

Development of 3D Human Body Measurement System
and Its Application to Biometric Gait Recognition

A THESIS SUBMITTED TO THE KEIO UNIVERSITY
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE GRADUATE SCHOOL OF SCIENCE AND TECHNOLOGY



March 2010

Koichiro Yamauchi
SCHOOL OF SCIENCE FOR OPEN AND ENVIRONMENT SYSTEMS

Abstract

This thesis presents development of 3D human body measurement system and its application to biometric gait recognition. A triangulation-based structured light system consisting of one camera and one projector obtains range data of a target object. In recent years, the measurement accuracy, speed, and physical size of the system have been dramatically improved, introducing a new class of products known as 3D camera. Unlike a typical camera which acquires 2D pixel data, a 3D camera captures range data as a large number of 3D points on an object's surface. The geometric model is defined such that the camera model is based on the pin-hole model, and the projector model is based on the equation of a plane model. Both camera and projector parameters are estimated by observing a planar object from three arbitrary viewpoints. One 3D camera cannot completely cover the entire body because of limited range and single view. In our system, four 3D cameras are installed in a pole as a measuring unit. The pole unit measures the whole body from head to toe and has a measurement range which is wider than a single 3D camera. We have built a human body measurement system consisting of three pole units for a total of twelve 3D cameras which can capture up to one million 3D points on the human body in only 2 seconds with 1.88 mm measurement accuracy. This sophisticated system is available for security applications, in particular for deriving the human gait from whole human body data. In our approach, we first measure range data of the entire body which consists of representative poses during a gait cycle and then extract gait features by fitting kinematic models to the data set. Unlike attempts which utilize silhouette images which are affected by clothing and self-occlusion, the positions of joints and body parts can be directly inferred. The experimental results show high identification rates in recognizing a human subject and his/her pose.

Acknowledgments

I wish to thank my supervisors, Hideo Saito and Yukio Sato, for their superb guidance. I could not have asked for better support, motivation, or enthusiasm, and working with them has been a great pleasure. I am immensely grateful to committee members, Masafumi Hagiwara, Hiroshi Shigeno, and Yoshimitsu Aoki, for accepting to evaluate this work. I must express my warmest gratitude to Shinji Ozawa and Jun Sato for giving me valuable comments and advices at conferences. I would also like to thank Bir Bhanu for giving me the opportunity of a visiting scholar at University of California, Riverside. Both Saito Laboratory and VISLab are filled with the most intellectually stimulating, helpful, and affable people that I have ever worked with. It would take a thesis in itself to thank individually everyone that has helped me through the course of my PhD. I am truly grateful to Mitsuru Ambai for providing me encouragement to overcome adverse challenges. I would also like to thank Takeharu Inukai and Hideo Narita for mutual trust and respect in our friendship. Last, my heartfelt thanks to my family for their encouragement, support, and love along the way.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Objective	2
1.3	Context	3
1.4	Contributions	5
1.5	Thesis Outline	6
2	About Range Imaging	9
2.1	Applications	9
2.2	Related Work	12
2.2.1	Sensing	13
2.2.2	Modeling	18
2.3	Thesis Position	21
3	Structured Light System	25
3.1	Introduction	25
3.2	Geometric Model	26
3.2.1	Camera Model	27
3.2.2	Projector Model	29
3.2.3	Triangulation Principle	30
3.3	Calibration Method	31
3.3.1	Camera Calibration	31
3.3.2	Projector Calibration	33
3.4	Experimental Results	35
3.4.1	Calibration	36

3.4.2	Evaluation	40
3.5	Discussion and Conclusion	41
4	3D Human Body Measurement	43
4.1	Introduction	43
4.2	Human Body Measurement System	45
4.2.1	Projector-Camera Pair	45
4.2.2	Pole Unit	47
4.2.3	System Configuration	47
4.2.4	Measurement Time	49
4.3	Calibration	51
4.3.1	Projector-Camera Pair	51
4.3.2	Coordinate System Alignment	51
4.4	Experimental Results	52
4.4.1	Calibration	53
4.4.2	Evaluation	55
4.4.3	Measurement	56
4.5	Discussion and Conclusion	57
5	Human Gait Recognition	61
5.1	Introduction	61
5.2	Related Work	63
5.3	3D Human Body Data	66
5.4	3D Human Body Model	67
5.5	Model Fitting	69
5.5.1	Body Axes	69
5.5.2	Torso	70
5.5.3	Arms and Legs	72
5.5.4	Head and Neck	72
5.6	Gait Reconstruction	73
5.7	Feature Matching	73
5.8	Experimental Results	74
5.8.1	Sensing and Modeling	75

5.8.2	Gait Reconstruction	76
5.8.3	Recognition	77
5.8.4	Comparisons	79
5.9	Discussion and Conclusion	80
6	Conclusion and Future Work	83
6.1	Summary of the Thesis	83
6.2	Future Work and Possible Extensions	85
	Bibliography	89

List of Figures

1.1	Framework for range imaging of a human body.	3
1.2	Range data of entire body.	5
2.1	Examples of structured light systems.	10
2.2	Examples of human body measurement systems.	11
2.3	Binary coded pattern and gray coded pattern.	13
2.4	De Bruijn pattern and M-array pattern.	15
2.5	Modeling results of CAESAR subjects.	18
2.6	Digital Michelangelo project.	19
2.7	Bayon digital archival project.	20
2.8	Classification of sensing and modeling technologies.	22
3.1	Geometric model of a structured light system.	27
3.2	Calibration scene.	31
3.3	Camera calibration.	32
3.4	Projector calibration.	34
3.5	Three sets of color images and stripe images.	36
3.6	Projector parameters.	37
3.7	Evaluation spheres.	38
3.8	Measurement results of five spheres.	39
4.1	Projector-camera pair.	45
4.2	8-bit gray coded patterns.	46
4.3	Drawing of pole unit.	48
4.4	Pole unit and control computer.	49
4.5	Measurement result of a man by one pole unit.	49

4.6	Human body measurement system.	50
4.7	Timing diagram.	50
4.8	Integration into world coordinate system.	51
4.9	Calibration cylinder.	53
4.10	Measurement result of the cylinder.	54
4.11	Horizontal cross-section.	54
4.12	Vertical cross-section.	54
4.13	Measurement result of a mannequin.	56
4.14	Measurement result of a man.	57
4.15	Measurement result of a clothed man.	58
5.1	Gait cycle expressed by swing phase and stance phase.	66
5.2	3D human body model.	67
5.3	Body axes estimation by PCA.	70
5.4	3D human body data of walking humans.	75
5.5	3D human body model fitted to four poses.	76
5.6	Gait sequence composed of twenty frames.	77
5.7	Examples of training data and testing data.	78
5.8	Virtual frontal view and side view.	80
6.1	UCR 3D gait database.	86

List of Tables

2.1	Comparisons of structured light patterns for range sensing.	16
3.1	Camera parameters.	36
3.2	Measurement accuracy of structured light system.	38
4.1	4-bit gray encoding and decoding.	47
4.2	Measurement accuracy of human body measurement system.	55
5.1	Comparisons of biometrics approaches for human recognition.	64
5.2	Bounding angles of rotational joints.	68
5.3	3D gait biometrics performance.	79
5.4	2D gait biometrics performance.	80

Chapter 1

Introduction

1.1 Motivation

The human body has fascinated scientists for thousands of years. Studying the shape of the human body offers opportunities to open up entirely new areas of research. The shape of the human body can be used to infer personal characteristics and features. Body type and muscle strength, for instance, can be used to distinguish gender. The presence or absence of wrinkles around the eyes and loose facial skin suggests a person's age. In addition, the size and shape of a person's face, belly, thighs, and arms can determine a body habitus: slim, healthy, or overweight. The length of individual limbs such as legs and their postural sway when a person walks suggests an underlying misalignment of the bone structure. It is in fact possible to identify people by their physical body shape of entire

body. Unlike traditional physical measures of height, width, and length of body parts, the body shape is represented by a closed surface of the entire human body as a 2-manifold. The surface is digitally captured and described by geometric primitives such as vertices, lines, and curves. It is useful for health professionals and technical experts to retrieve personal data from shared database whenever the need arises. Using a large number of body shape measurements, valuable personal characteristics can be statistically analyzed. If there is a strong correlation between body shape and medical disease, for instance, we can avoid invasive diagnostic procedures such as medical imaging methods which utilize electromagnetic radiation. Measurement of the shape of the human body also plays a leading role in medical diagnosis; it may even be possible to discover unexpected diseases through studying body shape. The differences among people of both sexes, young and old vary considerably in size and volume. In addition, the target is not only body parts such as face, finger, and ear, but also entire body including all of them. The focal issue involved in the use of entire body is how to exploit a vast amount of information. Therefore, it is challenging to know human mechanism from the appearance of human body and establish basic technologies for human understanding. A breakthrough in the fields of computer vision and pattern recognition will cause a paradigm shift well known as the structure of scientific revolutions.

1.2 Objective

This thesis provides new techniques for 3D human body measurement and recognition of walking humans. We address four sub-goals of the measurement and recognition: system calibration, range sensing, body modeling, and feature recognition. These tasks, illustrated in Fig. 1.1, are defined as follows:

System calibration aims to estimate intrinsic and extrinsic parameters representing the geometry of camera and projector and the relationships among the projector-camera pairs.

Range sensing addresses shape recovery from images of the camera and lights of the projector. Single or multiple projector-camera pairs are assigned around a target depending on the size of object.

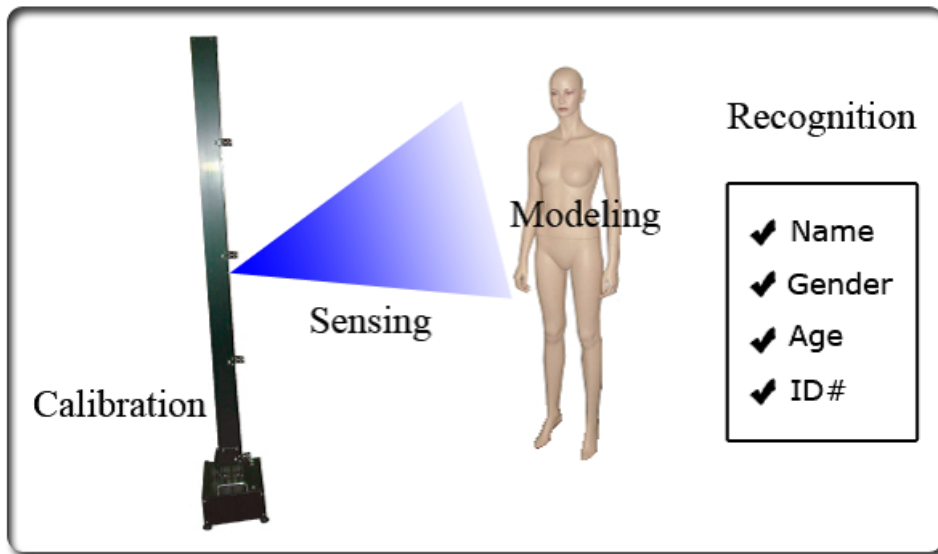


Figure 1.1: Framework for range imaging of a human body.

Body modeling aims to approximate body data by a simplified body model indicating the position, length, and volume of main parts of a human body used as physical features.

Feature recognition addresses identifying individuals through the process of feature matching to retrieve personal information such as name, gender, age, and ID number that is stored in the database.

The concept of four terms will be introduced throughout this thesis to form a framework which contains four-tiered structure. Here, the terms/tiers are numbered in order of calibration, sensing, modeling, and recognition from bottom to top. The upper tier task works properly when the lower tier tasks have been performed.

1.3 Context

Three-dimensional shape measurement is a technique for recovering the surface of a target object. Attempts have been successful in digitalizing the object represented by Cartesian coordinates in three dimensions. For shape reconstruction,

the combination of camera and projector is known as a minimum configuration. Source of light, i.e. projector, emits a plane of light which creates a narrow stripe on the scene. Since line of sight, i.e. camera, intersects with the light stripe, coordinates of all points along the stripe are obtained by triangulation principle. The geometry of projector-camera pair is represented by parameters corresponding to one side and two angles of a triangle. The parameters are estimated through calibration process which is designed to obtain accurate measurement results. The projection of light stripe offers several advantages in human body measurement. Especially, noncontact system does not take psychological and psychological pressures for a subject. Without any severe pose restrictions, the subject can change his/her pose and expression. Then, it is considered to finish the measurement as soon as possible to avoid the effects of body sways or shaky hand increasing every second. The measurement time is fast enough to cover for everyone from children to the elderly. In addition, the body surface is represented by a large number of points with high degree of accuracy. The points indicate not only anatomical landmarks but also as 2-manifold, i.e. closed surface. Furthermore, the light stripe has little adverse effect on the human body, in particular eyes. We can repeat the measurement as many times as we want with a sense of security.

Subsequently, three-dimensional image processing is a technique for analyzing the surface of a target object. Range data is comprised of range image whose pixels obtain coordinates value and color image whose pixels obtain color value. The resolution and accuracy depend on the number of pixels of the camera and the number of light stripes of the projector. For human body measurement, this type of system has targeted for small object such as face, finger, and ear. Recently, whole body which has large amount of information is subjected to study of sensing and modeling. Fig. 1.2 shows the range data of entire body for three subjects: mannequin, man, and clothed man. The data has approximately one million points, and also color information on every point. Of course, whole human body data includes all of body parts which have been individually captured by a range scanner. Biometrics system based on whole body comprises multimodal biometrics approaches and enhances the performance of personal identification. Especially, we focus on human gait consisting of combined physiological and behavioral features. The amount information of range data which includes depth

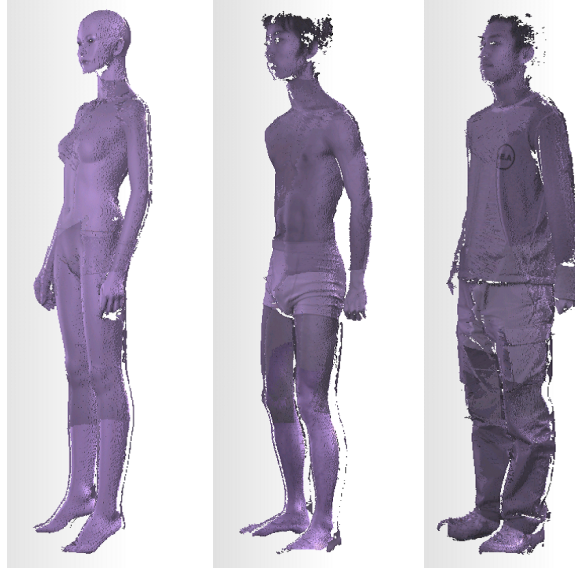


Figure 1.2: Range data of entire body.

and color, i.e. three-dimensional data, is not half as much again as that of video sequences which include only color, i.e. two-dimensional data. Intuitively, the video sequences only grab a piece of information, but range data acquires all of them, in other words, we can observe a target object from arbitrary viewpoints in three dimensions. Although it has been challenging to advance from the research and development stage, existing devices and applications do not become widely used. The seeds of technology need to meet the needs of consumers by developing a new market and exploring the capability. We have to distinguish what one really needs from what one thinks one needs. This thesis attempts to tackle range imaging technology of measuring and processing.

1.4 Contributions

The primary contributions of this thesis are threefold:

- Structured light system is defined that the camera is based on the pinhole model and the projector is based on the equation of a plane model. The parameters are estimated by observing a planar object from three viewpoints.

The geometric model provides highly accurate range data and the calibration method facilitates the procedure of user's task.

- Human body measurement system captures range data of entire body within a couple of seconds with high resolution and high accuracy. The pole unit consists of four projector-camera pair to capture range data from head to toe. Whole human body data is acquired by two or more independent pole units placed around a subject.
- Human gait recognition is one of the most recent biometrics approaches for personal identification by the manner of walking. 3D gait biometrics captured by the human body measurement system provides highly reliable authentication unlike 2D gait biometrics approaches and motion capture methods using a single or multi-camera system.

1.5 Thesis Outline

The body of this thesis is divided into six chapters, the first of which is this introduction.

Chapter2. This chapter investigates range imaging commercial products, representative applications, coded structured light projections, and sensing and modeling technologies.

Chapter3. This chapter describes geometric model and calibration method for a structured light system which consists of one camera and one projector through a planar object.

Chapter4. This chapter introduces compact, high speed, high resolution, and high accuracy human body measurement system using multiple projector-camera pairs.

Chapter5. This chapter presents 3D gait biometrics recognition based on interpolation of continuous motion using whole human body data of representative poses during walking.

Finally, we conclude with a summary of the thesis and some suggestions of future work.

Chapter 2

About Range Imaging

2.1 Applications

A triangulation-based structured light system consists of a camera and a projector. The system is similar to passive stereo vision system whose camera is replaced by the projector. Range data encapsulating range image and color image is obtained by the camera observing a target object illuminated from the projector. The commercialization of research achievements in range imaging has already been realized. Fig. 2.1 is the examples of structured light systems. VIVID910 of KONICA MINOLTA captures range data in 0.3 seconds with 8, 14, 25 mm focal length and 640×480 pixels. The structured light based on light section method is made by galvanometer mirror. This system provides 3D CAD software to edit and process the data. Cartesia 3D Handy Scanner of SPACEVISION captures range data



KONICA MINOLTA
VIVID910



SPACEVISION
Cartesia

Figure 2.1: Examples of structured light systems. Courtesy of KONICA MINOLTA and SPACEVISION.

in 0.5 seconds with 8 mm focal length and 640×480 pixels. The structured light based on space encoding method is made by polygon mirror. This system is slim and compact using one USB 2.0 cable instead of any AC power cables and transmission cables.

While structured light system is designed for general use, some systems specialize in capturing range data of an entire body. Fig. 2.2 is the examples of human body measurement systems. Model WBX of Cyberware captures whole human body data in 17 seconds with 1.2 mm horizontal, 2.0 mm vertical, and 0.5 mm depth resolution. The data is represented by X, Y, and Z coordinates for shape and RGB values for color. Bodyline Scanner of Hamamatsu Photonics completes the measurement in 5 seconds using four scanning heads based on light section method. This system takes subject's size by detecting landmarks automatically. They are applicable to offer custom-made clothes. VITUS Smart XXL of HUMAN SOLUTIONS proposes to integrate avatars with real body dimensions and attributes into design programs. This system is part of the program incorporating three separate stages: measurement, simulation, and integration. Cartesia 3D Body Scanner of SPACEVISION captures whole human body data including one million points in 3.6 seconds using three pole units based on space encoding method. This system is not divided by a wall or curtain, and the installation space is reduced by independent measuring units. Here, the pictures in Fig. 2.1 and Fig. 2.2, which are the copyrights of Cyberware, Hamamatsu Photonics, HUMAN SOLUTIONS, KONICA MINOLTA, and SPACEVISION, are permitted to



Cyberware
Model WBX



Hamamatsu Photonics
Bodyline Scanner



HUMAN SOLUTIONS
VITUS Smart XXL



SPACEVISION
Cartesia

Figure 2.2: Examples of human body measurement systems. Courtesy of Cyberware, Hamamatsu Photonics, HUMAN SOLUTIONS, and SPACEVISION.

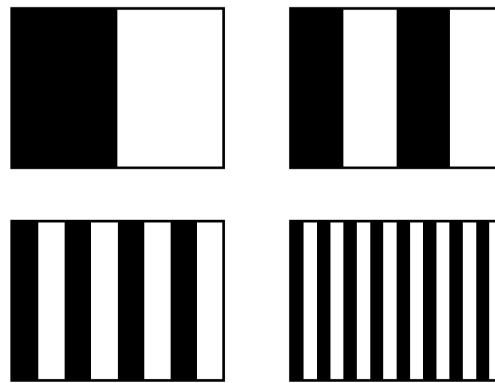
use for this thesis.

