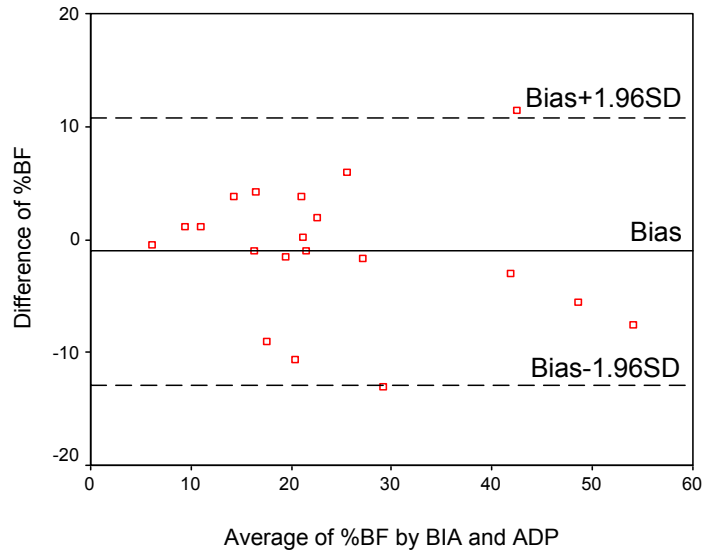


Figure 6.10: Bland-Altman plot of percent body fat (%BF) by bioimpedance analysis (BIA) and air displacement plethysmography (ADP).

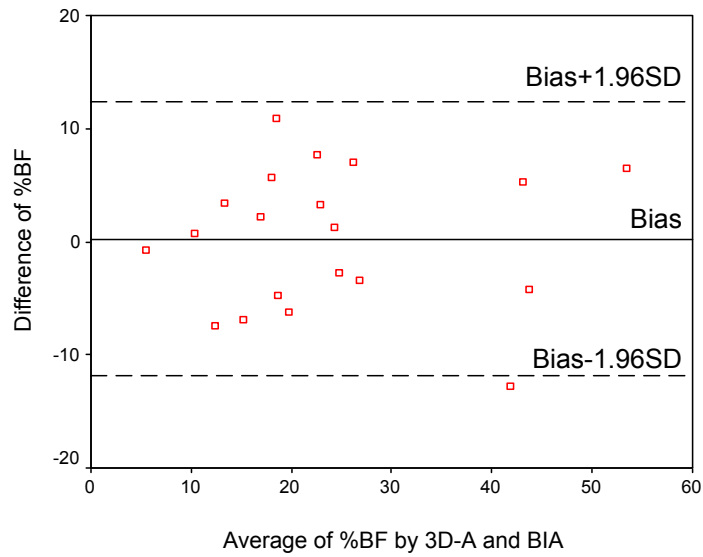


Figure 6.11: Bland-Altman plot of percent body fat (%BF) by 3D anthropometry (3D-A) and bioimpedance analysis (BIA).

Table 6.8: Bland-Altman analysis on percent body fat.

	Bias	SD	Limits of agreement	<i>P</i>
(3D-A) – ADP	-0.789	4.178	± 8.189	0.409
BIA – ADP	-1.040	5.923	± 11.609	0.442
(3D-A) – BIA	0.251	6.177	± 12.107	0.858

Note: Limits of agreement is defined as ± 1.96 SD. The *P* values were from paired-sample *t* tests.

6.4 DISCUSSION

We have presented an automatic body measurement system, which is an extension to its earlier version and dedicated to the needs of body composition assessment. The functions of 3D measurement can be enhanced by taking advantage of modern graphics hardware.

The overall performance of the developed 3D anthropometry system has been evaluated. The measurements were highly repeatable both in mannequins and human subjects. Relatively large differences were observed for circumferences, breadth and depth in human subjects. Most likely, the errors mainly originated from the inconsistency of locating the landmarks between different methods. For example, the level of waist is usually the narrowest part of the torso for individuals of normal weight. However, the location of waist in obese is not well defined. For manual measurement, we can determine waist as midway between iliac crest and lowest rib margin [166]. But we cannot take advantage of this skeletal information when performing measurement on a 3D body model. The error in the waist depth measurement was larger than that in the

waist breadth for human subjects, perhaps due to the fact that breathing had a greater effect on the depth. Once the uncertainties had been minimized, as in the measurements of mannequins, high accuracy could be reached.

The body volumes measured by 3D anthropometry and ADP were highly correlated, which was essential to effective body fat assessment. In body fat estimate, 3D anthropometry and ADP had closer agreement than BIA and ADP. However, we need to realize that the estimation of percent body fat is very sensitive to the accuracy of body volume measurement in the two-component body composition model. For example, Siri's equation (Equation 2.4) yields

$$\Delta(\%BF) = \frac{495\Delta V}{W}, \quad (6.1)$$

where W is the body weight in kg, and ΔV is the error of body volume measurement in L. If we assume $W = 60$ kg, then an error of 0.5 L in ΔV would lead to an over 4% difference in %BF. A small error in body volume measurement can readily result from inaccuracy of lung volume estimate or a slight movement of the body during imaging. However, body volume is only one of a number of variables that can be measured from a 3D anthropometry system. Its combination with other variables may offer better prediction of body fat.