Last decade, range imaging has proved as an emerging technology for a wide variety of applications. In the medical field [1, 2], X-ray [3], i.e. radiography, CT [4], i.e. computerized tomography, MRI [5], i.e. magnetic resonance imaging, represent volume data of a subject and visualize body shape. Although internal organs, vessels, and bones clarify health condition, some parts of that are diagnosed by surface shape. For example, BMI [6], i.e. body mass index, which compares subjects' weight and height, provides a healthy body weight depending on the body height. Range data from the neck to the waist derives such information. In addition, the differences before and after surgery and body changes while dieting are recorded. In the fashion field [7, 8], fashion designers measure body size with

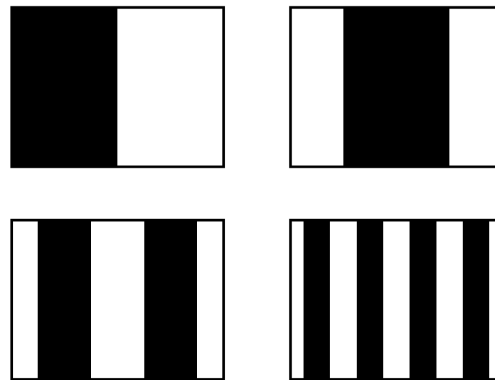
tools, such as a tape measure and a ruler. Their rich experiences through basic practices also support to custom-made clothes. Range data is useful to remove this laborious task and measure the size at different places without craftsmanship. For example, some human body measurement systems obtain landmarks of a human body, and then compute body size automatically. In other cases, the systems employ a semi-automatic approach using attached markers. After that, we virtually try on a lot of different clothes changing the size, color, and material. In the graphics field [9, 10], aesthetically pleasing virtual mannequins are based on a limited number of user-specific criteria such as a waist circumference, chest, and weight. For example, whole human body data are grouped into a user-specified number of clusters, which are formed through the clustering of the anthropometric database. The body data whose shape is closest to the virtual mannequin is retrieved and subsequently used to adjust the virtual mannequin. In contrast, some approaches employ simplified virtual models, e.g., stick figure model, cylinder model, and blob model. These models are directly fitted to whole human body data with/without a *priori* knowledge. In the security field [11, 12, 13], biometrics systems generally use color images or video sequences, and extract features for human recognition. Although they are successfully gaining and are available for security applications, range data holds enormous potential for the improvement of recognition rate and robustness. Biometrics approaches using range data of face, finger, and ear have proven the effectiveness and reliability. In the near future, whole human body data will be used for biometrics system.

2.2 Related Work

Related work is divided into two parts. In the first part, sensing technology covers representative structured light patterns: light section method, binary coded pattern, gray coded pattern, De Bruijn pattern, and M-array pattern. In the second part, modeling technology targets humans, sculptures, statues, and sites from small object to large-scale environment.



(a) Binary coded pattern



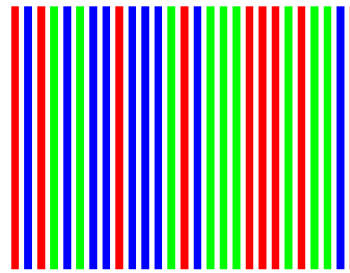
(b) Gray coded pattern

Figure 2.3: Binary coded pattern and gray coded pattern.

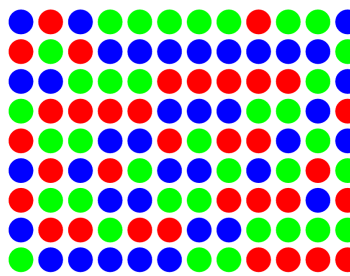
2.2.1 Sensing

Coded structured light projection is one of the highly reliable techniques for the acquisition of range data. The projector emits a structured light to a target object and encodes the scene captured by the camera. The pixels on the image are decoded by either zero or one. The range data is computed by triangulation principle among the camera, projector, and target object. This type of device is called as structured light system. In 1971, Shirai [14] proposed the light section method which scans a light stripe from one side to the other. The projected light stripe is approximated by line segments, the slope and distance of which are used for object recognition. Y. Sato *et al.* [15] proposed a method for the acquisition of complete

range data using a turntable. If both camera and projector observe a target object from one direction, they cannot observe the sides and the back of the object. Using the controlled turntable, one direction view is changed to omnidirectional view. Araki *et al.* [16] proposed a method for high speed range imaging within a few milliseconds. The integrated circuit consists of the optical part, control part, and signal processing part. This system performs to scan a light stripe and compute the position simultaneously. Kanade *et al.* [17] proposed the system based on VLSI sensor array. Unlike most systems which perform step-and repeat process of the camera and projector, the sensor includes both photosensitive and signal processing circuits. This system achieves range data up to one thousand frames per second. In stead of a single light stripe, several coding strategies have been studied so far. Fig. 2.3 shows the binary coded pattern and the gray coded pattern. In 1982, Posdamer and Altschuler [18] presented the binary coded pattern. The projection of p patterns are the same as the projection of 2^p light stripes. When we use the eight patterns, the number of scanning times is dramatically reduced from 256 to 8. Altschuler *et al.* [19] presented the system using dot matrix of binary light beams. Each column of the pattern is independently controlled to be either lighted or obscured. It allows coding any pattern dot as a sequential projection of different patterns. In 1984, Inokuchi *et al.* [20] presented the gray coded pattern. The advantage of this pattern is robust to noise or encoding/decoding error, because the consecutive codeword has the Hamming distance of one. The gray coded pattern is adapted for practical use and used for the following systems. K. Sato and Inokuchi [21] presented the system using an electrically controlled liquid crystal device. The pattern is generated with speed and precision compared to a slide projector. To improve the decoding error, reference image is used to define dynamic thresholds for all of pixels. Y. Sato and Otsuki [22] presented the system, called Cubicscope, using a semiconductor laser and a synchronized scanned mirror. The pattern is generated by scanning and switching of one light stripe. This system provides high power illumination with low heat radiation, and also encourages the reduction of size and weight. Gunaratne *et al.* [23] developed face measurement system, based on Cubicscope, consisting of two cameras and one projector. While other systems consisting of one camera and one projector fail to capture nose and face line, the system successfully acquires dense range



(a) De Bruijn pattern



(b) M-array pattern

Figure 2.4: De Bruijn pattern and M-array pattern.

data of them. Hasegawa *et al.* [24] developed electronic endoscope system for computer-aided diagnosis. This system captures range data of organ by emitting the pattern from the endoscope tip through flexible tube. Appearance of the affected part is represented by both shape and color. Aoki *et al.* [25] developed tooth measurement system for the visualization of jaw movement. Range data of upper and lower tooth are separately acquired, and then the positions of the data are brought into occlusion using attached markers. Kouchi and Mochimaru [26] developed foot measurement system for designing shoes. The foot scanner, which is the product of INFOOT, is consisted of four projectors and eight cameras. It is capable of acquiring range data of foot in 10 seconds at 1.0 mm intervals. Anatomical landmarks are measured by using attached markers. As stated above, it has been preferable to use light stripes to recover shape of an object with a high degree of accuracy. Most systems which utilize structured light achieve dense data compared to other type of range sensor.

Subsequently, pseudorandom sequence and array are used to encode patterns

Table 2.1: Comparisons of structured light patterns for range sensing.

Technique	Accuracy	Speed	Authors
Light section method	* *	*	[14, 15, 16, 17, 26]
Binary coded pattern	* *	* *	[18, 19]
Gray coded pattern	* * *	* *	[20, 21, 23, 24, 25]
De Bruijn pattern	*	* * *	[27, 28, 29, 30]
M-array pattern	*	* * *	[31, 32, 33, 34]

for acquiring range data at once. Fig. 2.4 shows the De Bruijn pattern and the M-array pattern. First, De Bruijn pattern is based on cyclic string to identify a fixed length substring over the string. The string is constructed by searching Hamiltonian path or Eulerian path over De Bruijn graph. Boyer and Kak [27] proposed three colored pattern consisting of vertical lines. This pattern is generated by a sequence of vertical lines which are colored with any combination of red, green, and blue. Their specific hardware obtains the index and position of three colors peaks in real time. Vuylsteke and Oosterlinck [28] proposed the monochrome pattern using four primitives. This pattern structure is a checkerboard where grid points are encoded by either bright or dark spot. The sequence which forms two rows is compacted and insensitive to surface discontinuities. Salvi *et al.* [29] proposed the six colored pattern consisting of vertical and horizontal lines. The vertical lines are represented by magenta, cyan, and yellow, and the horizontal lines are represented by red, green, and blue. The grid points are easily segmented, and also neighbors are detected by tracking along the edges. L. Zhang *et al.* [30] proposed the eight colored pattern consisting vertical lines. The correspondence between points in the pattern and pixels in the image is resolved by multi-pass dynamic programming. To increase the resolution, the pattern is shifted to the right in time and located the lines with sub-pixel accuracy. Second, M-array pattern is based on unique matrix to identify a fixed size submatrix over the matrix. Morita *et al.* [31] presented the two monochrome patterns using black dots. The first pattern contains all of black dots to determine their positions and the second pattern contains M-array representation. Although this method is restricted to static scenes, it is applicable to dynamic ones by only emitting the second pattern. Griffin *et al.* [32]

presented the monochrome pattern using five primitives. This pattern matrix is a special case of perfect maps which is made by a systematic process. Each element has a unique codeword formed by its value and the values of four neighbors. Morano *et al.* [33] presented the three colored pattern consisting red, green, and blue dot array. This pattern growth of the matrix consists of a random assignment of colors from the available palette for undefined elements adjoining the established pattern. Kimura *et al.* [34] presented foot measurement system for motion analysis during walking. The three colored pattern is represented by pixels and cells assigned red, green, and blue. Since foot has smooth surface with uniform reflectance, it is suitable as a target for the measurement.

Obviously, these systems which utilize colors and primitives sometimes fail to decode the projected pattern on the surface of an object especially for colorful and complex shapes, although they can obtain range data in real time. Table 2.1 compares five kinds of structured light patterns presented in this chapter. Here, the number of symbol indicates three-star scale. The binary coded pattern and the gray coded pattern provide high degree of accuracy because of the projection of a large number of light stripes. Especially, the gray coded pattern whose light stripe has the Hamming distance of one is robust to decoding error. The light section method also provides high degree of accuracy, but background subtraction which detects scanning light stripe is required. The De Bruijn pattern and the M-array pattern depend on decoding process to extract colors and primitives of the pattern. Their use is limited to neutral color scene or simple object, so that the measurement accuracy is not stable enough. In contrast, the light section method which scans a light stripe from one side to other takes a long time. The binary coded pattern and gray coded pattern which usually emit eight or more patterns are slightly better than the light section method. The De Bruijn pattern and the M-array pattern have successfully captured range data in real time so far. Definitely, we would argue for accepting the gray coded pattern. The first priority for range imaging is to obtain range data with a high degree of accuracy. We intend to derive valuable information about both inside and outside of the object from the surface shape, and also develop them for various purposes including anthropometric and archaeological applications. The accurate range data leads to system performance enhancement and significantly boosts the reliability of its



Figure 2.5: Modeling results of CAESAR subjects. Courtesy of Brett Allen [38].

results. The second priority is measurement time. Although the light section method is slower than De Bruijn pattern and M-array pattern, this problem has already been addressed by the integrated circuit implementation for sensing the scene and computing the position. The gray coded pattern is overcoming through increasing the shutter speed of the camera, projection speed of the projector, and switching speed between the camera and the projector.

2.2.2 Modeling

The anthropometry has been emerging as leading-edge applications for range imaging. Especially, human body attracts an awful lot of attention at the moment. The dedicated device which captures range data of entire body is known as human body measurement system. CAESAR project [35, 36] is a survey of whole body measurements for people ages 18-65 in three countries: United States, Netherlands, and Italy. Five types of data are recorded: demographic information, 40 measurements taken with a tape measure, 60 measurements captured by a body scanner, complete range data of three postures, and 73 coordinates of specified landmarks. The total number of samples is approximately fifteen thousand. For data collection, two scanners which are built by Cyberware and Vitronic are used. Both scanners use similar measurement principle based on light section method. The projector emits a light stripe to the surface of a subject and moves from the top to the bottom. Then, the camera observes the projected light stripe from two directions at the same time. Whole human body data is recovered by triangulation principle. AIST/HQL database [37] contains a total of 97 samples including

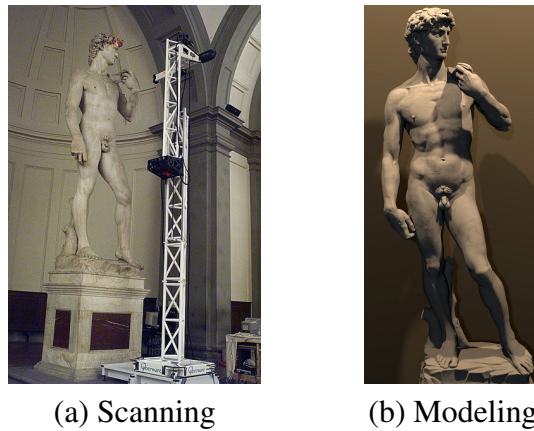


Figure 2.6: Digital Michelangelo project. Courtesy of Marc Levoy [40, 41].

49 men in the 20-30 age group and 48 women in the 20-35 age group in Japan. Five types of data are recorded: demographic information, 91 anthropometrical dimensions, 21 landmarks, 3D data, and 3D model. The demographic information is measurement date, birth date, age, and sex. The dimensions and landmarks are measured by two experts manually. Body data is collected by the commercial product of Hamamatsu Photonics. This system based on light section method captures body data in 11 seconds within ± 0.5 percent measurement accuracy. After that, body model is represented by not only vertices and polygons but also 26 positions of cross-section. The database is used for statistical analysis and open to the public on the website. However, starting from a range scan, we need to process the noisy and incomplete surface into a complete model suitable for applications. Further, the scanned data has holes caused by self-occlusions and grazing angle views. Allen *et al.* [38] proposed a method for fitting high resolution template meshes to detailed human body range scans with sparse markers. Fig. 2.5 is the modeling results of CAESAR subjects. Affine transformation at each template vertex is formulated as an optimization problem. The object function trades off fit to the range data, fit to scattered markers, and smoothness of the transformations over the surface. Anguelov *et al.* [39] introduced SCAPE method for building a human shape model which incorporates both articulated and non-rigid deformations. The SCAPE model is used for shape completion, partial view completion, and motion capture animation using just a single static scan and a marker mo-

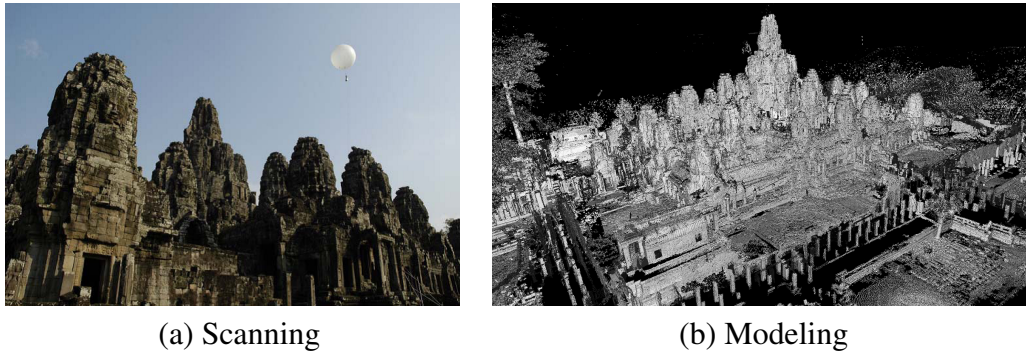


Figure 2.7: Bayon digital archival project. Courtesy of Katsushi Ikeuchi [44, 45].

tion capture sequence of the person. This method generates realistic meshes with muscle deformation for a wide range of subjects and their poses.

Large-scale environment in the archaeology field has been selectively targeted by improving the sensing and modeling technologies. Digital Michelangelo project [40, 41] is to involve in digitizing the sculptures and architectures of Michelangelo. Fig. 2.6 is the scanning scene and the modeling result. The statue of David is captured by a structured light system mounted on a motorized gantry. Since, the focus of this project is to protect both physical shape and geometric representation such as vertex coordinates, surface normals, and connectivity information, the system acquires a total of two billion polygons and seven thousand color images for three weeks. After scanning, pipeline processing aligns range data taken from different positions, and then the range data are combined into a unified surface mesh filling any holes automatically. Finally, the model containing eight million polygons is rendered. While it is necessary to measure human body at high speed, the resolution and accuracy are as important as the measurement time. This way of thinking leads to trade-off problem. Great Buddha project [42, 43] focuses on the preservation and restoration of Asuka, Kamakura, and Nara Buddha. The framework of geometric modeling incorporates three separate steps: acquisition, alignment, and merging. The first step is to acquire range data by laser range sensors. The sensors capture a set of partial mesh models, overlapping each other and covering the entire object surface. The second step is to align partial mesh models. Each sensor is located at an arbitrary position on data acquisition, so that relative

relations of these models are determined by considering resemblances in the data set. The third step is to merge the aligned multiple mesh into a complete mesh model. The entire object is represented by one surface from multiple overlapping surface observations. Asuka, Kamakura, and Nara Buddha contain three million meshes, ten million meshes, and seventy million meshes, respectively. Bayon digital archival project [44, 45] is to reconstruct the Bayon temple which is located at the center of Angkor-Thom. This is a huge structure, i.e. more than 150 meters long on all sides and up to 45 meters high, including 51 towers, 173 calm, smiling faces carved on the towers, and double corridors carved in beautiful and vivid bas-relief. Fig. 2.7 is the scanning scene and modeling result. Flying laser range sensor which is suspended beneath a balloon measures structures invisible from the ground. The obtained data has some distortion due to the movement of the sensor during the scanning process, so that alignment algorithm estimates not only rotation and translation but also motion parameters. The resulting 3D model consists of twenty thousand range images and its total size is about two hundred gigabytes.

2.3 Thesis Position

To clarify the position of this thesis in range imaging, sensing and modeling technologies including several levels are arranged in a treelike structure, as illustrated in Fig. 2.8. First, sensing technology is classified according to passive and active range sensing. In passive sensing, shape recovery techniques are called as shape from X such as stereo, motion, shading, texture, focus. Using single or multi-camera, they achieve range data in real time with low resolution and low accuracy. In active sensing, representative techniques are triangulation, time of flight, photometric stereo, and moiré topography. Combining camera and projector, they achieve range data in no real time with high resolution and high accuracy. Although there are a variety of techniques, the first and foremost priority is to obtain high resolution, highly accurate range data. In Chapter 3 and Chapter 4, we employ active range sensing by triangulation principle. Especially, gray coded pattern is the most reliable structured light pattern considering not only the resolution and accuracy but also the speed and robustness. Second, modeling technology

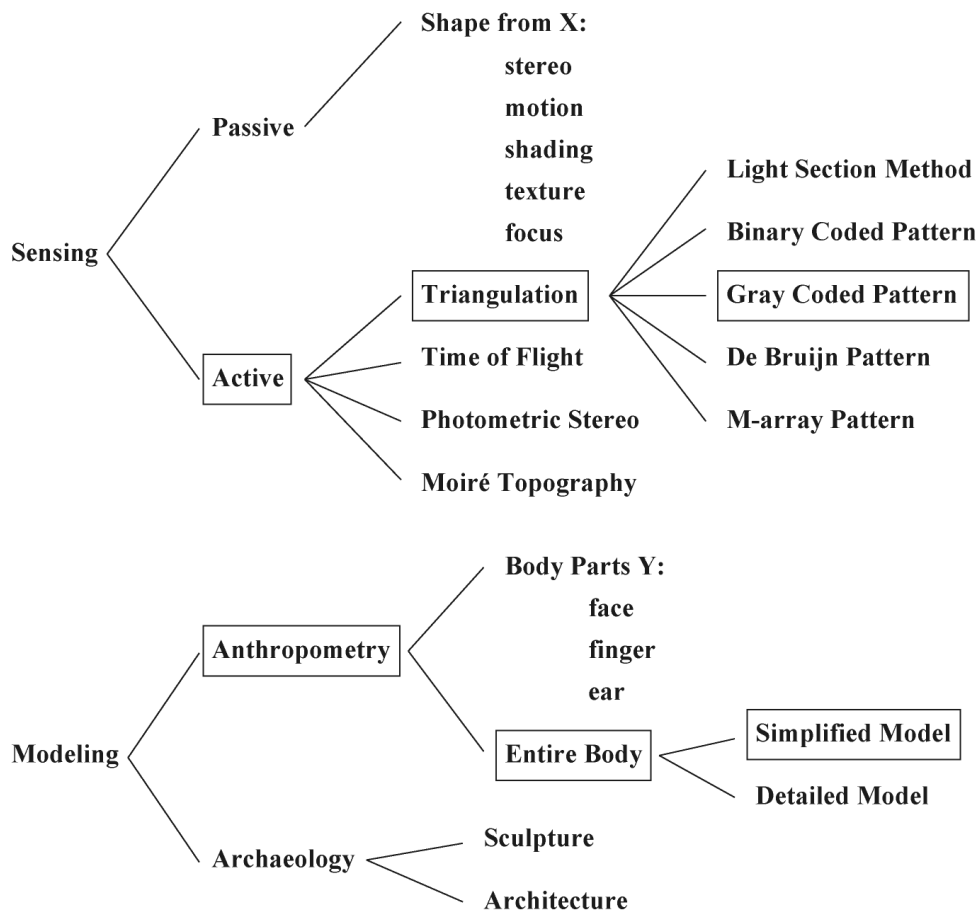


Figure 2.8: Classification of sensing and modeling technologies.

is classified into two categories: anthropometry and archaeology. In the anthropometry, the target is not only body parts Y such as face, finger, and ear, but also entire body including all of them. The measurement of the human body is useful for human understanding. In the archaeology, the target objects are sculpture and architecture. It aims to protect cultural assets which have been deteriorated over the centuries by digitalizing the object's surface. Although biometrics systems have used face, finger, ear, and their multimodal data, nobody tries security application using whole human body data. In Chapter 5, we focus on anthropometry covering entire body. Unlike detailed model which collect physical measures and anatomical landmarks, simplified model is used to extract the position, length, and

volume of main parts of a human body. In this thesis, these sensing and modeling technologies are related to development of 3D human body measurement and its application to biometric gait recognition.

Chapter 3

Structured Light System

3.1 Introduction

Pair of one camera and one projector, i.e. structured light system, is minimum configuration for range imaging. The geometry of the structured light system is approximated by either simple model or complex model. The simple model represents the system by simplifying the geometry with the small number of parameters. In contrast, the complex model represents the system by complicating the geometry with the large number of parameters. Naturally, the geometric model is designed for high accuracy measurement against the loss of versatility.

Typical geometric model encapsulates two separate models: camera model and projector model [46]. Generally, the pinhole model is the most commonly used representation [47, 48, 49]. The camera geometry is represented by a 3×4

matrix having 11 degrees of freedom and the projector geometry is represented by a 2×4 matrix having 7 degrees of freedom. The two projection matrices provide range data by the principle of binocular stereo. Although the pinhole model is suited for the camera geometry, it is not applicable to the projector geometry. For example, light stripes do not always pass through the optical center of the projector using a rotatable mirror, e.g., galvanometer mirror and polygon mirror.

Subsequently, the triangulation principle based on the baseline is also utilized. Given one side and two angles of a triangle determine the position of a target object. One side is the baseline which is defined as the distance between the camera and the projector. One of the angles indicates camera view and the other indicates projector view. The invariable baseline model [50, 51, 52] fails to represent some projectors using a rotatable mirror, but the variable baseline model [53, 54, 55] eases this problem. However, these models assume that all of light stripes are vertical to the baseline. It is preferable to express the light stripe by a three-dimensional plane disregarding the inner structure of the projector.

In this chapter, the problem of geometric model for a structured light system is addressed. The geometric model is defined such that the camera model is based on the pinhole model and the projector model is based on the equation of a plane model. If light stripes are projected in different directions, their projections are expressed accurately. Furthermore, the camera and projector parameters are estimated by observing a planar object from three viewpoints. Unlike other approaches using cube objects [56, 57], it facilitates the procedure of user's tasks and provides a high degree of accuracy.

3.2 Geometric Model

A structured light system consists of a camera and a projector. The system captures a range data by the camera observing a target object illuminated from the projector. Fig. 3.1 is the geometric model of a structured light system. The camera model is based on the pinhole model and the projector model is based on the equation of a plane model. The geometric model is represented in the camera coordinate system and the reference plane is represented in the reference plane coordinate system.

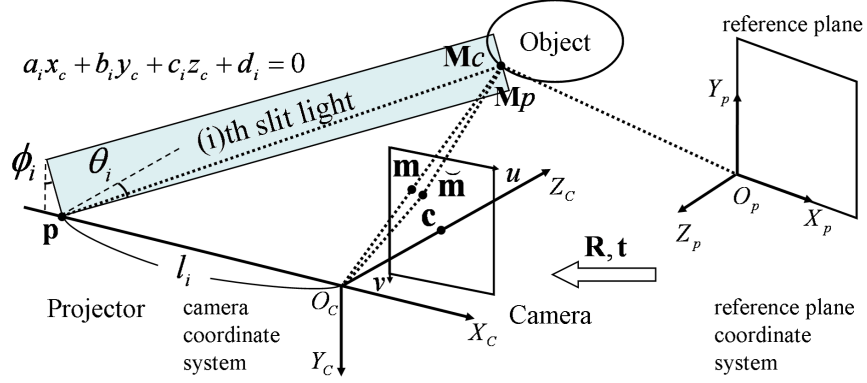


Figure 3.1: Geometric model of a structured light system.

3.2.1 Camera Model

Pinhole model is defined that light rays from an object pass through the optical center O_c for imaging. The principal point \mathbf{c} at the intersection of the optical axis with the image plane is denoted by $[u_0, v_0]$. X_c -axis, Y_c -axis, and Z_c -axis are parallel to horizontal axis, vertical axis, and optical axis of the image plane. Here, a 2D point, i.e. image coordinates, \mathbf{m} is denoted by $[u, v]$ in the image plane, and a 3D point, camera coordinates, \mathbf{M}_c is denoted by $[x_c, y_c, z_c]$ in the camera coordinate system (O_c - X_c - Y_c - Z_c). In addition, X_p -axis, Y_p -axis, Z_p -axis, and O_p are defined as horizontal axis, vertical axis, orthogonal axis, and the coordinate origin of the reference plane. Here, a 3D point, i.e. reference plane coordinates, \mathbf{M}_p is denoted by $[x_p, y_p, z_p]$ in the reference plane coordinate system (O_p - X_p - Y_p - Z_p). The perspective projection which maps the reference plane coordinates onto the image coordinates is given by

$$\tilde{\mathbf{m}} \simeq \mathbf{A} \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \tilde{\mathbf{M}}_p, \quad (3.1)$$

$$\text{with } \mathbf{A} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix},$$

where \mathbf{A} is the camera intrinsic matrix with the scale factors, α, β, γ , and the principal point, u_0, v_0 , i.e. the intrinsic parameters, and $\begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix}$ combines the rotation

matrix and the translation vector, i.e. the extrinsic parameters. The tilde indicates the homogeneous coordinate by adding 1 for the additional element: $\tilde{\mathbf{m}} = [u, v, 1]$ and $\tilde{\mathbf{M}}_p = [x_p, y_p, z_p, 1]$. The Euclidean transformation which transforms the reference plane coordinates to the camera coordinates is given by

$$\mathbf{M}_c = \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \tilde{\mathbf{M}}_p, \quad (3.2)$$

with $\mathbf{R} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix}$,

where $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ correspond to unit vectors to indicate the directions of X_p -axis, Y_p -axis, Z_p -axis, respectively. \mathbf{t} is the direction vector from O_p to O_c . Therefore, camera parameters provide the perspective projection and the Euclidian transformation.

Let us consider camera lens distortion and its removal. The radial distortion causes the inward or outward displacement of the image coordinates from their ideal locations. This type of distortion is mainly caused by flawed radial curvature curve of the lens elements [58, 59]. Here, a distorted 2D point, i.e. real image coordinates, $\check{\mathbf{m}}$ is denoted by $[\check{u}, \check{v}]$. The discrepancy between the ideal image coordinates and the real image coordinates considering first two terms of radial distortion is given by

$$\check{\mathbf{m}} = \mathbf{m} + [k_1 \hat{\mathbf{m}}^T \hat{\mathbf{m}} + k_2 (\hat{\mathbf{m}}^T \hat{\mathbf{m}})^2] (\mathbf{m} - \mathbf{c}), \quad (3.3)$$

where k_1 and k_2 are the coefficients of the radial distortion, the center of which is the principal point. Here, a normalized 2D point, i.e. normalized image coordinates, $\hat{\mathbf{m}}$ is denoted by $[\hat{u}, \hat{v}]$, the focal length of which is defined by 1 as seen in [60] is given by

$$\hat{\mathbf{m}} = \mathbf{A}^{-1} \mathbf{m}. \quad (3.4)$$

Therefore, camera lens distortion can be corrected from captured images. For more detail on camera geometry, refer to computer vision literatures [61, 62].

3.2.2 Projector Model

The projector emits one to hundreds of light stripes for the measurement. We consider the case in which the light stripes are projected in different directions. It is difficult to assume that the projector model is based on the pinhole model, because they do not pass through the optical center. Therefore, we use the equation of a plane model to accurately represent the projector instead of considering the projection of the light stripes which depend on the type of projector. In the camera coordinate system, the light stripe can be written as

$$a_i x_c + b_i y_c + c_i z_c + d_i = 0, \quad (3.5)$$

where i is the light stripe number, and a_i, b_i, c_i, d_i are the coefficients of the equation. There are an equal number of the equations of planes and the light stripes. Intuitively, the intersection of X_c -axis and the light stripe, \mathbf{p} , moves from the left to the right on the axis.

We define l_i is the baseline, i.e. the distance between the optical center of the camera and the light stripe of the projector, θ_i is the projection angle, i.e. the angle between Z_c -axis and the light stripe, and ϕ_i is the tilt angle, i.e. the angle between Y_c -axis and the light stripe. From the coefficients of the equation, these explicit parameters can be written as

$$l_i = d_i/a_i, \quad (3.6)$$

$$\theta_i = \arctan(-c_i/a_i), \quad (3.7)$$

$$\phi_i = \arctan(-b_i/a_i). \quad (3.8)$$

Projector parameters are expressed by both implicit and explicit representations. The coefficients are used for computation of range data, but their values do not exhibit distinct features. In contrast, the baselines, projection angles, and tilt angles provide characteristic distributions.

3.2.3 Triangulation Principle

To achieve range data, the projector emits light stripes to a target object, and then the camera observes the illuminated object. So, the camera coordinates is the intersection of the viewpoint of the camera and the equation of a plane of the projector. The linear equation $[x_c/z_c, y_c/z_c, 1/z_c]$ which is derived from (3.1), (3.2), and (3.6) is given by

$$\begin{bmatrix} \alpha & \gamma & 0 \\ 0 & \beta & 0 \\ a_i & b_i & d_i \end{bmatrix} \begin{bmatrix} x_c/z_c \\ y_c/z_c \\ 1/z_c \end{bmatrix} = \begin{bmatrix} u - u_0 \\ v - v_0 \\ -c_i \end{bmatrix}. \quad (3.9)$$

Consequently, we have

$$x_c = \frac{(u - u_0) - \frac{\gamma}{\beta}(v - v_0)}{\alpha} z_c, \quad (3.10)$$

$$y_c = \frac{v - v_0}{\beta} z_c, \quad (3.11)$$

$$z_c = \frac{\frac{d_i}{a_i}}{-\frac{c_i}{a_i} - \frac{(u - u_0) - \frac{\gamma}{\beta}(v - v_0)}{\alpha} - \frac{b_i}{a_i} \frac{(v - v_0)}{\beta}}. \quad (3.12)$$

The coordinate z_c is computed by the relationship between the viewpoint of the camera and the equation of a plane of the projector. Then, the coordinate x_c and the coordinate y_c are computed by the similar triangle related to the camera. Therefore, the camera coordinates can be recovered by the camera and projector parameters. The coordinate z_c which is expressed by the baseline, projection angle, and tilt angle instead of the coefficients can be written as

$$z_c = \frac{l_i}{\tan \theta_i - \frac{(u - u_0) - \frac{\gamma}{\beta}(v - v_0)}{\alpha} + \tan \phi_i \frac{(v - v_0)}{\beta}}. \quad (3.13)$$

It indicates the triangulation principle based on one side and two angles of a triangle.

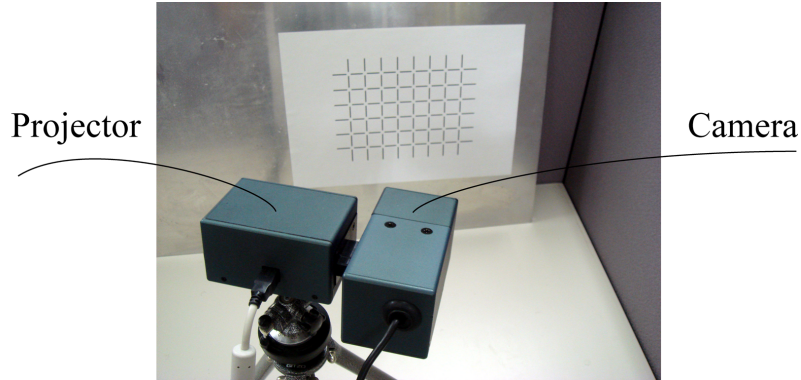


Figure 3.2: Calibration scene.

3.3 Calibration Method

In this section, we present a calibration method for a structured light system by observing a planar object from three viewpoints. Fig. 3.2 is the calibration scene of a structure light system. The planar object, called reference plane, contains a checkered pattern, so that calibration points are detected as the intersection of line segments. To perform the calibration, the reference plane coordinates is assigned to the calibration points. Three sets of color images and stripe images, which capture calibration points and light stripes on the reference planes respectively, are required. Our approach incorporates two separate stages: camera calibration and projector calibration.

3.3.1 Camera Calibration

In the camera calibration stage, camera parameters are obtained by Zhang's method [63]. Fig. 3.3 shows the relationship between the reference plane and the image plane. The camera parameters are estimated by the correspondence between the reference plane coordinates and the image coordinates. Note that three color images must be captured from different positions changing orientations. If the reference plane undergoes pure translation, the camera parameters cannot be estimated [64]. Here, the reference plane is on $z_p = 0$ and the homography matrix \mathbf{H}_p is denoted by $[\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3]$. From (3.1), the perspective projection which maps

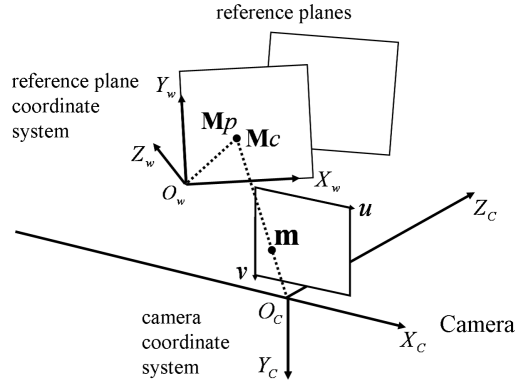


Figure 3.3: Camera calibration.

the reference plane coordinates onto the image coordinates can be written as

$$\tilde{\mathbf{m}} = \mathbf{H}_p \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix}, \quad (3.14)$$

$$\text{with } \mathbf{H}_p = \mathbf{A} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}.$$

Consequently, we have

$$\begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} = \lambda \mathbf{A} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}, \quad (3.15)$$

where λ is the constant parameter. The homography matrix has eight degrees of freedom which encapsulates five degrees of freedoms of intrinsic parameters and three degrees of freedoms of extrinsic parameters. Then, the symmetric matrix \mathbf{B} can be written as

$$\mathbf{B} = \kappa \mathbf{A}^{-T} \mathbf{A}, \quad (3.16)$$

$$\text{with } \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{12} & b_{22} & b_{23} \\ b_{13} & b_{23} & b_{33} \end{bmatrix},$$

where κ is the constant parameter and $(-T)$ indicates the transpose of an inverse matrix. Thus, the intrinsic parameters are given by

$$u_0 = \gamma v_0 / \alpha - b_{13} \alpha^2 / \kappa, \quad (3.17)$$

$$v_0 = (b_{12} b_{13} - b_{11} b_{23}) / (b_{11} b_{22} - b_{12}^2), \quad (3.18)$$

$$\alpha = \sqrt{\kappa / b_{11}}, \quad (3.19)$$

$$\beta = \sqrt{\kappa b_{11} / (b_{11} b_{22} - b_{12}^2)}, \quad (3.20)$$

$$\gamma = -b_{12} \alpha^2 \beta / \kappa, \quad (3.21)$$

$$\kappa = b_{33} - [b_{13}^2 + v_0 (b_{12} b_{13} - b_{11} b_{23})] / b_{11}. \quad (3.22)$$

After that, the extrinsic parameters are given by

$$\mathbf{r}_1 = \lambda \mathbf{A}^{-1} \mathbf{h}_1, \quad (3.23)$$

$$\mathbf{r}_2 = \lambda \mathbf{A}^{-1} \mathbf{h}_2, \quad (3.24)$$

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2, \quad (3.25)$$

$$\mathbf{t} = \lambda \mathbf{A}^{-1} \mathbf{h}_3, \quad (3.26)$$

$$\text{with } \lambda = 1 / \|\mathbf{A}^{-1} \mathbf{h}_1\| = 1 / \|\mathbf{A}^{-1} \mathbf{h}_2\|,$$

where (-1) indicates the inverse of a matrix. Finally, we refine all of parameters through maximum likelihood estimation based on Levenberg-Marquardt algorithm [65, 66].

3.3.2 Projector Calibration

In the projector calibration stage, projector parameters are estimated by image-to-camera transformation matrix based on the perspective projection and the Euclidian transformation of the camera parameters which encapsulate the position and orientation of the reference planes. Fig. 3.4 shows the relationship among the reference plane, the image plane, and the light stripe. Here, the reference plane is on $z_p = 0$ and the coupled matrix \mathbf{Q} is denoted by $[\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}]$. From (3.1), the perspective projection which maps the reference plane coordinates onto the image coordinates can be written as

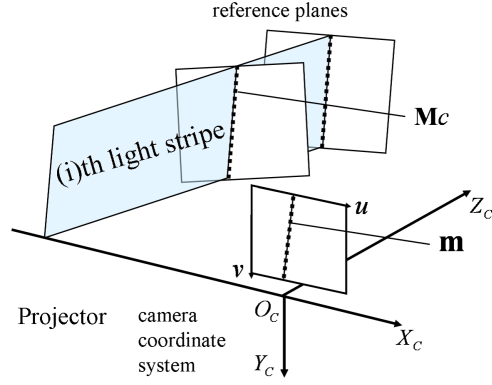


Figure 3.4: Projector calibration.

$$\tilde{\mathbf{m}} \simeq \mathbf{A}\mathbf{Q} \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix}. \quad (3.27)$$

From (3.2), the Euclidean transformation which transforms the reference plane coordinates to the camera coordinates can be written as

$$\mathbf{M}_c = \mathbf{Q} \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix}. \quad (3.28)$$

Furthermore, the inverse of the coupled matrix is given by

$$\begin{aligned} \mathbf{Q}^{-1} &= \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}^{-1} \\ &= \frac{1}{(\mathbf{r}_1 \times \mathbf{r}_2)^T \mathbf{t}} \begin{bmatrix} (\mathbf{r}_2 \times \mathbf{t})^T \\ (\mathbf{t} \times \mathbf{r}_1)^T \\ (\mathbf{r}_1 \times \mathbf{r}_2)^T \end{bmatrix} \\ &= \frac{1}{\mathbf{r}_3^T \mathbf{t}} \begin{bmatrix} (\mathbf{r}_2 \times \mathbf{t})^T \\ (\mathbf{t} \times \mathbf{r}_1)^T \\ \mathbf{r}_3^T \end{bmatrix}, \end{aligned} \quad (3.29)$$

where (T) indicates the transpose of a matrix. From (3.15), (3.16), and (3.17), the transformation matrix which maps the image coordinates into the camera coordinates is given by

$$\begin{aligned}
 \tilde{\mathbf{M}}_c &= \begin{bmatrix} \mathbf{M}_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{Q} \\ \mathbf{b}^T \end{bmatrix} \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} \\
 &\simeq \begin{bmatrix} \mathbf{Q} \\ \mathbf{b}^T \end{bmatrix} \mathbf{Q}^{-1} \mathbf{A}^{-1} \tilde{\mathbf{m}} \\
 &\simeq \begin{bmatrix} \mathbf{I} \\ (\mathbf{r}_3^T \mathbf{t})^{-1} \mathbf{r}_3^T \end{bmatrix} \mathbf{A}^{-1} \tilde{\mathbf{m}}, \tag{3.30}
 \end{aligned}$$

with $\mathbf{I} = \text{diag}(1, 1, 1)$ $\mathbf{b} = [0, 0, 1]$.