Chapter 7

Conclusions and Future Work

7.1 SUMMARY OF THE DISSERTATION

The prevalence of obesity has made it necessary to develop a convenient, reliable and safe tool for timely assessing and monitoring obesity in public health. After reviewing current techniques for body composition analysis, we suggested that 3D anthropometry can provide a convenient, accommodating and comprehensive means to body composition assessment.

A 3D anthropometry system based on stereo vision technology was developed. To make it more affordable and portable, the system was reduced to a two-stance design that is the minimum configuration required for whole body imaging. The system is calibrated in two stages: camera calibration and 3D registration. The first stage is relatively complicated, but it doesn't need to be repeated frequently. The relative position of two cameras in a stereo head can be readily fixed and intrinsic camera parameters can be stabilized using locking lenses. Therefore, only 3D registration needs to be redone when the system is transported. This property contributes to the portability of the system and also reduces cost of maintenance.

The hardware requirements of stereo vision are relatively low in comparison with laser scanning and structured light, the most popular technologies in body scanning. Additionally, fast image acquisition can be reached in stereo vision. However, the computation in stereo vision is complex and intensive, and is still a big challenge. In this study, we proposed a two-phase stereo matching algorithm. In the first phase, foreground is accurately segmented with the help of a predefined virtual interface, and a coarse

disparity map is generated with block matching. In the second phase, local least squares matching is combined to global optimization within a regularization framework, so both high accuracy and reliability can be reached.

To make the 3D data more interpretable and manageable, it is essential to convert the raw 3D data to a body surface model. For our system, a unique challenge is that there are large gaps in the data caused by occlusions. An effective surface reconstruction algorithm based on subdivision surface representation was developed. The algorithm consists of four steps. First, the data is resampled on a regular grid. Second, the sampled data are triangulated to create an initial mesh with gap closure. Third, the initial mesh is simplified to generate an approximate control mesh. Finally, the control mesh is optimized and subdivided to obtain a smooth model. Our results demonstrated the algorithm is reliable in gap closing, efficient in data compression, and accurate in representation.

To make the 3D anthropometry system ready for practical use, automatic body measurement is indispensable. A body measurement system dedicated to body composition assessment was developed based on an earlier system that was designed for applications in apparel fitting. The functions of 3D measurement were enhanced by taking advantage of modern graphics hardware. The measurable parameters include circumferences, frontal and sagittal diameters, whole body volume, segmental volumes, cross-sectional areas, and body surface area.

The overall performance of the presented system was evaluated. The measurements were highly repeatable. The feasibility of 3D anthropometry in body fat assessment was demonstrated in comparison to ADP and BIA. The results showed that the agreement between 3D anthropometry and ADP was higher than that between BIA and ADP

7.2 SUGGESTIONS ON FUTURE WORK

The current dimensions and portability of the system are constrained by the field of view of the projectors. The system can be made more compact if projectors with shorter throw distance are available. Furthermore, true portability can be realized if the need of projectors can be eliminated. In an unreported study, we have successfully applied a variant of our stereo matching algorithm to human face imaging using 10-megapixel consumer digital cameras without texture projection. However, we need to increase considerably the number of cameras for whole body imaging since each camera can only cover a limited field view under which skin texture is sufficient for reliable stereo matching. Here the major challenge is that the computation cost will increase dramatically due to the large image size and number of cameras. In addition, it will become more difficult to control the systematic error caused by imperfect calibration when more cameras are involved.

There is still room for improvement on the algorithms developed in this study. For example, our stereo matching algorithm is effective in foreground segmentation, but a slight “foreground fattening” effect is still noticeable in some regions. As for surface reconstruction, some features of the body are over smoothed, so it is preferred to make the algorithm more adaptive to local geometry.

The potential of the applications of 3D anthropometry in public health is enormous. For example, it will be of great value if we can develop new indexes for estimating the distribution of body fat or more directly predicting health risks. This technology is also ideal for tracing changes in body size and shape and monitoring related health conditions.