The image-to-camera transformation matrix is directly estimated by camera parameters unlike other methods which necessitate recalculations [67, 68, 69]. This matrix has eight degrees of freedom which is similar to the homography matrix in 2D.

For each light stripe, the image coordinates is transformed to the camera coordinates, so that the coefficients of the equation of a plane can be computed by the least square method at least three image coordinates. If the image coordinates of the light stripe are obtained from one reference plane, the equation of a plane cannot be computed. This is how all the light stripes are estimated.

3.4 Experimental Results

The data is captured by a structured light system, Cartesia 3D Handy Scanner of SPACEVISION. This system captures range data in 0.5 seconds with 8 mm focal length, 640×480 pixels, and 256 light stripes by the space encoding method [53, 54]. Here, two light stripes are not used for measurement. The light stripes based on the gray coded pattern are scanned by a single light stripe and a rotatable mirror. The reference plane with the checkered pattern includes 48 calibration points with

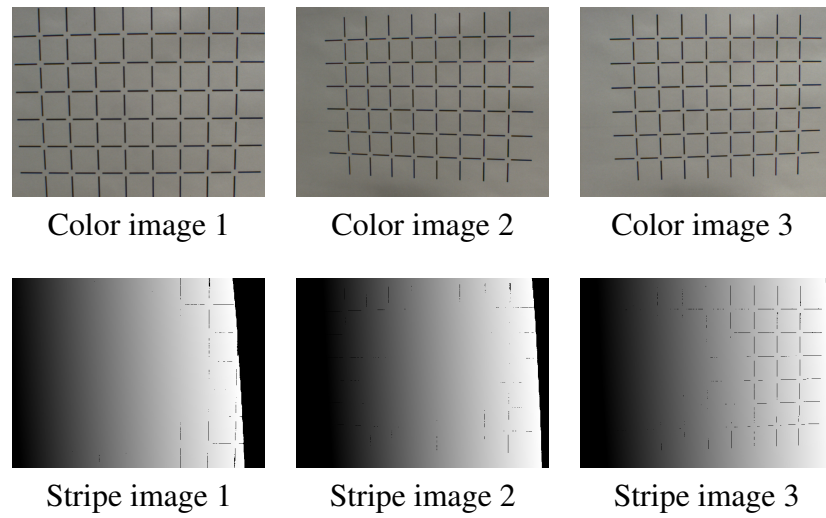


Figure 3.5: Three sets of color images and stripe images.

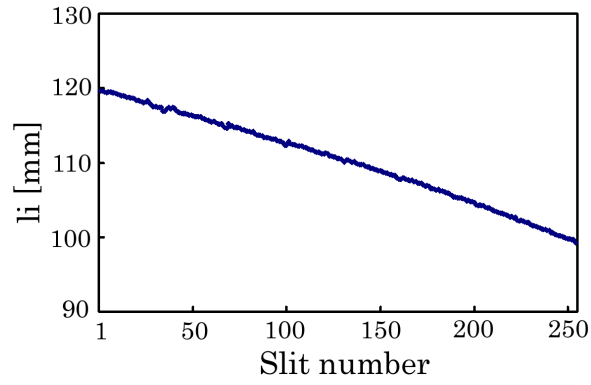
Table 3.1: Camera parameters.

\mathbf{A}	$\begin{bmatrix} 1061.71 & -0.562002 & 350.08 \\ 0 & 1064.09 & 286.547 \\ 0 & 0 & 1 \end{bmatrix}$
k_1	-0.140279
k_2	-0.0916363

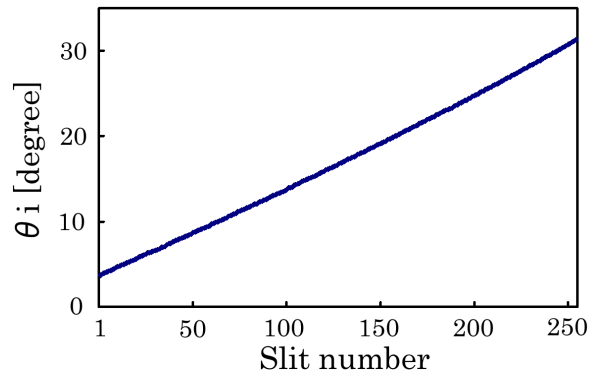
20 mm horizontal and vertical intervals. It is preferable to assign the reference plane within the measurement range and capture a large number of calibration points in the entire image.

3.4.1 Calibration

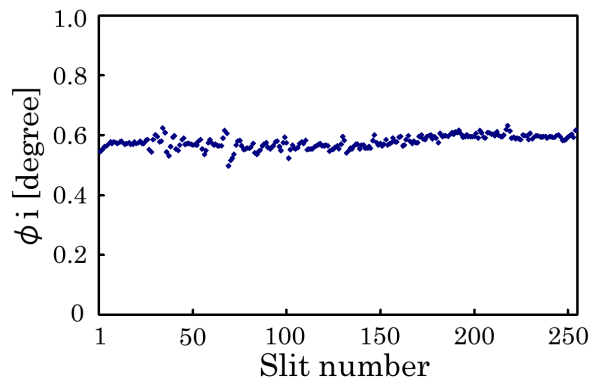
Three sets of color images and stripe images are used for calibration as shown in Fig. 3.5. For the color images, one straight line is fitted to two horizontal line segments and the other straight line is fitted to two vertical segments. The calibration point is detected as the intersection of two straight lines. For stripe



(a) Baselines



(b) Projection angles



(c) Tilt angles

Figure 3.6: Projector parameters.

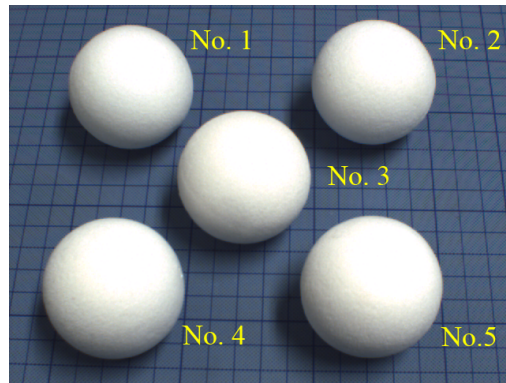


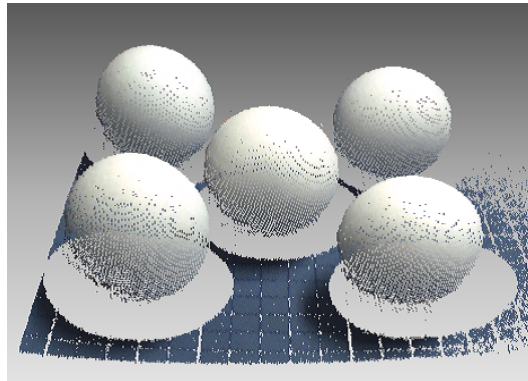
Figure 3.7: Evaluation spheres.

Table 3.2: Measurement accuracy of structured light system.

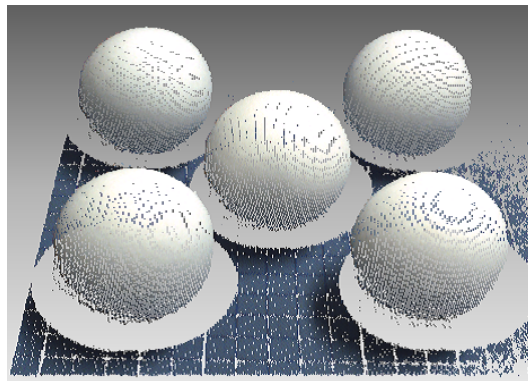
Sphere number	No. 1	No. 2	No. 3	No. 4	No. 5
Measuring points	15,629	15,629	19,405	19,861	19,861
Approach (i)					
$[mm^2]$	0.41	0.38	0.26	0.26	0.31
Approach (ii)					
$[mm^2]$	0.22	0.31	0.19	0.13	0.20
Approach (iii)					
$[mm^2]$	0.23	0.32	0.21	0.15	0.21

images, luminance values from 1 to 254 correspond to the light stripe number. The light stripes are projected to the reference plane vertically.

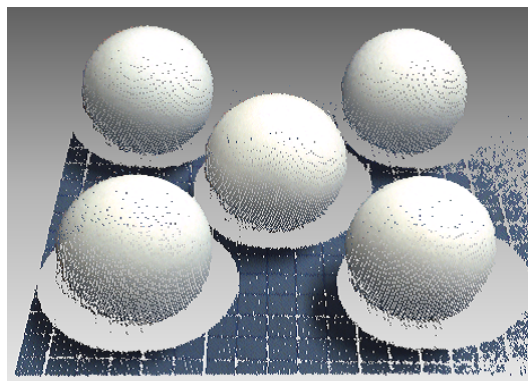
Table 3.1 shows the camera intrinsic matrix and the coefficients of the radial distortion of the camera parameters. Fig. 3.6 shows the baselines, projection angles, and tilt angles of the projector parameters. When the light stripe number increases, the baselines gradually reduce, the projection angles increase, and the tilt angles remains almost constant. The camera and projector parameters enable the system to recover the camera coordinates of a target object.



Approach (i)



Approach (ii)



Approach (iii)

Figure 3.8: Measurement results of five spheres.

3.4.2 Evaluation

We evaluated the measurement accuracy using five spheres with 25 mm radius placed in front of the system. Fig. 3.7 is the evaluation spheres. They are numbered from top left to bottom right. In our evaluation, the system captures range data, and then fit the ideal spheres to them. The measurement accuracy which is defined as the distance between the ideal radius \hat{r} and the real radius r_j is given by

$$E_s = \frac{1}{N_s} \sum_{j=1}^{N_s} (r_j - \hat{r})^2, \quad (3.31)$$

where N_s is the number of measuring points. To show the effectiveness, we evaluated our approach by comparing with two conventional approaches.

- (i) **The pinhole model calibrated by slide stage:** The camera is modeled by the 3×4 projection matrix, and the projector is modeled by the 2×4 projection matrix. The camera and projector parameters are estimated using the slide stage.
- (ii) **The equation of a plane model calibrated by slide stage:** The camera model is based on the pinhole model, and the projector model is based on the equation of a plane model. The camera parameters are obtained by Tsai's method [70], and the projector parameters are estimated using the reference plane.
- (iii) **The equation of a plane model calibrated by reference plane:** The camera model is based on the pinhole model, and the projector model is based on the equation of a plane model. The camera and projector parameters are estimated using the reference plane.

Fig. 3.8 is the measurement results of five spheres. In the approach (i), left two spheres, i.e. No. 1 and No. 4, and the ground are distorted in contrast to the approach (ii) and (iii). Table 3.2 shows the measurement accuracy of the structured light system. In the approach (i), the measurement accuracy is lower than

the approach (ii) and (iii). The approach (ii) and (iii) achieve similar performance. Therefore, the equation of a plane model is applicable to the structured light system. In addition, the reference plane as a planer object provides a high degree of accuracy and has a high degree of availability compared to the slide stage as a cubic object. The measurement accuracy is improved by 44, 16, 19, 42, and 32 percent for each sphere. The experimental results demonstrate the effectiveness and efficiency of our approach.

3.5 Discussion and Conclusion

We presented a novel geometric model and calibration method for a structured light system using a planar object. Typically, the geometry of the structured light system which is modeled by two projection matrices fails to represent its projections when using a rotatable mirror for the projector. In our approach, the geometric model is defined such that the camera model is based on the pinhole model and the projector is based on the equation of a plane model. Although the light stripes do not exactly pass through the optical center, our model can approximate the system geometry. In addition, the camera and projector parameters are estimated by observing the planar object from three viewpoints. Unlike other approaches using cube objects, it facilitates the procedure of user's tasks. The camera parameters are obtained by Zhang's method and the projector parameters are estimated by using image-to-camera transformation matrix. Furthermore, we verify our approach provides a high degree of accuracy in the experiments. The measurement accuracy is improved by 19 to 44 percent compared to traditional approach. To achieve high reliability and effectiveness, tests changing the positions and orientations of the reference plane are required thousands of times. The camera and projector parameters are statistically analyzed. Then, both parameters, in particular, baselines, projection angles, and tilt angles, can be revised and interpolated.

Chapter 4

3D Human Body Measurement

4.1 Introduction

Human body measurement system is classified as either multi-camera system or projector-camera system. We emphasize the advantages and drawbacks of both systems. The multi-camera system captures whole human body data in real time with low resolution and low accuracy. In contrast, the projector-camera system captures whole human body data over a couple of seconds with high resolution and high accuracy. These systems are used to make distinctions depending on application needs. For example, motion tracking systems require real-time data compared to biometrics recognition systems require high resolution and highly accurate data. The following are representative human body measurement systems using multiple cameras or projector-camera pairs.

Moezzi *et al.* [71] proposed a system for generating and replying photorealistic 3D digital video sequences of real events and performances by two or more cameras. The volume intersection method recovers an object's 3D shape by determining whether the corresponding pixel location is part of the object or the background. The model of a single person in any pose has about seventeen thousand voxels or seventy thousand triangles. Tanaka *et al.* [72] proposed a system for capturing time-series volume data of human body by eight cameras. The subject's image regions are extracted by background subtraction and shadow removal. Then, voxel-based visual hull algorithm is applied to reconstruct the volume data. Saito *et al.* [73] proposed a method for shape reconstruction and virtual view generation from multiple cameras. The multi-camera system, called 3D room, is consisted of 49 cameras, ten of which are mounted on each of four walls and nine of which are mounted on the ceiling in the room. The system acquires ten thousand triangles in the mesh to represent a subject on the sofa in real time.

Koo *et al.* [74] presented a contour triangulation based shape reconstruction from multiple projector-camera pairs. The projector-camera system, called 3D Model Studio, is consisted of four projector-camera pairs. Each projector-camera pair which is mounted in the pole moves from the top to the bottom. The system acquires two hundred thousand triangles in the mesh to represent a standing subject in 16.7 seconds. Funatomi *et al.* [75] presented a clustering based surface reconstruction method for upright standing postures. The range data is captured by the projector-camera system with 2.0 mm depth resolution and 1.0 mm measurement accuracy. The system exploits light section method which scans a light stripe from head to toe, so that measuring time is longer than other systems. Treleven [76] reported a whole body scanner and its applications using multiple projector-camera pairs. The projector-camera system, which is the product of $(TC)^2$, is consisted of six projector-camera pairs. Each projector-camera pair is assigned around a subject. The system acquires two hundred thousand points located on the subject's surface within 10 seconds. However, there are not enough viewpoints to capture whole human body data due to occlusion.

In this chapter, a compact and high speed human body measurement system is proposed. Four projector-camera pairs are installed in a pole as a measuring unit which covers the range from head to toe. Each pair captures partial body data in

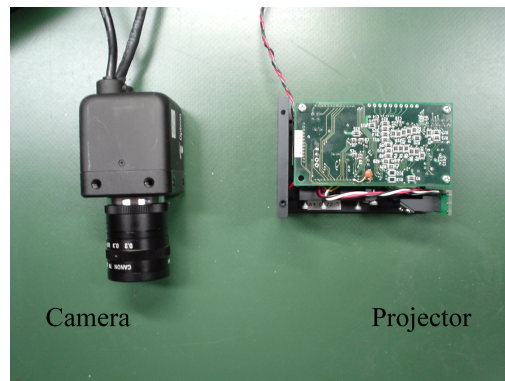


Figure 4.1: Projector-camera pair.

0.5 seconds with 2.0 mm measurement accuracy. Three pole units which have a total of twelve projector-camera pairs are assigned around a subject. Two of them are positioned diagonally forward right and left, and the other one is positioned backward. The system acquires whole human body data in 2 seconds with 1.88 mm measurement accuracy.

4.2 Human Body Measurement System

In this section, we present a method for shape reconstruction of entire body using multiple projector-camera pairs. The small number of projector-camera pairs cannot completely cover entire body and resolve occlusion problem. In addition, it is preferable to finish the measurement in the shortest possible time, because body sway of a subject affects measurement accuracy. In our system, whole human body data is captured from multiple viewpoints by simultaneously measuring.

4.2.1 Projector-Camera Pair

Fig. 4.1 is the structured light system consisting of one camera and one projector. This system captures range data in 0.5 seconds with 6 mm focal length, 640×480 pixels, and 256 light stripes by the space encoding method [53, 54]. The light stripes based on the gray coded pattern are generated by scanning and switching of a single light stripe emitted from the semiconductor laser. The wavelength

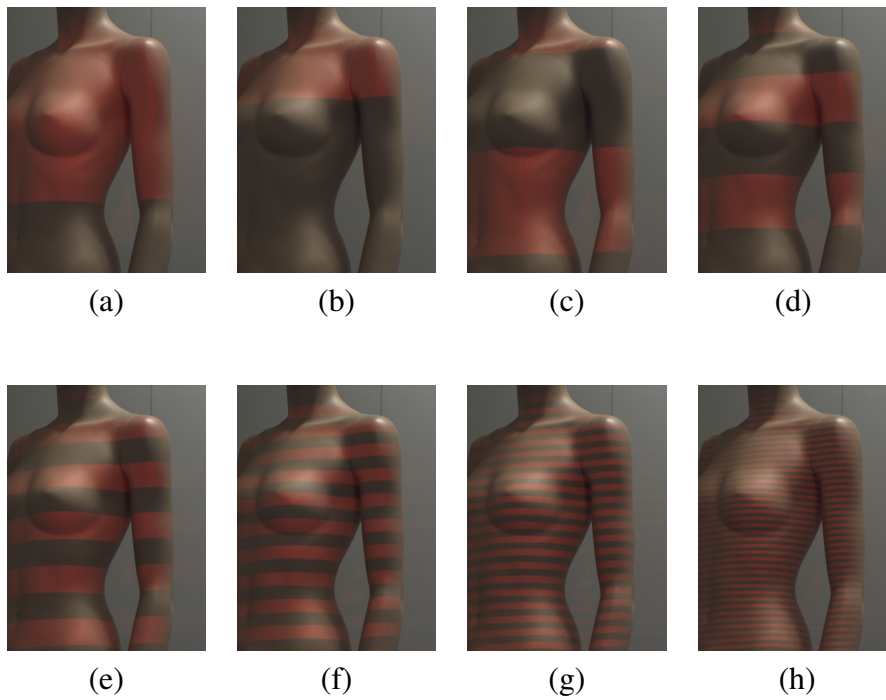


Figure 4.2: 8-bit gray coded patterns.

and power of the laser light are 650 nm and 50 mW, respectively. Using 650 nm bandpass filter, the projector-camera pair is insensitive to illuminations such as fluorescent lights. The measurement error is within 2.0 mm when a target object is located one meter ahead. Fig. 4.2 is the projection of 8-bit gray coded patterns. From (a) to (h), lighted and obscured regions are decreased in width eight times. To simplify the explanation of the algorithm, we indicate 4-bit gray coded patterns in the Table 4.1. Pattern (a), (b), (c), (d) obtain two, three, five, nine horizontal regions, either lighted or obscured region, and indicate first, second, third, fourth bit as a bit sequence. Here, 0 is denoted by lighted region and 1 is obscured region. This four-digit gray code is converted to decimal numbers corresponding to light stripes from 0 to 15. It means that 4-bit gray coded patterns generate 16 light stripes. Of course, 8-bit gray coded patterns generate 256 light patterns in the same way. Each light stripe has the Hamming distance of one between adjacent light stripes, so that it is robust to decoding error.

Table 4.1: 4-bit gray encoding and decoding.