Bibliography

- [1] WHO. Obesity: Preventing and managing the global epidemic. Report of WHO Consultation on Obesity, Geneva, June 1997.
- [2] NIH. Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: the evidence report. NIH Publication No. 98-4083, September 1998.
- [3] HHS/NIH. Strategic plan for NIH obesity research. NIH Publication No. 04-5493, August 2004. <http://www.obesityresearch.nih.gov/About/strategic-plan.htm>.
- [4] Peeters A, Barendregt JJ, Willekens F, Mackenbach JP, Manun AAL, and Bonneux L. Obesity in adulthood and its consequences for life expectancy: a life-table analysis. *Annals of Internal Medicine*, 138: 24-32, January, 2003.
- [5] Bjorntorp P. The regulation of adipose tissue distribution in humans. *International Journal of Obesity*, 20: 291-302, 1996.
- [6] Bjorntorp P. Centralization of body fat. In *International Textbook of Obesity*, ed. Bjorntorp, pp. 213-224. New York, NY: John Wiley & Sons, 2001.
- [7] Lean ME, Han TS, and Morrison CE. Waist circumference as a measure for indicating need for weight management. *BMJ*, 311: 158-64, 1995.
- [8] Behnke AR, Feen BG, and Welham WC. The specific gravity of healthy man. *Journal of the American Medical Association*, 118: 495-498, 1942.
- [9] Dempster P and Aitkens S. A new air displacement method for the determination of human body composition. *Medicine & Science in Sports & Exercise*, 27: 1692-1697, 1995.
- [10] Thomasset A. Bio-electrical properties of tissue impedance measurements. *Lyon Medical*, 1962. 207: 107-118.
- [11] Mazess RB, et al. Dual-energy X-ray absorptiometry for total-body and regional bone-mineral and soft-tissue composition. *American Journal of Clinical Nutrition*, 51: 1106-1112, 1990.
- [12] Sjostrom L. A computer-tomography based multi-compartment body composition technique and anthropometric predictions of lean body mass, total and subcutaneous adipose tissue. *International Journal of Obesity*, 15: 19-30, 1990.
- [13] Ross R, et al. Quantification of adipose tissue by MRI: Relationship with anthropometric variables. *Journal of Applied Physiology*, 72: 787-795, 1992.

- [14] Wells JCK, Ruto A, and Treleaven P. Whole-body three-dimensional photonic scanning: a new technique for obesity research and clinical practice. *International Journal of Obesity*, 32: 232-238, 2008.
- [15] Istook CL and Hwang S-J. 3D body scanning systems with application to the apparel industry. *Journal of Fashion Marketing and Management*, 5(2): 120-132, 2001.
- [16] Thalmann D, Shen J, and Chauvineau E. Fast realistic human body deformations for animation and VR applications. In *Computer Graphics International*, Pohang, Korean, June 1996.
- [17] Gardner G and Halweil B. Underfed and overfed: the global epidemic of malnutrition. Washington, DC: Worldwatch Institute, 2000. (Worldwatch paper no. 150).
- [18] National Center for Health Statistics. *Health, United States, 2007*, with Chartbook on Trends in the Health of Americans. Hyattsville, MD: 2007.
- [19] Hossain P, Kavar B, Nahas ME. Obesity and diabetes in the developing world—a growing challenge. *New England Journal of Medicine*, 356(3): 213-215, 2007.
- [20] Kolanowski J. Obesity and hypertension: from pathophysiology to treatment. *International Journal of Obesity*, 23(Suppl. 1): 42-46, 1999.
- [21] Mertens IL and Van Gaal LF. Overweight, obesity and blood pressure: the effects of moderate weight reduction—a review. *Obesity Research*, 8: 270-278, 2000.
- [22] Rahmouni K, Correia MLG, Haynes WG and Mark AL. Obesity-associated hypertension: new insights into mechanisms. *Hypertension*, 45: 9-14, 2005.
- [23] Mokdad AH, Ford ES, Bowman BA, et al. Prevalence of obesity, diabetes, and obesity-related health factors. *JAMA*, 2003(289) 76-79, 2001.
- [24] Smyth S and Heron A. Diabetes and obesity: the twin epidemics. *Nature Medicine*, 12: 75-80, 2006.
- [25] Wilson PW. Established risk factors and coronary artery disease: the Framingham study. *American Journal of Hypertension*, 7: 7S-12S, 1994.
- [26] Sundquist J and Winkleby MA. Cardiovascular risk factors in Mexican American adults: a transcultural analysis of NHANES III, 1988-1994. *American Journal of Public Health*, 89: 723-730, 1999.

- [27] US Surgeon General. Overweight and obesity: health consequences. http://www.surgeongeneral.gov/topics/obesity/calltoaction/fact_consequences.htm
- [28] American Dietetic Association. Position of the American Dietetic Association: Weight management. *Journal of American Dietetic Association*, 97: 71-74, 1997.
- [29] Haslam D and James W. Obesity. *The Lancet*, 366(9492): 1197-1209, 2005.
- [30] Wang ZM, Pierson RNJ, and Heymsfield SB. The five level model: A new approach to organizing body composition research. *American Journal of Clinical Nutrition*, 56: 19-28, 1992.
- [31] Heymsfield SB, Lohman TG, Wang ZM, and Going SB. *Human Body Composition*, 2nd Edition. Champaign, IL: Human Kinetics, 2005.
- [32] Wilmore JH. A simplified method for determination of residual lung volume. *Journal of Applied Physiology*, 27: 96-100, 1969.
- [33] Buskirk, ER. Underwater weighing and body density: A review of procedures. In *Techniques for Measuring Body Composition*, ed. J. Brozek and A. Henschel, pp. 90-105. Washington, DC: National Academy of Sciences, 1961.
- [34] Siri WE. Body composition from fluid spaces and density: Analysis of methods. In *Techniques for measuring body composition*, ed. J. Brozek and A. Henschel, pp. 223-244. Washington, DC: National Academy of Sciences, 1961.
- [35] Brozek J, et al. Densitometric analysis of body composition: revision of some quantitative assumptions. *Annals of the New York Academy of Sciences*, 110: 113-140, 1963.
- [36] Crapo RO, Morriss AH, Clayton PD, and Nixon CR. Lung volumes in healthy nonsmoking adults. *Bull Eur Physiopath Respir*, 18: 419-425, 1982.
- [37] Dubois D and Dubois EF. A formula to estimate the approximate surface area if height and weight be known. *Arch Intern Med*, 17: 863-871, 1916.
- [38] Jackson AS and Pollock ML. Generalized equations for predicting body density of men. *British Journal of Nutrition*, 40: 487-504, 1978.
- [39] Lohman TG. Skinfolds and body density and their relation to body fatness: a review. *Human Biology*, 53: 181-225, 1981.
- [40] Tran ZV and Weltman A. Predicting body composition of men from girth measurements. *Human Biology*, 60: 167-175, 1988.

- [41] Tran ZV and Weltman A. Generalized equation for predicting body density of women from girth measurements. *Medicine & Science in Sports & Exercise*, 21: 101-104, 1989.
- [42] Weltman A, et al. Accurate assessment of body composition in obese females. *American Journal of Clinical Nutrition*, 48: 1179-1183, 1988.
- [43] Weltman A, et al. Practical assessment of body composition in adult obese males. *Human Biology*, 59: 523-535, 1987.
- [44] Seip R and Weltman A. Validity of skinfold and girth based regression equations for the prediction of body composition in obese adults. *American Journal of Human Biology*, 3: 91-95, 1991.
- [45] Hoffer EC, et al. Correlation of whole-body impedance with total body water volume. *Journal of Applied Physiology*, 27: 531-534, 1969.
- [46] Houtkooper LB, et al. Why bioelectrical impedance analysis should be used for estimating adiposity. *American Journal of Clinical Nutrition*, 64: 436S-448S, 1996.
- [47] Cornish B, Thomas B and Ward L. Improved prediction of extracellular and total body water using impedance loci generated by multiple frequency bioelectrical impedance analysis. *Physics in Medicine and Biology*, 38: 337-346, 1993.
- [48] Deurenberg P. Limitation of the bioelectrical impedance method for the assessment of body fat in severe obesity. *American Journal of Clinical Nutrition*, 64: 449S-452S, 1996.
- [49] Pietrobelli A, Formica C, Wang Z, and Heymsfield SB. Dual-energy X-ray absorptiometry body composition model: review of physical concepts. *American Journal of Physiology*, 271: E941-E951, 1996.
- [50] Sutcliffe JF. A review of in vivo experimental methods to determine the composition of the human body. *Physics in Medicine and Biology*, 41: 791-833, 1996.
- [51] Borkan GA, Gerzof SG, Robbins AH, Hults DE, Silbert CK, and Silbert JE. Assessment of abdominal fat content by computed tomography. *American Journal of Clinical Nutrition*, 36: 172-177, 1982.
- [52] Foster MA, Hutchison JMS, Mallard JR, Fuller M. Nuclear magnetic resonance pulse sequence and discrimination of high- and low-fat tissues. *Magnetic Resonance Imaging*, 2: 187-192, 1984.