Patten (a)	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	
Patten (b)	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	
Patten (c)	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	
Patten (d)	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	
Light stripe	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

4.2.2 Pole Unit

In our system, four projector-camera pairs are installed in a pole and they are used as a measuring unit. Fig. 4.3 is the drawing of pole unit. For each projector-camera pair, the projector is assigned over the camera. The baseline which is the distance between the camera and the projector is 330 mm. The control computer with Intel D865GRH motherboard, Pentium 4 2.4GHz, 512MB memory is installed in the pole unit. Fig. 4.4 is the pole unit and control computer. The pole unit is 300 mm wide, 265 mm long, 2135 mm high, and 30 kg weight. The measurement range is 800 mm wide and 1800 mm high when a target object is located one meter ahead. It is much wider than the range of one projector-camera pair. Each projector-camera pair is connected to control computer through two USB 2.0 cables. The computer synchronizes the actions of the camera and projector, and generates range data from four sets of color images and stripe images. Fig. 4.5 is the measurement result of a man by using one pole unit. Here, left figure shows texture representation and right figure shows point cloud representation. The pole unit acquires range data from head to toe. Therefore, range data of an entire body can be captured by using two or more pole units.

4.2.3 System Configuration

Let us consider human body measurement and its occlusion problem. The large number of projector-camera pairs can capture range data of an entire body, but their measurement ranges are severely overlapped. In contrast, the small number of projector-camera pairs cannot cover well particular body parts, such as the sub-

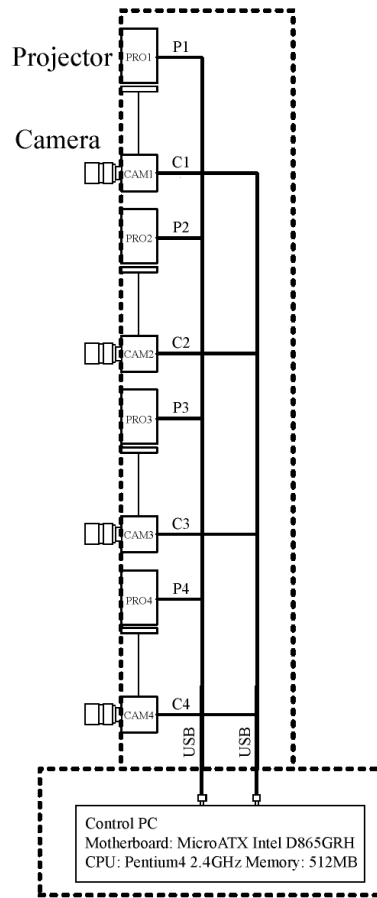


Figure 4.3: Drawing of pole unit.

mental region, axillary region, groin region, and side of a human body. Therefore, we appropriately assign multiple projector-camera pairs around a subject.

Fig. 4.6 is the human body measurement system consisted of three pole units. Twelve projector-camera pairs allow shape reconstruction of entire body. The system is 1200 mm wide and 2000 mm long to cover both men and women in standard proportions. The distance between the pole unit and a subject is approximately 1000 mm and the measurement range is 800 mm round. The number of measuring points is about 1/2 to one million depending on the subject and its pose. The three pole units are movable to change the measurement range. The measurement accuracy is improved by increasing the number of pole units. In contrast, the

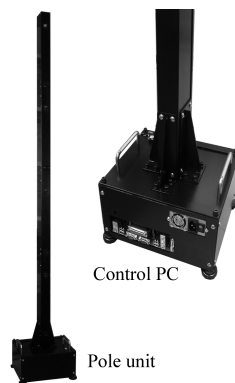


Figure 4.4: Pole unit and control computer.

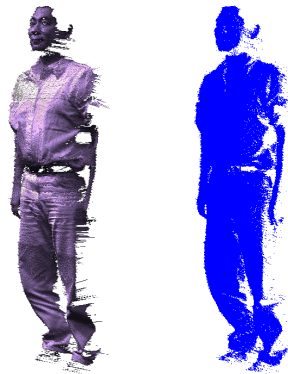


Figure 4.5: Measurement result of a man by one pole unit.

installation space is reduced by decreasing the number of pole units. Therefore, we freely construct the system according to the circumstances.

4.2.4 Measurement Time

During the measurement, subjects need to retain posture as still as possible. The body sways increase every second, so that the measurement should be finished within one second, preferably in real time. If we operate a total of twelve projector-camera pairs one by one, the measurement time will be lengthened. And if we operate some projector-camera pairs simultaneously, range data will not be acquired due to light interference. Fig. 4.7 is the timing diagram. Here, three pole



Figure 4.6: Human body measurement system.

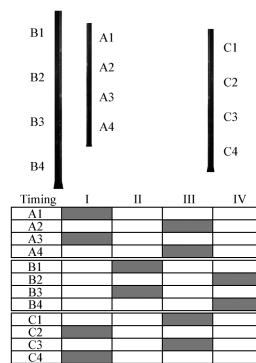


Figure 4.7: Timing diagram.

units are indicated by A, B, C, and four projector-camera pair of the pole unit are numbered from top to bottom. Two or four projector-camera pairs are used simultaneously, so that whole human body data is captured only four times. In timing I, four projector-camera pairs, i.e. A1, A3, C2, C4 differ in height and opposite each other, so that we can avoid light interference. Since the measurement time for each timing is 0.5 seconds, the measurement time is approximately 2 seconds.

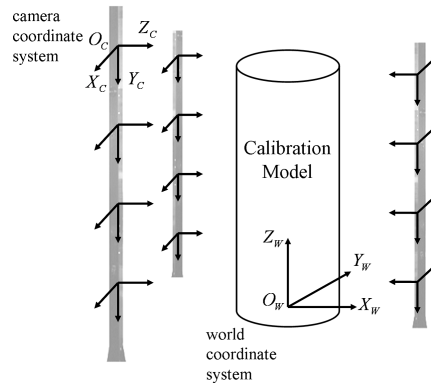


Figure 4.8: Integration into world coordinate system.

4.3 Calibration

In this section, we present calibration method of a projector-camera pair and its coordinate system alignment. For the projector-camera pair, the camera parameters are estimated by Tsai's method and the projector parameters are determined by mechanical specification. After that, coordinate systems of projector-camera pairs are integrated together.

4.3.1 Projector-Camera Pair

The projector-camera pair consists of one camera and one projector. The camera model is based on the pinhole model. The camera parameters, i.e. intrinsic, extrinsic, and radial distortion parameters, are obtained by Tsai's method [70]. The projector model is based on the baseline model. The baseline is defined by product drawing and the projection angles are computed by clock speed of the polygon mirror. Therefore, the projector-camera pair allows us to capture range data by triangulation principle in the camera coordinate system.

4.3.2 Coordinate System Alignment

The human body measurement system is consisted of three pole units with twelve projector-camera pairs. Each projector-camera pair is represented in the camera coordinate system, so that twelve camera coordinate systems are integrated into

the world coordinate system as shown in Fig. 4.8. For the coordinate system alignment, calibration object which is represented in the world coordinate system such as a cylinder and a cube is required. Here, a 3D point, i.e. camera coordinates, \mathbf{M}_c is denoted by $[x_c, y_c, z_c]$ in the camera coordinate system (O_c - X_c - Y_c - Z_c), and a 3D point, i.e. world coordinates, \mathbf{M}_w is denoted by $[x_w, y_w, z_w]$ in the world coordinate system (O_w - X_w - Y_w - Z_w). The affine transformation which transforms the camera coordinates to the world coordinates is given by

$$\tilde{\mathbf{M}}_w = \mathbf{H}_a \tilde{\mathbf{M}}_c, \quad (4.1)$$

$$\text{with } \mathbf{H}_a = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where \mathbf{H}_a is the affine transformation matrix. The tilde indicates the homogeneous coordinate by adding 1 for the additional element: $\tilde{\mathbf{M}}_c = [x_c, y_c, z_c, 1]$ and $\tilde{\mathbf{M}}_w = [x_w, y_w, z_w, 1]$. Affine transformation which has a total of 12 degrees of freedom encapsulates Euclid transformation which has a total of 6 degrees of freedom describing rotation and translation. The twelve parameters $[h_{11}, \dots, h_{34}]$ can be estimated by the least square method at least four camera coordinates. To achieve accurate range data, it is necessary to use a lot of camera coordinates. Since twelve camera coordinate systems are integrated into the world coordinate system, the system captures whole human body data. The assignment and the number of pole units have no constraints as long as all of projector-camera pairs observe a calibration object.

4.4 Experimental Results

The data is captured by our human body measurement system consisting of three pole units with twelve projector-camera pairs. We calibrated the system in two separate steps, and then evaluated the system using representative projector-camera pairs. Furthermore, three subjects are measured to indicate the performance of the system.

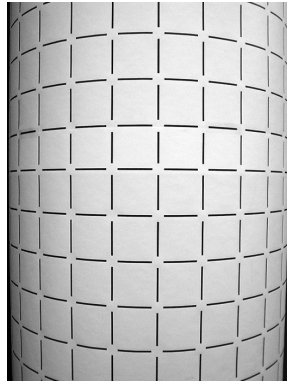


Figure 4.9: Calibration cylinder.

4.4.1 Calibration

The calibration method incorporates two separate stages: projector-camera pair calibration and coordinate system alignment calibration. The camera and projector parameters of twelve projector-camera pairs are estimated, and then their camera coordinate systems are integrated to the world coordinate system. The world coordinate system need to be defined on a cubic object to perform Tsai's method and estimate the affine transformation matrix. Thus, we utilize the calibration cylinder with 415 mm round and 1800 mm height as shown in Fig. 4.9. The cylinder is slightly similar in form and size to a subject compared to cubes and cones, so that it is expected to improve the measurement accuracy. The pole units are located around the cylinder to observe the curved surface at any views. The calibration cylinder contains a checkered pattern, i.e. line segments, with 50 mm horizontal and vertical intervals whose vertices are invisible by erasing a part of lines. The calibration point is detected as the intersection of two straight lines, one of which is fitted to two horizontal line segments and the other one of which is fitted to two vertical segments. Finally, about 80 to 100 calibration points are used for the calibration process.

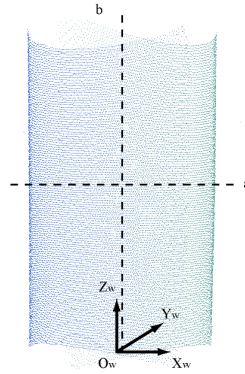


Figure 4.10: Measurement result of the cylinder.

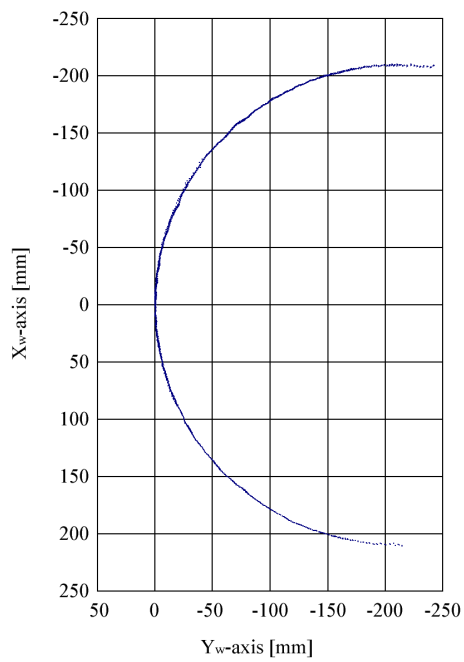


Figure 4.11: Horizontal cross-section.

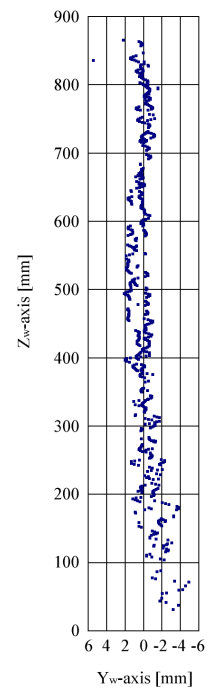


Figure 4.12: Vertical cross-section.

Table 4.2: Measurement accuracy of human body measurement system.

Calibration points	33
Average error [mm]	1.88
Standard deviation [mm]	0.79

4.4.2 Evaluation

We evaluated the measurement accuracy using the calibration cylinder. Fig. 4.10 is the measurement result of the cylinder captured by two projector-camera pairs. In the figure, the symbol, a, indicates the horizontal cross-section and the symbol, b, indicates the vertical cross-section. Here, a 3D point captured by left projector-camera pair is denoted by \mathbf{M}_w^l and a 3D point captured by right projector-camera pair is denoted by \mathbf{M}_w^r . The measurement accuracy which is defined by observing same calibration points from two projector-camera pairs is given by

$$E_h = \frac{1}{N_h} \sum_{k=1}^{N_h} \|\mathbf{M}_{w,k}^l - \mathbf{M}_{w,k}^r\|, \quad (4.2)$$

where N_h is the number of calibration points. Table 4.2 shows the measurement accuracy of the human body measurement system. We used 33 calibration points and obtained the result with 1.88 mm average error and 0.79 mm standard deviation. The error is within 0.2 percent of the distance to a subject. Fig. 4.11 is the horizontal cross-section of the cylinder. The overlapped range of two projector-camera pairs is $x_w = -130 \sim 130$ mm. Subjectively, mapped points are depicted as a smooth curve. Fig. 4.12 is the vertical cross-section of the cylinder. The ideal value is $y_w = 0$ mm. Almost mapped points are distributed around $y_w = -2 \sim 2$ mm. However, on both $z_w = 0 \sim 200$ mm and $z_w = 700 \sim 900$ mm, mapped points are distorted due to radial lens distortion. It turns out that the maximum error is approximately 6 mm.

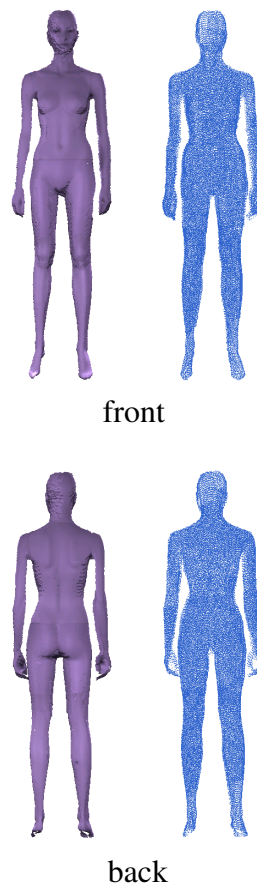


Figure 4.13: Measurement result of a mannequin.

4.4.3 Measurement

Three subjects, i.e. 165 cm tall mannequin, 175 cm tall man, and 173 cm tall clothed man, are measured by our system. Fig. 4.13 is the measurement result of a mannequin. The system successfully acquires range data of entire body, especially occluded parts, i.e. the submental region, axillary region, groin region, and side of a human body. Fig. 4.14 is the measurement result of a man. The body sways increase gradually when capturing, but the system remains nearly unaffected because of simultaneously measuring. Fig. 4.15 is the measurement result of a clothed man. The system mostly acquires range data of head hair and clothes which are difficult to measure due to low reflectance of structured lights emitted from the projector.

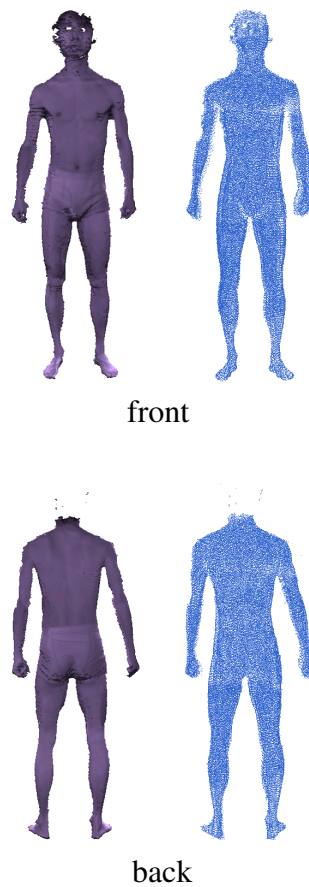


Figure 4.14: Measurement result of a man.

We consider the effect of a structured light on retina of the eye. In our system, the structured light is emitted from class 3B semiconductor laser. Even if the distance between the laser and the eye is 50 mm, the energy density incident upon the pupil is thousandth less legal standard. Since the structured light is generated by expanding the laser beam and scanning it at high speed, the irradiance level per unit area per unit time is substantially below.

4.5 Discussion and Conclusion

We developed a human body measurement system using multiple viewpoints. Although the projector-camera pair captures accurate range data, the measurement

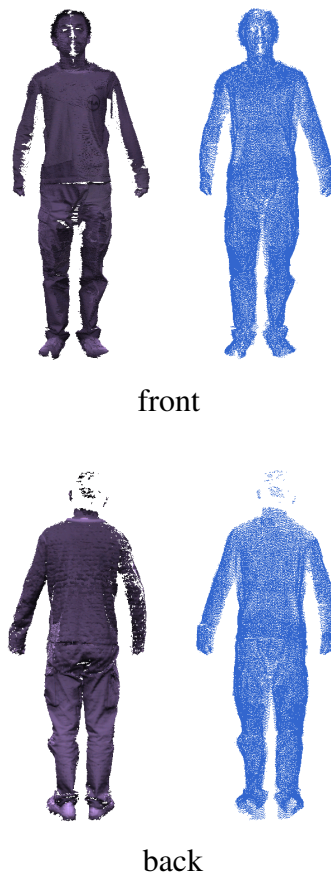


Figure 4.15: Measurement result of a clothed man.

range cannot cover entire body. In addition, it is difficult to measure occluded parts, especially the submental region, axillary region, groin region, and side of a human body. Therefore, we use multiple projector-camera pairs to obtain whole human body data. First, four projector-camera pairs are installed in pole unit. Then, human body measurement system consists of three pole units with twelve projector-camera pairs. The system captures range data of entire body including about 1/2 to one million points in 2 seconds with 1.88 mm average error. The result is within 0.2 percent of the distance to a subject. Unlike other human body measurement systems, the system configuration can be changed freely. Specifically, the pole units are movable to change the measuring range depending on the subject and its pose. The measurement accuracy is improved by increasing the

number of pole units, and also the installation space is reduced by decreasing the number of pole units. However, it is required to perform the calibration whenever the pole units are moved. The operator takes much time and energy for this task. Therefore, a robust and fully automatic calibration method is needed for accurate measurement.

Chapter 5

Human Gait Recognition

5.1 Introduction

Gait recognition aims for personal identification based on walking style. Recognition based on human gait has several advantages as a contactless, exposed, and characteristic biometrics. If the habit of walking is changed consciously, the motion seems unnatural. In addition, gait involves not only surface shape, called static feature, but also continuous motion, called dynamic feature. Nixon *et al.* [77, 78] introduced a total walking cycle defined that the action of walking can be thought of as a periodic signal. The gait cycle is the time interval between instances of initial foot-to-floor contact, called heel strike, for the same foot. Each leg has two periods: swing phase, when the foot is off the floor moving forward to the next step, and stance phase, when the foot is in contact with the

floor. In the medical field, Murray *et al.* [79] introduced standard movement patterns of healthy subjects compared to disabled subjects pathologically. For data collection, required markers are attached to anatomical landmarks of human body. They advocate that the pelvic and thorax rotations are highly variable from one subject to another.

Single-camera based gait recognition methods have been investigated. Usually, they use a silhouette image extracted by background subtraction. Liu and Sarkar [80] developed hidden Markov model-based gait representation by estimating the stance state and then averaging silhouette images. Goffredo *et al.* [81] developed view-independent markerless gait analysis based on the anthropometric propositions of human limbs and the characteristics of gait. To overcome the non-frontal pose problem more recently multi-camera based gait recognition methods have also been developed. These methods exhibited higher recognition accuracy for multi views than that of a single view. Zhao *et al.* [82] presented gait tracking and recognition by matching 3D body model to video sequences. Gait feature is defined by the lengths of key segments and the motion trajectory of lower limbs which is obtained from the inferred model. Huang *et al.* [83] presented the exploitation of the availability of multi views in a gait recognition system. The combination of the results of different views is evaluated to find the improvement in the recognition accuracy. Seely *et al.* [84] presented the University of Southampton Multi-Biometrics Tunnel. The subject's gait is recorded by eight synchronized cameras, and also the face and ear are captured by two videos separately.