- [53] Robinette KM, Vannier MW, and Jones PRM. 3-D Surface Anthropometry: Review of Technologies. AGARD, Neuilly-sur-Seine, 1997.
- [54] D'Apuzzo N. State of the art of the methods for static 3D scanning of partial or full human body. In Proceedings of Conference on 3D Modeling, Paris, France, June 13-14, 2006.
- [55] Daanen HAM and van de Water GJ. Whole body scanners. *Displays*, 19: 111-120, 1998.
- [56] Wells JCK, Douros I, Fuller NJ, Elia M, Dekker L. Assessment of body volume using three-dimensional photonic scanning. *Annals of the New York Academy of Sciences*, 904: 247-254, 2000.
- [57] Wang J, Gallagher D, Thornton JC, Yu W, Horlick M, and Pi-Sunver FX. Validation of a 3-dimensional photonic scanner for the measurement of body volumes, dimensions, and percentage body fat. *American Journal of Clinical Nutrition*, 83(4): 809-816, 2006.
- [58] Heyward VH and Wagner DR. *Applied Body Composition Assessment*. 2nd Edition. Champaign, IL: Human Kinetics, 2004.
- [59] Zamboni M, et al. Sagittal abdominal diameter as a practical predictor of visceral fat. *International Journal of Obesity and Related Metabolic Disorders*, 22: 655-660, 1998.
- [60] Ohrvall M, Berglund L, and Vessby B. Sagittal abdominal diameter compared with other anthropometric measurements in relation to cardiovascular risk. *International Journal of Obesity and Related Metabolic Disorders*, 24: 497-501, 2004.
- [61] Hertzberg HTE, Dupertuis CW, and Emanuel I. Stereophotogrammetry as an anthropometric tool. *Photogrammetric Engineering*, 32: 942-947, 1957.
- [62] Herron RE. Biostereometric measurement of body form. *Yearbook of Anthropometry*, 16: 80-121, 1972.
- [63] Whittle MW, Herron R, and Cuzzi J. Biostereometric analysis of body form. In *Biomedical Results from Skylab*, ed. R.S. Johnston and L.F. Dietlein. Washington: Scientific and Technical Information Office, NASA, 1977.
- [64] Sheffer DB, Price TE, Loughry CW, Bolyard BL, Morek WM, and Varga RS. Validity and reliability of biostereometric measurement of the human female breast. *Annals of Biomedical Engineering*, 14: 1-14, 1986.

- [65] Loughry CW, Sheffer DB, and Price TE. Breast volume measurement of 598 women using biostereometric analysis. *Annals of Plastic Surgery*, 22(5): 380-385, 1989.
- [66] Addleman D and Addleman L. Rapid 3D digitizing. *Computer Graphics World*, 8: 42-44, 1985.
- [67] Jones PRM, West GM, Harris DH, and Read JB. The Loughborough anthropometric shadow scanner (LASS). *Endeavour*, 13: 162-168, 1989.
- [68] Nahas M, Huitric H, Rioux M, and Domey J. Registered 3D-texture imaging. In *Proceedings of the Computer Animation '90 Conference*, Geneva, Switzerland, April 25-27, 1990, pp. 81-91.
- [69] Demers MH, Hurley JD, Wulpern RC, and Grindon JR. Three-dimensional surface capture for body measurements using projected sinusoidal patterns. In *IS&T/SPIE's 9th Annual Symposium on Electronic Imaging: Science and Technology*, San Jose, CA, USA, 1997.
- [70] InSpeck Inc. <http://www.inspeck.com>.
- [71] Chen F, Brown GM, and Song M. Overview of three-dimensional shape measurement using optical methods. *Optical Engineering*, 39(1): 10-22, 2000.
- [72] Siebert JP and Marshall SJ. Human body 3D imaging by speckle texture projection photogrammetry. *Sensor Review*, 20(3): 218-226, 2000.
- [73] 3dMD. <http://www.3dmd.com>.
- [74] D'Apuzzo N. 3D body scanning technology for fashion and apparel industry. In *Proceedings of SPIE-IS&T Electronic Imaging*, vol. 6491, San Jose, CA, 2007.
- [75] Paquette S. 3D Scanning in apparel design and human engineering. *IEEE Computer Graphics and Applications*, 16(5): 11-15, 1996.
- [76] Cordier F, Hyewon S, and Magnenat-Thalmann N. Made-to-measure technologies for an online clothing store. *IEEE Computer Graphics and Applications*, 23(1): 38-48, 2003.
- [77] Magnenat-Thalmann N, Seo H, and Cordier F. Automatic modeling of animatable virtual humans—a survey. In *Fourth International Conference on 3-D Digital Imaging and Modeling (3DIM 2003)*, pp. 2-10, 2003.
- [78] Galdino M, et al. Clinical applications of three-dimensional photography in breast surgery. *Plastic and Reconstructive Surgery*, 110(1): 58-70, 2002.