Similarly, full body motion capture methods extract a gait feature by identifying joint angles and lengths of body parts. Unlike gait recognition methods, the effectiveness and correctness of the inference and extraction are verified across time. Mikić *et al.* [85] generated a fully automated system for human body model acquisition and tracking using multiple cameras. The system performs the tracking on 3D voxel reconstructions computed from silhouette images. Kalman filter estimates body model parameters based on the measurements made on the labeled voxel data. Deutscher and Reid [86] generated a modified particle filter for searching high-dimensional configuration spaces encountered in visual tracking. Articulated body motion is estimated from simple image features, which are combined with the edge and the silhouette likelihood measures, in video sequences. Yang

and Lee [87] generated a hierarchical search method for recovering human body pose from stereo image sequences. The method searches a cluster for the best matching silhouette image with an input silhouette history image. Human body pose is reconstructed with a linear combination of input range data of a subject. Cheung *et al.* [88, 89] proposed the theory and algorithms for performing temporal shape-from-silhouette in the applications of human modeling and motion tracking. The algorithm is based on the Visual Hull alignment algorithm and utilizes both color and silhouette video sequences. Kehl and Gool [90] proposed an algorithm capable of tracking a person's full body pose during complex motions. The algorithm performs volumetric reconstruction using edges and color information. The articulated model built from super ellipsoids is matched against the image edges and color to overcome ambiguous situations such as touching limbs or strong occlusion. Caillette *et al.* [91] proposed a full body tracker based on Monte-Carlo Bayesian framework. The volumetric reconstruction method follows shape-from-silhouette paradigm. The appearance model represented by Gaussian blobs is fitted onto voxels using the K-means algorithm. The voxels are assigned to the nearest blob using Mahalanobis distances between blobs and voxels using both color and position. Horaud *et al.* [92] proposed a new metric to register a model surface to body data. The model surface of ellipsoids is used to represent body parts, and the body data include point and normal vectors. The metric is defined by the Euclidean distance from the ellipsoid-point to the data-point under the constraint that the ellipsoid-normal and the data-normal are parallel. While multi-camera systems can obtain 3D gait biometrics in real time, the low resolution, low accuracy, and the small number of measuring points are not well suited and qualified to make precise models. In addition, it is important to determine the number of cameras and its assignment in the system.

5.2 Related Work

Biometrics systems generally use single or multi-camera, and extract individual features for human recognition. They are successfully gaining ground and are available for security applications. Recently, biometrics modalities with depth information are an attractive resource. Range scanners have become popular in-

Table 5.1: Comparisons of biometrics approaches for human recognition.

Authors	System		Dimension		Biometrics
	Mul-Cam	Pro-Cam	2D	3D	
Liu and Sarkar [80]	✓		✓		Gait
Goffredo <i>et al.</i> [81]	✓		✓		Gait
Zhao <i>et al.</i> [82]	✓		✓		Gait
Huang <i>et al.</i> [83]	✓		✓		Gait
Seely <i>et al.</i> [84]	✓		✓		Gait
Kakadiaris <i>et al.</i> [93]		✓		✓	Face
Samir <i>et al.</i> [94]		✓		✓	Face
Woodard and Flynn [95]		✓		✓	Finger
Malassiotis <i>et al.</i> [96]		✓		✓	Finger
Chen and Bhanu [97]		✓		✓	Ear
Yan and Bowyer [98]		✓		✓	Ear
Tsalakanidou <i>et al.</i> [99]		✓		✓	Face & Finger
Theoharis <i>et al.</i> [100]		✓		✓	Face & Ear
This thesis		✓		✓	Gait

creasing the measurement accuracy and speed. The following are representative recognition approaches using 3D biometrics.

Kakadiaris *et al.* [93] developed a fully automatic 3D face recognition system based on a composite alignment algorithm to register 3D facial scans with a 3D facial model. The geometry image and normal map image are created by the deformed facial model. They are analyzed by a wavelet transform and the coefficients are used for authentication. Samir *et al.* [94] developed a geometric approach for comparing the shapes of facial surfaces via the shapes of facial curves. The facial surface is represented by a union of level curves of the height function. The metric on shapes of facial surfaces is derived by accumulating distances between corresponding facial curves.

Woodard and Flynn [95] presented a approach for personal identification and identity verification which utilize 3D finger surface features. Three fingers, i.e. the index, middle, and ring finger, are used for comparison to determine subject similarity. To compute a match score, curvature based shape index is used to

represent the finger's surface. Malassiotis *et al.* [96] presented an authentication system based on measurements of 3D finger geometry using a real-time and low-cost sensor. The similarity between training and testing set are computed by the finger width and curvature measurements sampled by the finger length.

Chen and Bhanu [97] proposed a complete human recognition system using 3D ear biometrics. The ear helix/antihelix representation and the local surface patch representation are used to estimate the initial rigid transformation between a gallery-probe pair. A modified iterative closest point algorithm is performed to iteratively refine this transformation. Yan and Bowyer [98] proposed a fully automatic ear biometrics system including automated ear region segmentation and 3D ear shape matching for recognition. The system uses an improved interactive closest point method, combined with point-to-point and point-to-surface approaches, to align a probe surface with a gallery surface.

Multimodal 3D biometrics approaches have been developed in recent years. Tsalakanidou *et al.* [99] presented an authentication system based on the fusion of 3D face and finger biometrics. Theoharis *et al.* [100] presented a unified approach to fuse 3D facial and ear data. These methods achieve high recognition rate when compared to a single modality approach.

As compared to all the works presented in Table 5.1, there are various approaches for 2D and 3D biometrics. Here, Mul-Cam indicates a single or multi-camera system and Pro-Cam indicates a projector-camera system. While biometrics approaches using 3D face, finger, ear, and their multimodal data have been proposed, gait recognition methods still utilize video sequences. Therefore, we attempt to tackle human recognition using 3D gait biometrics where both the modeling and the test data are obtained in 3D.

In this chapter, we present a recognition method using 3D gait biometrics from a projector-camera system. 3D human body data consisting of representative poses over one gait cycle are captured. 3D human body model is fitted to the body data using a bottom-up approach. Since the body data is dense and it is at a high resolution, we can interpolate the entire sequence to fill in between gait acquisitions. Gait features are defined by both dynamic features and static features. The similarity measure based on gait features is used for recognition of a subject and its pose.

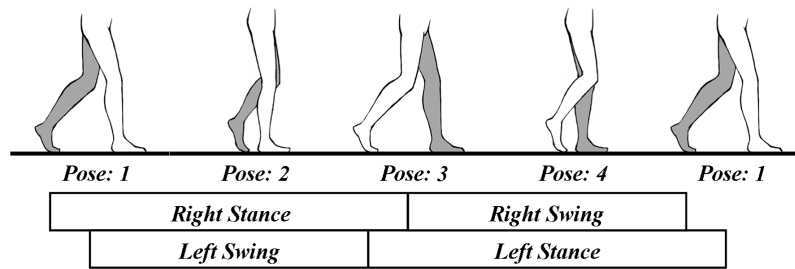


Figure 5.1: Gait cycle expressed by swing phase and stance phase.

5.3 3D Human Body Data

Gait has two distinct periods: a swing phase, when the foot does not touch the ground moving the leg forward, and a stance phase, when the foot touches the ground. Fig. 5.1 is the gait cycle expressed by the swing phase and the stance phase. The cycle begins with *foot touch* which marks the start of the swing phase. The body weight is transferred onto the other leg and the leg swings forward to meet the ground in front of the other foot. The cycle ends with the *foot touch*. The start of stance phase is when the heel strikes the ground. The ankle flexes to bring the foot flat on the ground and the body weight transferred onto it. The end of stance phase is when the heel leaves the ground.

We measure four poses during the cycle by a projector-camera system. The human body measurement system captures high resolution and highly accurate range data of entire body. It includes approximately one million points in a couple of seconds. Here, a 3D point, i.e. model coordinates, \mathbf{M}_m is denoted by $[x_m, y_m, z_m]$ in the model coordinate system (O_m - X_m - Y_m - Z_m). A subject has the following posture conditions:

1. Right foot touches the ground. Right leg (left hand) is in front of the torso and left leg (right hand) is at the back of the torso. The length of stride is the longest during walking.
2. Right foot touches the ground and left foot leaves the ground. Right leg is vertical to the ground and left leg is at the back of the torso. Both hands are along the sides.

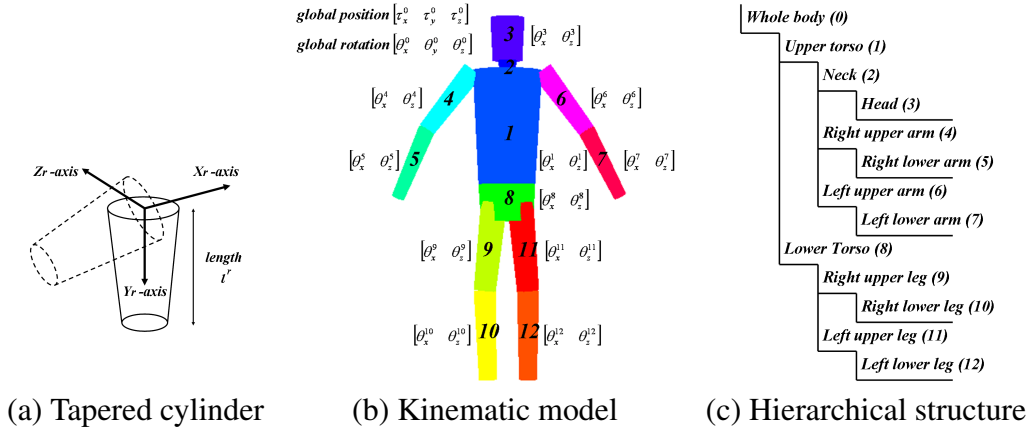


Figure 5.2: 3D human body model.

3. Left foot touches the ground. Left leg (right hand) is in front of the torso and right leg (left hand) is at the back of the torso. The length of stride is the longest during walking.
4. Left foot touches the ground and right foot leaves the ground. Left leg is vertical to the ground and right leg is at the back of the torso. Both hands are along the sides.

Currently, gait databases have tens of images during a gait cycle [101, 102]. We assumed that the measured poses are four of them.

5.4 3D Human Body Model

The model of the human body is based on a kinematic tree consisting of 12 segments, as illustrated in Fig. 5.2. The body segment, r , is approximated by a 3D tapered cylinder which has one free parameter, l^r : the cylinder length. It has two degrees of freedom rotational joint, $[\theta_x^r, \theta_z^r]$, in the local coordinate system $(O_r-X_r-Y_r-Z_r)$. Upper torso is the root segment, i.e. the parent of lower torso, right upper leg, and left upper leg. Similarly, other segments are linked to parent segments by the rotational joints. Table 5.2 shows the bounding angles of rotational joints. The constraints are enforced under the form of bounding values on the

Table 5.2: Bounding angles of rotational joints.

Parts ID	Description	X_r -axis	Y_r -axis	Z_r -axis	Parent
0	Whole body	$[-\pi, \pi]$	$[-\pi, \pi]$	$[-\pi, \pi]$	Root
1	Upper torso	$[-\pi/2, \pi/2]$	$[-0, 0]$	$[-\pi/2, \pi/2]$	0
2	Neck	$[-0, 0]$	$[-0, 0]$	$[-0, 0]$	1
3	Head	$[-\pi/2, \pi/2]$	$[-0, 0]$	$[-\pi/2, \pi/2]$	2
4	Right upper arm	$[-\pi/6, \pi/6]$	$[-0, 0]$	$[-\pi/18, \pi/18]$	1
5	Right lower arm	$[-\pi/4, \pi/4]$	$[-0, 0]$	$[-\pi/18, \pi/18]$	4
6	Left upper arm	$[-\pi/6, \pi/6]$	$[-0, 0]$	$[-\pi/18, \pi/18]$	1
7	Left lower arm	$[-\pi/4, \pi/4]$	$[-0, 0]$	$[-\pi/18, \pi/18]$	6
8	Lower torso	$[-\pi/2, \pi/2]$	$[-0, 0]$	$[-\pi/2, \pi/2]$	0
9	Right upper leg	$[-\pi/4, \pi/4]$	$[-0, 0]$	$[-\pi/18, \pi/18]$	8
10	Right lower leg	$[-\pi/6, \pi/6]$	$[-0, 0]$	$[-\pi/18, \pi/18]$	9
11	Left upper leg	$[-\pi/4, \pi/4]$	$[-0, 0]$	$[-\pi/18, \pi/18]$	8
12	Left lower leg	$[-\pi/6, \pi/6]$	$[-0, 0]$	$[-\pi/18, \pi/18]$	11

joint angles. But, the model has enough range of movement to represent various poses. The whole body is rotated around three axes and other segments are rotated around two axes. Here, neck is the fixed segment between head and upper torso, so that we do not consider the neck joint angles.

The articulated structure of the human body has a total of 40 degrees of freedom (DOFs). The pose is described by a 6-D vector, \mathbf{p} , representing global position and rotation, a 22-D vector, \mathbf{q} , representing the joint angles, and a 12-D vector, \mathbf{r} , representing the lengths of body part as follows.

$$\mathbf{p} = [\tau_x^0, \tau_y^0, \tau_z^0, \theta_x^0, \theta_y^0, \theta_z^0], \quad (5.1)$$

$$\mathbf{q} = [\theta_x^1, \theta_z^1, \theta_x^3, \theta_z^3, \theta_x^4, \theta_z^4, \theta_x^5, \theta_z^5, \theta_x^6, \theta_z^6, \theta_x^7, \theta_z^7, \theta_x^8, \theta_z^8, \theta_x^9, \theta_z^9, \theta_x^{10}, \theta_z^{10}, \theta_x^{11}, \theta_z^{11}, \theta_x^{12}, \theta_z^{12}], \quad (5.2)$$

$$\mathbf{r} = [l^1, l^2, l^3, l^4, l^5, l^6, l^7, l^8, l^9, l^{10}, l^{11}, l^{12}]. \quad (5.3)$$

The combination of the representative four poses is denoted by s . Joint DOF values concatenated along the kinematic tree define the kinematic pose, \mathbf{k} , as a

tuple, $[\mathbf{p}, \mathbf{q}, \mathbf{r}, s]$, where $\mathbf{p} \in \mathbb{R}^6$, $\mathbf{q} \in \mathbb{R}^{22}$, $\mathbf{r} \in \mathbb{R}^{12}$, $s = \{s_1, s_2, s_3, s_4\}$. In the previous works, segments are linked to parent segments by either 1-DOF (hinge), 2-DOF (saddle) or 3-DOF (ball and socket) rotational joints [103]. We use only 2-DOF rotational joints, because the 3D tapered cylinder has rotational symmetry along the direction orthogonal to the radial direction. As a result, we eliminate the twist of body parts as a unnecessary variable.

5.5 Model Fitting

Let us consider human body modeling and its problems. Modeling methods, which use ideal data, sometimes fail when applied to real data [104]. The real data captured by projector-camera systems have some problems. For example, the projector-camera system cannot cover well particular body parts, such as the submental region, axillary region, groin region, and side of a human body, so that the real data are not independently and identically distributed [105]. In addition, the body sways and deep color clothes also have detrimental effects such as holes and gaps. In this section, a modeling method for dealing with the problems occurring in real data is proposed. Our approach to modeling a walking human incorporates four separate steps: body axes estimation, torso detection, arms/legs detection, and head/neck detection.

5.5.1 Body Axes

The intuition behind the principal component analysis (PCA) is to find a set of base vectors, so that they explain the maximum amount of variance of the data [106]. PCA is applied to determine coronal axis (X_m -axis), vertical axis (Y_m -axis), and sagittal axis (Z_m -axis), as illustrated in Fig. 5.3. Our approach to determining the three axes and the centroid incorporates two separate steps. First, we compute the eigenvectors and the mean vector using whole human body data. The first eigenvector, \mathbf{e}_1 , and the mean vector, \mathbf{e}_0 , define the vertical axis and the centroid. The range data of arms and legs do not affect the estimation of the vertical axis and the centroid adversely because they are at symmetric positions in a horizontal direction. Second, we compute the eigenvectors using the extracted

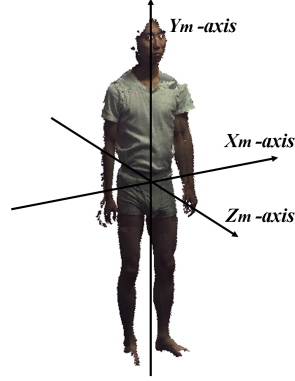


Figure 5.3: Body axes estimation by PCA.

range data of torso after torso detection. The second eigenvector, \mathbf{e}'_2 , and the third eigenvector, \mathbf{e}'_3 , define the coronal axis and the sagittal axis, respectively. The range data of torso is convex and has symmetrical shape even if a subject is walking, so that the two axes can be estimated robustly.

Here, a captured 3D point which is obtained by the human body measurement system is world coordinates and a normalized 3D point which is defined by the three axes and the centroid is model coordinates. The normalization which transforms world coordinates to model coordinates can be written as

$$\mathbf{M}_m = \begin{bmatrix} \mathbf{e}'_2 & \mathbf{u}_2 & \mathbf{e}'_3 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{u}_1 & \mathbf{e}_1 & \mathbf{u}_3 \end{bmatrix}^{-1} (\mathbf{M}_w - \mathbf{e}_0), \quad (5.4)$$

with $\mathbf{u}_1 = [1, 0, 0]$ $\mathbf{u}_2 = [0, 1, 0]$ $\mathbf{u}_3 = [0, 0, 1]$.

Regardless of subject's pose, head and torso face forward and arms and legs swing back and forth. In addition, the model coordinate system is represented by associating the coronal axis, the vertical axis, the sagittal axis, and the centroid with X_m -axis, Y_m -axis, Z_m -axis, and the origin O_m .

5.5.2 Torso

We use the cross-section of a human body to detect upper torso and lower torso. Usually, body data has some holes and gaps, so that the sampling and interpolation process for the extraction of cross-section are required. First, the cross-section is

divided into R regions radiating from a 3D point, \mathbf{w}_h , which is the intersection of the Y_m -axis, with the X_m - Z_m plane. Here, the index h corresponds to the body height. For each region, the closest point from the intersection point is left and the others are removed. If there are no points, one linear interpolated point is calculated using the neighbors. Next, the sample points are connected by the line segments. The area inside the line segments defined by a_l are used for the detection of upper torso and lower torso. The height of the centroid is denoted by g . We assume the cross-sectional area, a_g , is the boundary between the upper torso and lower torso.

The cross-sectional area of the top of upper torso, a_ξ , and the cross-sectional area of the base of lower torso, a_ζ , are given by

$$a_\xi = \delta_{ut} a_g, \quad (5.5)$$

$$a_\zeta = \delta_{lt} a_g, \quad (5.6)$$

where δ_{ut} and δ_{lt} are height parameters. We compute all the cross-sectional areas, and then search for the similar values in a vertical direction. When a_l is smaller than a_ξ (or a_ζ), l is the height of the top of the upper torso (or the base of the lower torso). Then, directional vectors, \mathbf{n}_{ut} and \mathbf{n}_{lt} , of upper torso and lower torso are given by

$$\mathbf{n}_{ut} = \text{median}\left(\frac{\mathbf{w}_h - \mathbf{w}_g}{\|\mathbf{w}_h - \mathbf{w}_g\|}\right) \quad g < h \leq \xi, \quad (5.7)$$

$$\mathbf{n}_{lt} = \text{median}\left(\frac{\mathbf{w}_h - \mathbf{w}_g}{\|\mathbf{w}_h - \mathbf{w}_g\|}\right) \quad \zeta \leq h < g. \quad (5.8)$$

The joint angles of upper torso and lower torso are obtained by the directional vectors. One is the angle between \mathbf{n}_{ut} (or \mathbf{n}_{lt}) and the X_m -axis, and the other is the angle between \mathbf{n}_{ut} (or \mathbf{n}_{lt}) and the Z_m -axis. Therefore, tapered cylinders can be fitted along the directional vectors.

5.5.3 Arms and Legs

We use the fitting of tapered cylinders to detect arms and legs. In this subsection, right/left-upper/lower-arm and right/left-upper/lower-leg are detected by using the same method. Thus, the detection of two of them, which are called the *upper part* and the *lower part* (e.g. right upper arm and right lower arm), takes place. The line segments between the top and base of the cylinders of the upper part and lower part are defined as $L_{up} = \{\mathbf{o}_{up} + \lambda \mathbf{n}_{up} \mid \lambda \in \mathbb{R}^3\}$ and $L_{lp} = \{\mathbf{o}_{lp} + \lambda \mathbf{n}_{lp} \mid \lambda \in \mathbb{R}^3\}$. Here, \mathbf{o}_{up} and \mathbf{o}_{lp} are joint points (shoulder and elbow, or hip and knee). \mathbf{n}_{up} and \mathbf{n}_{lp} are the corresponding directional vectors, which are used as \mathbf{n}_{rua} and \mathbf{n}_{rla} for the right arm, and \mathbf{n}_{rul} and \mathbf{n}_{rll} for the right leg. The distance between the line segment and model coordinates can be written as

$$d_{up} = \sum_{\mathbf{M}_{up} \in \text{upper part}(\mathbf{M}_m)} \left\| \mathbf{M}_{up} - \mathbf{o}_{up} + \frac{\mathbf{M}_{up} \mathbf{n}_{up} - \mathbf{o}_{up} \mathbf{n}_{up}}{\mathbf{n}_{up} \mathbf{n}_{up}} \mathbf{n}_{up} \right\|^2, \quad (5.9)$$

$$d_{lp} = \sum_{\mathbf{M}_{lp} \in \text{lower part}(\mathbf{M}_m)} \left\| \mathbf{M}_{lp} - \mathbf{o}_{lp} + \frac{\mathbf{M}_{lp} \mathbf{n}_{lp} - \mathbf{o}_{lp} \mathbf{n}_{lp}}{\mathbf{n}_{lp} \mathbf{n}_{lp}} \mathbf{n}_{lp} \right\|^2, \quad (5.10)$$

where \mathbf{M}_{up} and \mathbf{M}_{lp} are the model coordinates within each of the parts. We first seek the direction vectors to minimize each of the functions, and then two sets of two joint angles are estimated from the directional vectors. Accordingly, tapered cylinders can be fitted to the arms and legs along the directional vectors.

5.5.4 Head and Neck

Let us consider the difference between head and the other body parts. In the measurement of the head, it is sometimes difficult to capture the shape of hair on the head, because of the low sensitivities to deep color. Therefore, face shape and neck shape are used for the detection. First, the sizes of head and neck are estimated from the distribution of mapped points in the X_m - Y_m plane and Y_m - Z_m plane. Next, the directional vector of head, denoted by \mathbf{n}_h , is determined, and then tapered cylinder is fitted to the head. As stated above, neck does not rotate in our model independently, so that neck's tapered cylinder is placed on upper torso.

5.6 Gait Reconstruction

Gait sequence composed of tens or hundreds of poses is required to analyze and recognize. The representative four poses obtained by fitting body models to body data are used to recover the other poses. Assuming that the motion between pose α and pose β varies linearly, kinematic pose, $\mathbf{k}_f = [\mathbf{p}_f, \mathbf{q}_f, \mathbf{r}_f, s]$, at frame f ($\alpha < f < \beta$) can be written as

$$\mathbf{p}_f = \mathbf{p}_\alpha + (f - \alpha)\mathbf{v}, \quad (5.11)$$

$$\mathbf{q}_f = \mathbf{q}_\alpha + \frac{f - \alpha}{\beta - \alpha}(\mathbf{q}_\beta - \mathbf{q}_\alpha), \quad (5.12)$$

$$\mathbf{r}_f = (\mathbf{r}_{s_1} + \mathbf{r}_{s_2} + \mathbf{r}_{s_3} + \mathbf{r}_{s_4})/4, \quad (5.13)$$

where \mathbf{v} is velocity vector which includes speed and direction, and combination of α and β is expressed by $\{\alpha, \beta\} \in \{\{s_1, s_2\}, \{s_2, s_3\}, \{s_3, s_4\}, \{s_4, s_1\}\}$. The equations allow interpolation of joint angles and lengths of body parts. Therefore, arbitrary poses between representative poses can be recovered.

5.7 Feature Matching

Gait features are divided into two types: (a) dynamic features and (b) static features. For example, the length of stride is one of significant features of human gait. It can be computed by the leg length and its varying angles between poses. In addition, all of joint positions can be computed by using the same method. Therefore, both dynamic feature and static feature are used for recognition. We define the dynamic feature as joint angles, $\mathbf{q}_{m,n}$, and static feature as lengths of the body parts, $\mathbf{r}_{m,n}$. Here, m is the personal identification number, and n is the pose index. To transform these values into a common domain, the normalization is given by

$$\mathbf{q}'_{m,n} = \frac{\mathbf{q}_{m,n} - \boldsymbol{\mu}_q}{\boldsymbol{\sigma}_q}, \quad (5.14)$$

$$\mathbf{r}'_{m,n} = \frac{\mathbf{r}_{m,n} - \boldsymbol{\mu}_r}{\boldsymbol{\sigma}_r}, \quad (5.15)$$

$$\text{where } \boldsymbol{\mu}_q = \frac{1}{M} \frac{1}{N} \sum_m \sum_n \mathbf{q}_{m,n} \quad \boldsymbol{\mu}_r = \frac{1}{M} \frac{1}{N} \sum_m \sum_n \mathbf{r}_{m,n},$$

$$\boldsymbol{\sigma}_q = \left(\frac{1}{M} \frac{1}{N} \sum_m \sum_n (\mathbf{q}_{m,n} - \boldsymbol{\mu}_q)(\mathbf{q}_{m,n} - \boldsymbol{\mu}_q)^T \cdot \mathbf{E} \right)^{1/2} \cdot \mathbf{d},$$

$$\boldsymbol{\sigma}_r = \left(\frac{1}{M} \frac{1}{N} \sum_m \sum_n (\mathbf{r}_{m,n} - \boldsymbol{\mu}_r)(\mathbf{r}_{m,n} - \boldsymbol{\mu}_r)^T \cdot \mathbf{E} \right)^{1/2} \cdot \mathbf{d}.$$

In the formulations, $\boldsymbol{\mu}_q$, $\boldsymbol{\mu}_r$ are the arithmetic means of dynamic and static features, and $\boldsymbol{\sigma}_q$, $\boldsymbol{\sigma}_r$ are the standard deviations of dynamic and static features, M , N are the numbers of people and poses, with the matrix $\mathbf{E} = \text{diag}(1, 1, 1, \dots, 1)$ and the vector $\mathbf{d} = [1, 1, 1, \dots, 1]$. Both features are concatenated on a feature vector $\boldsymbol{\phi}_{m,n} = [\mathbf{q}'_{m,n}, \mathbf{r}'_{m,n}]$. If dynamic feature is only used, a feature vector is defined as $\boldsymbol{\phi}_{m,n} = [\mathbf{q}'_{m,n}]$. Suppose that unknown feature vector, $\boldsymbol{\phi}_U$, is one of $M \times N$ feature vectors, $\boldsymbol{\phi}_{m,n}$. The minimum value of matching scores can be written as

$$E_r = \min_{m,n} \|\boldsymbol{\phi}_U - \boldsymbol{\phi}_{m,n}\|. \quad (5.16)$$

The matching score is computed as L_2 distance. For unknown data, the personal identification number and pose index are recognized.

5.8 Experimental Results

The experiments were performed on the body data set collected by the human body measurement system. It contains twenty-four body data from the representative four poses of six subjects $X \in \{A, B, C, D, E, F\}$.

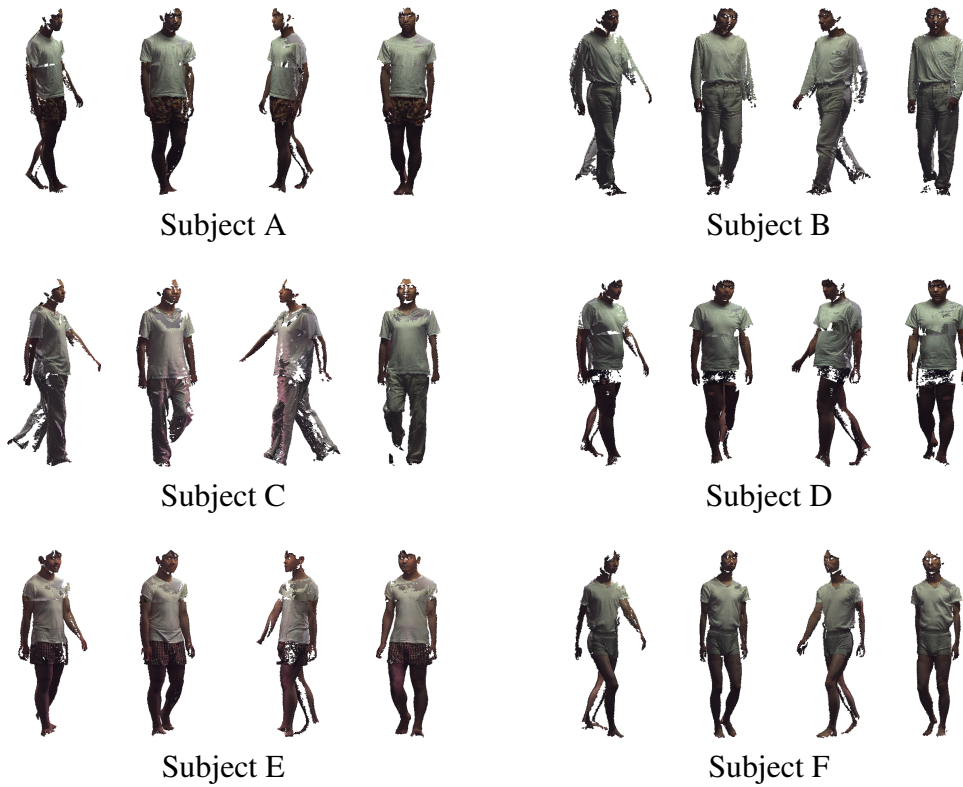


Figure 5.4: 3D human body data of walking humans.

5.8.1 Sensing and Modeling

The body data of representative poses are captured by a human body measurement system, Cartesia 3D Body Scanner of SPACEVISION. The system consisted of nine projector-camera pairs, which acquires nine range data in 3.6 seconds with 640×480 pixels, 3 mm depth resolution, and 3 mm measurement accuracy. We have developed this commercial product based on research results achieved up to now. Projector-camera pairs are calibrated by the proposed geometric model and calibration method in Chapter 4 and their camera coordinate systems are integrated by the improved alignment approach in Chapter 5. Fig. 5.4 is the measurement results of walking humans. The number of measuring points is about 1/2 to one million depending on the subject and the pose. Fig. 5.5 is the results of human body modeling. For the modeling in the experiment we used three parameters: $R = 36$, $\delta_{ut} = 0.25$, and $\delta_{lt} = 0.5$. The body model is fitted to the captured

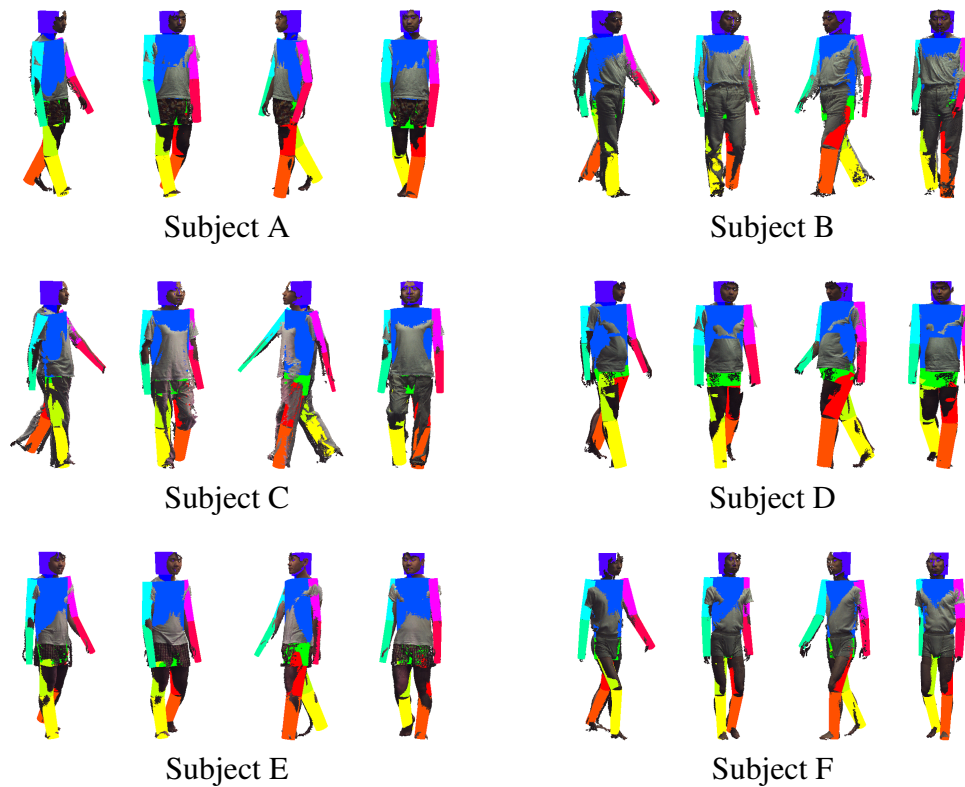


Figure 5.5: 3D human body model fitted to four poses.

body data, so that their joint angles and lengths of body parts are obtained.

5.8.2 Gait Reconstruction

Fig. 5.6 is the results of gait reconstruction. We define that the one gait cycle is composed of twenty frames $Y \in \{1, 2, \dots, 20\}$. The speed is given by dividing the stride length by the number of poses and the direction is given manually. Four of them are representative poses, indicated by the frame index 1, 6, 11, and 16, and the others are interpolated poses, indicated by the frame index 2-5, 7-10, 12-15, and 17-20.

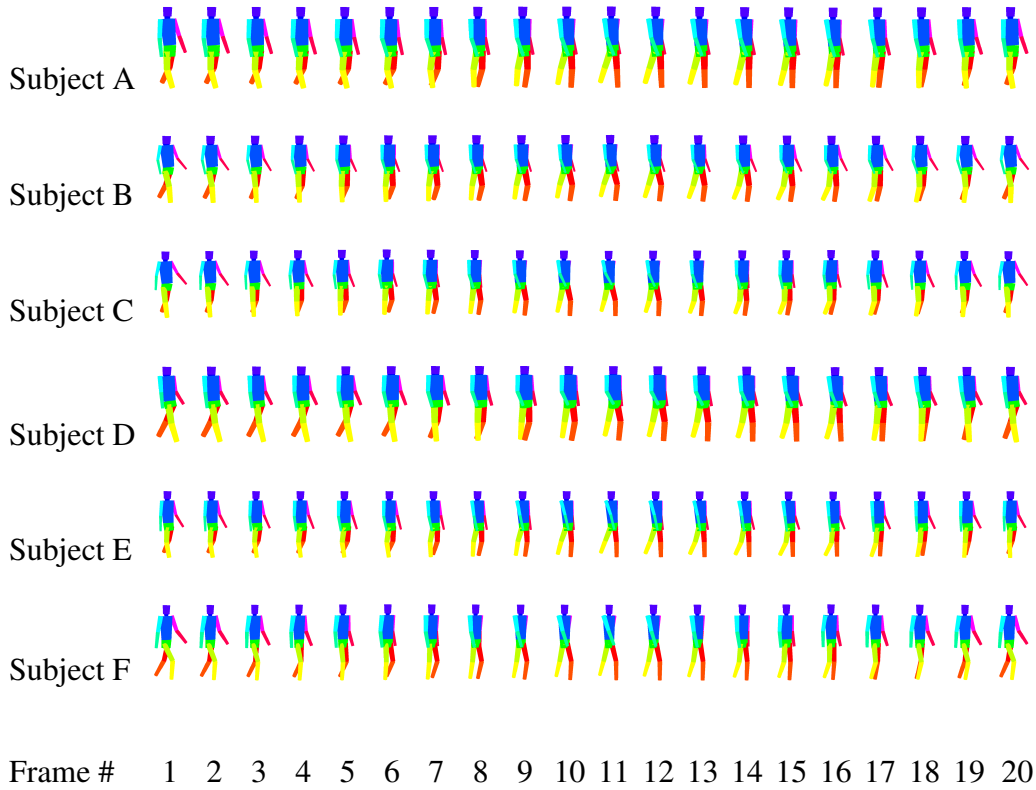


Figure 5.6: Gait sequence composed of twenty frames.

5.8.3 Recognition

The representative poses $s = \{s_1, s_2, s_3, s_4\}$ and their symmetric poses $\bar{s} = \{\bar{s}_1, \bar{s}_2, \bar{s}_3, \bar{s}_4\}$ are used for the experiment. The symmetric poses $\bar{s}_1, \bar{s}_2, \bar{s}_3, \bar{s}_4$ are symmetric to s_3, s_4, s_1, s_2 , respectively. They are synthesized by allocating right (or left) side parameters of representative poses to left (or right) side parameters of symmetrical poses.

For the training data, two gait sequences are recovered by using two combinations of representative poses and symmetrical poses. Fig. 5.7(a) shows the training data of six subjects. One gait sequence is recovered by four poses $\eta_1 = \{s_1, \bar{s}_2, s_3, \bar{s}_4\}$, and the other one is recovered by four poses $\eta_2 = \{\bar{s}_1, s_2, \bar{s}_3, s_4\}$. Each subject has 40 poses, so that training data contains a total of 240 kinematic poses.

For the testing data, one gait sequence is recovered by representative poses

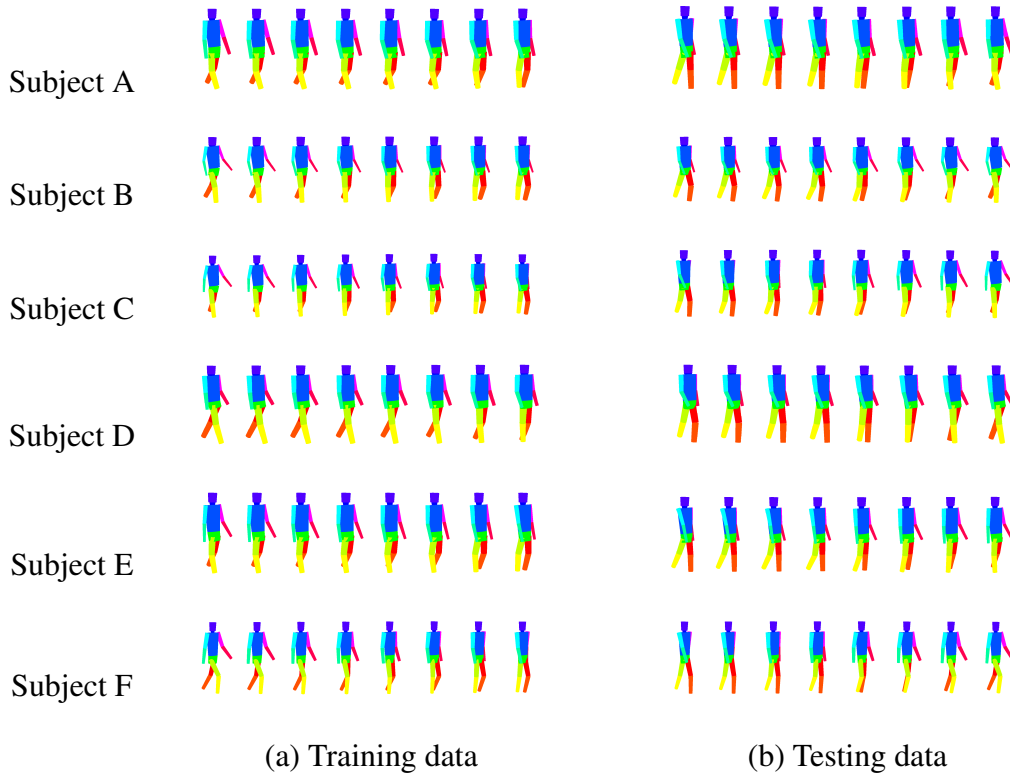


Figure 5.7: Examples of training data and testing data.