- [79] Lee S. Three-dimensional photography and its application to facial plastic surgery. *Archives of Facial Plastic Surgery*, 6: 410-414, 2004.
- [80] Tikuisis P, Meunier P, and Jubenville CE. Human body surface area: measurement and prediction using three dimensional body scans. *European Journal of Applied Physiology*, 85: 264-271, 2001.
- [81] Robinette KM, Daanen H, and Paquet E. The CAESAR project: a 3-D surface anthropometry survey. In *Proceedings of the Second International Conference on 3-D Digital Imaging and Modeling*, pp. 380-386, 1999.
- [82] Wells JCK, Treleaven P, and Cole TJ. BMI compared with 3-dimensional body shape: the UK national sizing survey. *American Journal of Clinical Nutrition*, 85(2): 419-425, 2007.
- [83] Wells JCK, Cole TJ, Bruner D, and Treleaven P. Body shape in American and British adults: between country and inter-ethnic comparisons. *International Journal of Obesity*, 32: 152-159, 2008.
- [84] Treleaven PC. Sizing us up. *IEEE Spectrum*, 41: 28 –31, 2004.
- [85] TC². <http://www.tc2.com>.
- [86] Cyberware. <http://www.cyberware.com>.
- [87] Besl PJ. Active optical range imaging sensors. In *Advances in Machine Vision*, ed. J.L.C. Sanz, pp. 1-63. New York, NY: Springer-Verlag, 1989.
- [88] Borghese N, et al. Autoscan: a flexible and portable 3D scanner. *IEEE Computer Graphics and Applications*, 18(3): 38-41, 1988.
- [89] Battle J, Mouaddib E, and Salvi J. Recent progress in coded structured light as a technique to solve the correspondence problem: a survey. *Pattern Recognition*, 31(7): 963-982, 1998.
- [90] Salvi J, Pages J, and Battle J. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4): 827-849, 2004.
- [91] Gühring J. Dense 3-d surface acquisition by structured light using off-the-shelf components. *Proc. SPIE Videometrics and Optical Methods for 3D Shape Measurement*, 4309, pp. 220–231, 2001.
- [92] Dhond U and Aggarwal JK. Structure from stereo—a review. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6): 1489-1510, 1989.

- [93] Scharstein D and Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1/2/3): 7-42, 2002.
- [94] Xu B and Huang Y. 3D technology for apparel mass customization, Part I: Rotary body scanning. *Journal of Textile Institute*, 94(1): 72-80, 2003
- [95] Konolige K and Beymer D. *Small Vision System Calibration Supplement to the User's Manual*, SRI International, May, 2007.
- [96] Tsai RY. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, RA-3(4): 323-344, 1987.
- [97] Heikkila J and Silven O. A four-step camera calibration procedure with implicit image correction. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, pp. 1106, 1997.
- [98] Bouguet J-Y. *Camera Calibration Toolbox for Matlab*, http://www.vision.caltech.edu/bouguetj/calib_doc/.
- [99] Zhang Z. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11): 1330-1334, 2000.
- [100] Horn BKP. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, 4(4): 629-642, 1987.
- [101] Brown MZ, Burschka D, and Hager GD. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8): 993-1008, 2003
- [102] Cox IJ, Hingorani SL, Rao SB, and Maggs BM. A maximum likelihood stereo algorithm. *Computer Vision and Image Understanding*, 63(3): 542-567, 1996.
- [103] Scharstein D. Matching images by comparing their gradient fields. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, 1: 572-575, 1994.
- [104] Sanger TD. Stereo disparity computation using Gabor filters. *Biological Cybernetics*, 59: 405-418, 1988.
- [105] Fleet DJ. Phase-based disparity measurement. *CVGIP: Image Understanding*, 53(2): 198-210, 1991.
- [106] Chen T-Y, Bovik AC, and Cormack LK. Stereoscopic ranging by matching image modulations. *IEEE Transactions on Image Processing*, 8(6): 785-797, 1999.

- [107] Grimson WEL. Computational experiments with a feature-based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(1): 17-34, 1985.
- [108] Pollard SB, Mayhew JEW, and Frisby JP. PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14: 449-470, 1981.
- [109] Medioni G and Nevatia R. Segment-based stereo matching. *Computer Vision, Graphics, and Image Processing*, 31(1): 2-18, 1985.
- [110] Kim NH and Bovik AC. A contour-based stereo matching algorithm using disparity continuity. *Pattern Recognition*, 21(5): 505-514, 1988.
- [111] Poggio T, Torre V, and Koch C. Computational vision and regularization theory. *Nature*, 317(26): 314-319, 1985.
- [112] Okutomi M and Kanade T. A locally adaptive window for signal matching. *International Journal of Computer Vision*, 7(2): 143-162, 1992.
- [113] Veksler O. Stereo matching by compact windows via minimum ratio cycle. In *Proceedings of Eighth IEEE International Conference on Computer Vision (ICCV 2001)*, 1: 540-547, 2001.
- [114] Marr D and Poggio T. Cooperative computation of stereo disparity. *Science*, 194: 209-236, 1976.
- [115] Zitnick CL and Kanade T. A cooperative algorithm for stereo matching and occlusion detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7): 675-684, 2000.
- [116] Terzopoulos D. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(4): 413-424, 1986.
- [117] Tardon-Garcia L-J, Portillo-Garcia J, and Alberola-Lopez C. Markov random field and the disparity gradient constraint applied to stereo correspondence. In *Proceedings of 1999 International Conference on Image Processing*, 3: 901-905, 1999.
- [118] Geiger D, Ladendorf B, and Yuille A. Occlusions and binocular stereo. *International Journal of Computer Vision*, 14: 211-226, 1996.
- [119] Sun C. Fast stereo matching using rectangular subregioning 3D maximum-surface techniques. *International Journal of Computer Vision*, 47(1/2/3): 99-117, 2002.

- [120] Boykov Y, Veksler O, and Zabih R. Fast approximation energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11): 1222-1239, 2001.
- [121] Kolmogorov V and Zabih R. Computing visual correspondence with occlusions using graph cuts. In *Proceedings of Eighth IEEE International Conference on Computer Vision (ICCV 2001)*, 2: 508-515, 2001.
- [122] Roy S. Stereo without epipolar lines: a maximum-flow formulation. *International Journal of Computer Vision*, 34(2/3): 147-161, 1999.
- [123] Sun J, Zheng N-N, and Shum H-Y. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7): 787-800, 2003.
- [124] Felzenszwalb PF and Huttenlocher DR. Efficient belief propagation for early vision. In *Proc. IEEE CVPR 2004*, p. I-261-I-268, 2004.
- [125] Yedidia JS, Freeman WT, and Weiss Y. Bethe free energy, Kikuchi approximations, and belief propagation algorithms. Technical Report TR-2001-16, Mitsubishi Electric Research, 2001.
- [126] Pearl J. *Probabilistic reasoning in intelligent systems: networks of plausible inference*, San Mateo, CA: Morgan Kaufmann Publishers, 1988.
- [127] Srinivas M and Patnaik LM. Genetic algorithms: a survey. *Computer*, 27(6): 17-26, 1994.
- [128] Gong M and Yang Y-H. Genetic-based stereo algorithm and disparity map evaluation. *International Journal of Computer Vision*, 47(1/2/3): 63-77, 2002.
- [129] Han K-P, Song K-W, Chung E-Y, Cho S-J, and Ha Y-H. Stereo matching using genetic algorithm with adaptive chromosomes. *Pattern Recognition*, 34(9): 1729-1740, 2001.
- [130] Lin M and Tomasi C. Surface with occlusions from layered stereo. *Conference on Computer Vision and Pattern Recognition*, pp. 710-717, 2003.
- [131] Kolmogorov V, Criminisi A, Blake A, Cross G, and Rother C. Bi-layer segmentation of binocular stereo video. *Proc. IEEE Computer Vision and Pattern Recognition (CVPR) 2005*, San Diego, CA.
- [132] Bleyer M and Gelautz M. A layered stereo matching algorithm using image segmentation and global visibility constraints. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(3): 128-150, 2005.

- [133] Lewis JP. Fast Template Matching. *Vision Interface*, pp. 120-123, 1995.
- [134] Hartley R and Zisserman A. *Multiple View Geometry in Computer Vision*, Cambridge University Press: Cambridge, UK, 2000.
- [135] Faugeras O and Luong Q-T. *The Geometry of Multiple Images*. Cambridge, MA: MIT Press, 2001.
- [136] Greig D, Porteous B, and Seheult A. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B*, 51(2): 271-279, 1989.
- [137] Ford L and Fulkerson D. *Flows in Networks*. Princeton University Press, 1962.
- [138] Huang TS. *Two-Dimensional Signal Processing II: Transforms and Median Filters*. Berlin: Springer-Verlag, 1981.
- [139] Dougherty ER. *Hands-on Morphological Imaging Processing*. Bellingham, Washington: SPIE Press, 2003.
- [140] Westerweel J. *Digital Particle Image Velocimetry: Theory and Application*. Delft University Press, 1993.
- [141] Shimizu M and Okutomi M. Sub-pixel estimation error cancellation on area-based matching. *International Journal of Computer Vision*, 63(3): 207-224, 2005.
- [142] Nehab D, Rusinkiewicz S and Davis J. Improved sub-pixel stereo correspondences through symmetric refinement. *International Conference on Computer Vision (ICCV)*, October 2005.
- [143] Stein A, Huertas A, and Matthies L. Attenuating stereo pixel-locking via affine window adaptation. *IEEE International Conference on Robotics and Automation*, May, 2006, pp. 914-921.
- [144] Lucas BD and Kanade T. An iterative image registration technique with an application to stereo vision. In *Proceedings of Imaging Understanding Workshop*, pp. 121-130, 1981.
- [145] Shewchuk JR. *An Introduction to the Conjugate Gradient Method without the Agonizing Pain*, Carnegie Mellon University, Pittsburgh, PA, 1994.
- [146] Piegl L. On NURBS: a survey, *IEEE Computer Graphics and Applications*, 11(1): 55-71, 1991.

- [147] Warren J and Weimer H. *Subdivision Methods for Geometric Design: A Constructive Approach*, San Francisco, Calif.: Morgan Kaufmann Publishers, 2002.
- [148] Douros I, Dekker L, and Buxton B. Reconstruction of the surface of the human body from 3D scanner data using B-splines. In *SPIE Proceedings*, 3640: 235-245, 1999.
- [149] Douros I, Dekker L, and Buxton B. An improved algorithm for reconstruction of the surface of the human body from 3D scanner data using local B-spline patches. In *Proceedings of the IEEE International Workshop on Modelling People*, pp. 29-36, 1999.
- [150] Xu B, Yu W, and Chen T. 3D technology for apparel mass customization, Part II: Human body modeling from unorganized range data. *Journal of Textile Institute*, 94(1): 81-911, 2003.
- [151] Allen B, Curless B, and Popovic Z. The space of human body shapes: reconstruction and parameterization from range scans. In *Proceedings of ACM SIGGRAPH*, pp. 587-594, 2003.
- [152] Hoppe H, et al. Surface reconstruction from unorganized points. In *Proceedings of ACM SIGGRAPH'92*, pp. 71-78, 1992.
- [153] Hoppe H, et al. Piecewise smooth surface reconstruction. In *Computer Graphics Proceedings, Annual Conference Series*, pp. 295-302, 1994.
- [154] Fuchs H, Kedem ZM, and Uselton SP. Optimal surface reconstruction from planar contours. *Communications of the ACM*, 20(10): 693-702, 1977.
- [155] Garland M. *Quadric-Based Polygonal Surface Simplification*, Ph.D. thesis, Carnegie Mellon University, School of Computer Science, 1999.
- [156] Loop C. *Smooth Subdivision Surfaces Based on Triangles*, Master's thesis, University of Utah, Department of Mathematics, 1987.
- [157] Pargas RP, Staples NJ, and Davis JS. Automatic measurement extraction for apparel from a three-dimensional body scan. *Optics and Lasers in Engineering*, 28:157-172, 1997.
- [158] Dekker L, Douros I, Buxton BF, and Treleaven PC. Building symbolic information for 3D human body modeling from range data. In *Proc. the Second International Conference on 3D Digital Imaging and Modeling*, IEEE Computer Society, Ottawa, ON, Canada, 1999, pp. 388-397.

- [159] Ju X, Weighi N and Siebert JP. Automatic segmentation of 3D human body scans. In Proc. IASTED Int. Conf. on Computer Graphics and Imaging 2000, Las Vegas, USA, 2000.
- [160] Xiao Y, Siebert P and Werghi N. Topological segmentation of discrete human body shapes in various posture based on geodesic distance. In Proc. the 17th International Conference on Pattern Recognition, 2004, vol. 3, pp. 131-135.
- [161] Leong L-F, Fang J-J, and Tsai M-J. Automatic body feature extraction from a maker-less scanned human body. *Computer-Aided Design*, 39: 568-582, 2007.
- [162] Werghi N. Segmentation and modeling of full human body shape from 3-D scan data: a survey. *IEEE Transactions on Systems, Man, and Cybernetics–Part C: Applications and Reviews*, 37(6): 1122-1136, 2007.
- [163] Zhong Y and Xu B. Automatic segmentation and measurement on scanned human body. *International Journal of Clothing Science and Technology*, 18(1): 19-30, 2006.
- [164] Shreiner D, et al. *OpenGL Programming Guide: The Official Guide to Learning OpenGL, version 2.1*. Upper Saddle River, NJ: Addison-Wesley, 2008.
- [165] Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, i: 307-310, 1986.
- [166] Han TS and Lean MEJ. Anthropometric indices of obesity and regional distribution of fat depots. In *International Textbook of Obesity*, ed. Bjorntrop, pp. 51-56. New York, NY: John Wiley & Sons, 2001.

Vita

Wurong Yu was born in Zhejiang, China on October 11, 1975, the son of Xutang Yu and Xiemei Yu. After completing his study at No.2 High School of Quzhou, Zhejiang, in 1992, he entered Shanghai Jiaotong University, Shanghai, China, where he received the B.S. degree in Biomedical Engineering in 1996 and the M.S. degree in Electrical Engineering in 1999. He researched algorithms for generating digital elevation models from aerial photography under the supervision of Prof. Yuanhua Zhou. In August 2000, he entered The Department of Biomedical Engineering of The Graduate School at The University of Texas at Austin. During the academic year of 2001-2002, he studied at Texas A&M University and worked on optical coherence tomography under the supervision of Dr. Lihong V. Wang. After he received the M.S. degree in Biomedical Engineering at The University of Texas at Austin in 2003, he continued to pursue the Ph.D. degree in Biomedical Engineering under the supervision of Dr. Bugao Xu. Throughout most of time at The University of Texas at Austin, he worked on 3D surface imaging techniques as a Research Assistant at Dr. Xu's Lab. His current research interests include image processing, computer vision and computer graphics.

Permanent Address: Xiazhou 196, Chihuai
Kaihua County, Zhejiang Province, 324305
P. R. China

This thesis was typed by the author.