$\eta_3 = \{s_1, s_2, s_3, s_4\}$. Fig. 5.7(b) shows the testing data of six subjects. This sequence includes the representative four poses and sixteen interpolated poses. The sixteen interpolated poses are unique and also they are not included in the training data. Therefore, we utilize 96 kinematic poses of six subjects for testing. There is absolutely no overlap between the training and testing data.

In order to evaluate the proposed method, identification rate and average pose error are obtained. The identification rate is obtained by dividing the number of recognized subject by the number of testing data. The pose error is the frame difference between the estimated pose and the ideal pose. Table 5.3 shows that we achieve 98.96 percent using dynamic feature and 100.0 percent using both dynamic and static features for the identification rate. When only dynamic feature is used, the method fails to recognize testing data Subject D with pose 14 who should not be recognized as the training data for Subject B with pose 13.

Table 5.3: 3D gait biometrics performance.

Features	Identification rate (%)	Average pose error (frame)
Dynamic	98.96	0.41
Dynamic & Static	100.0	1.31

Although body types between two subjects are different, their joint angles, i.e. leg and arm swings, are quite similar. In contrast, we achieve 0.41 frame using dynamic feature and 1.31 frame using both features for the average pose error. The experiment using dynamic feature has acceptable results, because it focuses on estimating poses, i.e. dynamic feature cannot consider lengths of body parts. Therefore, both dynamic and static features are useful for gait recognition.

5.8.4 Comparisons

Most of the existing gait recognition methods rely on a single view, i.e. 2D gait biometrics. To compare 3D gait biometrics with 2D gait biometrics, frontal view and side view are synthesized from the training data and testing data by orthographic projection. Fig. 5.8 is the virtual frontal view and side view. For the frontal view, body segment has one free parameter in the X_m - Y_m plane. It has one degree of freedom rotational joint rotating around Z_m -axis. For the side view, body segment has one free parameter in the Y_m - Z_m plane. It has one degree of freedom rotational joint rotating around X_m -axis. Each pose is described by an 11-D vector representing the joint angles, i.e. dynamic feature, and a 12-D vector representing the lengths of body parts, i.e. static feature. 2D gait biometrics is used for the following experiment.

Table 5.4 shows that we achieve 93.75 percent, 95.83 percent using dynamic feature and 100.0 percent, 100.0 percent using both dynamic and static features for the identification rate. In addition, we achieve 1.42 frame, 0.57 frame using dynamic feature and 1.49 frame, 1.29 frame using both features for the average pose error. The experiment using dynamic feature exhibits higher identification rate for 3D gait biometrics which misclassifies 1 subject than that of 2D gait bio-

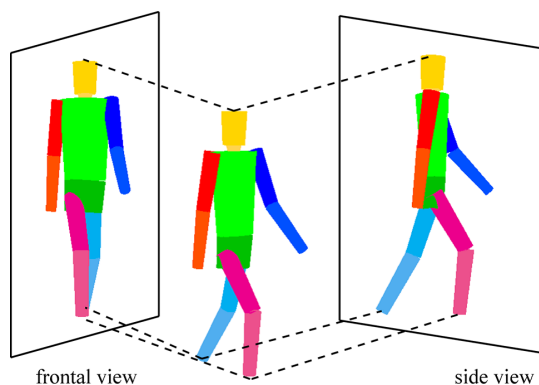


Figure 5.8: Virtual frontal view and side view.

Table 5.4: 2D gait biometrics performance.

View	Features	Identification rate (%)	Average pose error (frame)
Frontal	Dynamic	93.75	1.42
	Dynamic & Static	100.0	1.49
Side	Dynamic	95.83	0.57
	Dynamic & Static	100.0	1.29

metrics, i.e. the frontal view and side view, which misclassifies 6 and 4 subjects, respectively. In addition, it indicates lower average pose error for 3D gait biometrics, which expresses 29 and 72 percent of the frontal view and side view, than that of 2D gait biometrics. In contrast, the experiment using dynamic feature and static feature exhibits the same identification rate and similar average pose error for 3D gait biometrics and 2D gait biometrics which misclassify 0 subjects with over 1 frame. Although pose estimation is not improved, it indicates both dynamic and static features provide highly reliable authentication compared to dynamic feature.

5.9 Discussion and Conclusion

We proposed a new approach for biometric authentication based on 3D human gait. One of the significant weaknesses of current gait recognition methods and

tracking methods has been lack of enough pixels on the human body to fit the body model accurately in the video sequences or the range data obtained by a multi-camera system. In our approach, the body data are captured by a sophisticated projector-camera system. Unlike the multi-camera passive stereo systems used to date, the range data including one million points is high resolution and high accuracy. As a result, the fitted body models and the reconstructed gait sequences are quite accurate, so that the joint angles and lengths of body parts in a gait cycle, i.e. dynamic and static features, possess high reliability. Using training and testing experiments we verified that 3D gait biometrics provide high identification rate to recognize a human subject and his/her pose. Currently, we focus on the issues of modeling and database. In the human body modeling, the fitting approach does not provide matching score, so that we cannot judge whether the fitted model is appropriate or not. It is necessary to evaluate the fitted model or incorporate the notation of uncertainty. In the database, there are a total of twenty-four body data which are consisted of four poses of six subjects. The small size of database used for the evaluation is insufficient to convince about the efficacy. We intend to expand our database to collect one hundred subjects at least and make it publicly available.

Chapter 6

Conclusion and Future Work

6.1 Summary of the Thesis

We presented a framework for range imaging of a human body. The study of this thesis performed a series of tasks related to calibration, sensing, modeling, and recognition. Basically, technical parts, i.e. Chapter 3, Chapter 4, and Chapter 5, are organized in order of four tiers. The first tier is system calibration to estimate intrinsic and extrinsic parameters representing the geometry of camera and projector and the relationships among the projector-camera pairs. The second tier is range sensing to address shape recovery from the images and the lights of single or multiple projector-camera pairs which are assigned around a target object. The third tier is body modeling to approximate body data by a simplified body model indicating the position, length, and volume of main parts of a human body used

as physical features. The fourth tier is feature recognition to address identifying individuals and their looks through the process of feature matching to retrieve personal information that is stored in the database. The following are individual achievements for each part.

The first part of this thesis is the mechanism for calibration of structured light system. A structured light system consists of one camera and one projector. The geometric model is defined such that the camera model is based on the pinhole model and the projector model is based on the equation of a plane model. The measurement accuracy is improved by 19 to 44 percent compared to traditional approach. Next, both camera and projector parameters are estimated by observing a planar object from three arbitrary viewpoints. Unlike other approaches using cube objects, it facilitates the procedure of user's tasks.

The second part of this thesis is the formulation of human body measurement system for acquiring range data of entire body. Projector-camera pair is one of sophisticated structured light systems. In our system, four projector-camera pairs are installed in the pole as a measuring unit. Human body measurement system consists of three pole units with twelve projector-camera pairs. This system obtains whole human body data including about 1/2 to one million points in 2 seconds with 1.88 mm average error. We advocate that the high resolution, highly accurate range data is well suited for security applications.

The third part of this thesis is the introduction of gait biometrics for identification of individuals and their poses. Whole human body data which comprises of representative poses during a gait cycle are captured, and then the positions of joints and body parts are inferred by fitting kinematic models to data set. Unlike attempts which utilize silhouette images have been affected by clothing and self-occlusions, we use range data of entire body to directly extract gait features. In the experiments, we achieve 98.96 percent using dynamic feature and 100.0 percent using both dynamic and static features for the identification rate.

We demonstrated human body measurement system consisting of multiple projector-camera pairs. System calibration is applicable to all of them if reference plane can be seen from three viewpoints. This flexible calibration does not depend on the type of projector as far as light stripes are used for the projection. In addition, range sensing provides high degree of accuracy as experimental re-

sults verified. The gray coded pattern is utilized for the measurement because of high speed encoding process and robust decoding process. We proposed gait recognition method using range data of walking human. Body modeling extracts physical features by fitting body model to the body data. Our simplistic model focuses on feature extraction unlike detailed model in the computer graphics field. Furthermore, feature recognition is performed based on similarity measure for individual human recognition. Both static feature and dynamic feature are used as a continuous motion of human gait.

6.2 Future Work and Possible Extensions

Using a large number of projector-camera pairs, the measurement system covers wide range of the target from small object to large-scale environment. It is applicable to preventing product failures such as car parts which are checked by single camera against resolution or probe against time. It is also appropriate for protecting cultural assets like sculptures that have been deteriorated over the centuries. If one hundred or more pairs are assigned in public spaces such as airports and offices, spatiotemporal recognition based on range data of the scene is realized. At the same time, it is difficult to capture the data in real time and measure the target under the sun. The measurement time is improved by increasing the speed of emitting light stripes of the projector and setting the maximum shutter speed of the camera. To avoid increased illuminations, we intend to seek out solutions by using high power laser at various wavelength bands. Overcoming mechanical limitations on the system, the ideas of system calibration and range sensing need to be changed. Imaginarily, artificial darkness which can be controlled will achieve a tremendous breakthrough. Unlike switching on the light in a dark room, we push the button to darken the place in bright light. This way of thinking is unrealizable in the modern world of technology and science.

As for recognition of walking humans, we are building a large-scale database of body data, called UCR 3D gait database. The database contains range data, name, gender, age and assigned ID number. The range data of five selected poses, i.e. one standing posture and four walking postures, are captured by Cartesia 3D Body Scanner of SPACEVISION. Currently, we have a total of 150 body data



Figure 6.1: UCR 3D gait database.

of 30 subjects. Fig. 6.1 is the measurement results of twelve subjects which are stored in the database. Body data has not only three-dimensional coordinates but also color information every coordinate. They are used for the study of body modeling and gait recognition to verify the effectiveness of our approach using one hundred subjects at least. We intend to dismiss some claims that identification rate on a small database is not appropriate for biometrics authentication. Then, proposed recognition algorithm based on similarity measure is needed to revise in several respects. Although dynamic feature and static feature are simply combined as a feature vector, the extraction of valuable elements, based on principal component analysis, overcomes limitations of the algorithm. In addition, linear/multiple discriminant analysis which distinguishes between two or more groups is useful for classification of input data. Human gait is attractive modality for recognizing

people and becoming widely used in surveillance systems. Therefore, the goal is to establish an automatic gait recognition algorithm on the large-scale database.

Bibliography

- [1] P. Treleaven and J. Wells. 3D body scanning and healthcare applications. *IEEE Computer*, 40(7):28–34, 2007.
- [2] N. D’Apuzzo and H. Mitchell. Medical applications. In Z. Li, J. Chen, and E. Baltsavias, editors. *Advances in Photogrammetry, Remote Sensing and Spatial Information Sciences*. 425–438, Taylor & Francis Group, London, United Kingdom, 2008.
- [3] W. C. Röntgen. On a new kind of rays. *Nature*, 53:274–276, 1896.
- [4] A. M. Cormack. Representation of a function by its line integrals, with some radiological applications. *Journal of Applied Physics*, 34(9):2722–2727, 1963.
- [5] P. C. Lauterbur. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242:190–191, 1973.
- [6] A. Keys, F. Fidanza, M. J. Karvonen, N. Kimura, and H. L. Taylor. Indices of relative weight and obesity. *Journal of chronic diseases*, 25(6-7):329–343, 1972.
- [7] N. D’Apuzzo. 3D body scanning technology for fashion and apparel industry. In *Proceedings of SPIE 6491*, 1–12, 2007.
- [8] N. Werghi. Segmentation and modeling of full human body shape from 3-D scan data: a survey. *IEEE Trans. Systems, Man, and Cybernetics, Part C*, 37(6):1122–1136, 2007.
- [9] M. Nixon, J. Carter, D. Cunado, P. Huang, and S. Stevenage. Automatic Gait Recognition. In A. Jain, R. Bolle, and S. Pankanti, editors. *Biometrics: personal identification in networked society*. 273–285, Springer-Verlag, New York, NY, United States, 1999.

- [10] E. Paquet and H. L. Viktor. Adjustment of virtual mannequins through anthropometric measurements, cluster analysis, and content-based retrieval of 3-D body scans. *IEEE Transactions on Instrumentation and Measurement*, 56(5):1924–1929, 2007.
- [11] K. W. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. *Computer Vision and Image Understanding*, 101(1):1–15, 2006.
- [12] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2D and 3D face recognition: a survey. *Pattern Recognition Letters*, 28(14):1885–1906, 2007.
- [13] B. Bhanu and H. Chen. *Human ear recognition by computer*. Springer-Verlag, New York, NY, United States, 2008.
- [14] Y. Shirai. Recognition of polyhedrons with a range finder. *Pattern Recognition*, 4(3):243–250, 1972.
- [15] Y. Sato, H. Kitagawa, and H. Fujita. Shape measurement of curved objects using multiple slit-ray projections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(6):641–646, 1982.
- [16] K. Araki, Y. Sato, N. Tanaka, and T. Fujino. A method for high speed 3-D range measurement and its trail instrumentation. In *Proceedings of International Conference on Pattern Recognition*, 2:755–757, 1988.
- [17] T. Kanade, A. Gruss, and L. R. Carley. A very fast VLSI rangefinder. In *Proceedings of International Conference on Robotics and Automations*, 2:1322–1329, 1991.
- [18] J. L. Posdamer and M. D. Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18:1–17, 1982.
- [19] M. D. Altschuler, B. R. Altschuler, and J. Taboada. Laser electro-optic system for rapid three-dimensional (3-D) topographic mapping of surfaces. *Optical Engineering*, 20(6):953–961, 1981.
- [20] S. Inokuchi, K. Sato, and F. Matsuda. Range-imaging system for 3D object recognition. In *Proceedings of International Conference on Pattern Recognition*, 806–808, 1984.

- [21] K. Sato and S. Inokuchi. Range-imaging system utilizing nematic liquid crystal mask. In *Proceedings of International Conference on Computer Vision*, 657–661, 1987.
- [22] Y. Sato and M. Otsuki. Three-dimensional shape reconstruction by active rangefinder. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 142–147, 1993.
- [23] P. Gunaratne, H. Shogo, and Y. Sato. 3-D face modeling with multiple range images. In *Proceedings of International Conference on Pattern Recognition*, 4:559–562, 2000.
- [24] K. Hasegawa, K. Noda, and Y. Sato. Electronic endoscope system for shape measurement. In *Proceedings of International Conference on Pattern Recognition*, 1:792–795, 2002.
- [25] Y. Aoki, M. Terajima, and A. Nakasima. Development of 4D jaw movement visualization system for dental diagnosis support. In *Proceedings of SPIE 5603*, 1–8, 2004.
- [26] M. Kouchi and M. Mochimaru. Development of a low cost foot-scanner for a custom shoe making system. In *Proceedings of ISB Symposium on Footwear Biomechanics*, 58–59, 2001.
- [27] K. L. Boyer and A. C. Kak. Color-encoded structured light for rapid active ranging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(1):14–28, 1987.
- [28] P. Vuylsteke and A. Oosterlinck. Range image acquisition with a single binary-encoded light pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):148–164, 1990.
- [29] J. Salvi, J. Batlle, and E. Mouaddib. A Robust-Coded Pattern Projection for Dynamic 3D Scene Measurement. *Pattern Recognition Letters*, 19(11):1055–1065, 1998.
- [30] L. Zhang, B. Curless, and S. M. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In *Proceedings of International Symposium on 3D Data Processing Visualization and Transmission*, 24–36, 2002.

- [31] H. Morita, K. Yajima, and S. Sakata. Reconstruction of surfaces of 3-D objects by M-array pattern projection method. In *Proceedings of International Conference on Computer Vision*, 468–473, 1988.
- [32] P. M. Griffin, L. S. Narasimhan, and S. R. Yee. Generation of uniquely encoded light patterns for range data acquisition. *Pattern Recognition*, 25(6):609-616, 1992.
- [33] R. A. Morano, C. Ozturk, R. Conn, S. Dubin, S. Zietz, and J. Nissano. Structured light using pseudorandom codes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):322–327, 1998.
- [34] M. Kimura, M. Mochimaru, and T. Kanade. Measurement of 3D foot shape deformation in motion. In *Proceedings of International Workshop on Projector-Camera Systems*, 1–8, 2008.
- [35] K. M. Robinette, H. Daanen, and E. Paquet. The Caesar project: a 3-D surface anthropometry survey. In *Proceedings of International Conference on 3-D Digital Imaging and Modeling*, 380–386, 1999.
- [36] Z. B. Azouz, C. Shu, R. Lepage, and M. Rioux. Extracting main modes of human body shape variation from 3D anthropometric data. In *Proceedings of International Conference on 3-D Digital Imaging and Modeling*, 335–342, 2005.
- [37] M. Mochimaru, M. Kouchi, and M. Dohi. Analysis of 3-D human foot forms using the Free Form deformation method and its application in grading shoe lasts. *Ergonomics*, 43(9):1301–1313, 2000.
- [38] B. Allen, B. Curless, and Z. Popović. The space of human body shapes: reconstruction and parameterization from range scans shapes. In *Proceedings of ACM SIGGRAPH 2003*, 587–594, 2003.
- [39] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. SCAPE: shape completion and animation of people. *ACM Transactions on Graphics*, 24(3):408–416, 2005.
- [40] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital Michelangelo project: 3D scanning of large statues. In *Proceedings of ACM SIGGRAPH 2000*, 131–144, 2000.

- [41] D. Koller, M. Turitzin, M. Levoy, M. Tarini, G. Croccia, P. Cignoni, and R. Scopigno. Protected interactive 3D graphics via remote rendering. *ACM Transactions on Graphics*, 23(3):695–703, 2004.
- [42] K. Ikeuchi, A. Nakazawa, K. Hasegawa, T. Ohishi. The Great Buddha project: modeling cultural heritage for VR systems through observation. *International Symposium on Mixed and Augmented Reality*, 7–16, 2003.
- [43] K. Ikeuchi, T. Oishi, J. Takamatsu, R. Sagawa, A. Nakazawa, R. Kurazume, K. Nishino, M. Kamakura, and Y. Okamoto. The Great Buddha project: digitally archiving, restoring, and analyzing cultural heritage objects. *International Journal of Computer Vision*, 75(1):189–208, 2007.
- [44] K. Ikeuchi and D. Miyazaki. *Digitally archiving cultural objects*. Springer-Verlag, New York, NY, United States, 2007.
- [45] A. Banno, T. Masuda, T. Oishi, and K. Ikeuchi. Flying laser range sensor for large-scale site-modeling and its applications in Bayon digital archival project. *International Journal of Computer Vision*, 78(2-3):207–222, 2008.
- [46] R. C. Bolles, J. H. Kremers, and R. A. Cain. A simple sensor to gather three-dimensional data. *Technical Report 249*, Stanford Research Institute, 1981.
- [47] Y. F. Li and S. Y. Chen. Automatic recalibration of an active structured light vision system. *IEEE Transactions on Robotics and Automation*, 19(2):259–268, 2003.
- [48] M. Fukuda, T. Miyasaka, and K. Araki. A prototype system for 3D measurement using flexible calibration method. In *Proceeding of SPIE 6056*, 304–312, 2006.
- [49] B. Zhang, Y. F. Li, and Y. H. Wu. Self-recalibration of a structured light system via plane-based homography. *Pattern Recognition*, 40(4):1368–1377, 2007.
- [50] M. Matsuki and T. Ueda. A real-time sectional image measuring system using time sequentially coded grating method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11):1225–1228, 1989.
- [51] G. Sansoni, M. Carocci, and R. Rodella, Calibration and performance evaluation of a 3-D imaging sensor based on the projection of structured light. *IEEE Transactions on Instrumentation and Measurement*, 49(3):628–636, 2000.

- [52] G. Godin, F. Blais, L. Cournoyer, A. Beraldin, J. Domey, J. Taylor, M. Rioux, and S. El-Hakim. Laser range imaging in archaeology: issues and results. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop*, 1–6, 2003.
- [53] K. Hattori and Y. Sato. Handy rangefinder for active robot vision. In *Proceedings of International Conference on Robotics and Automations*, 2:1423–1428, 1995.
- [54] K. Hattori and Y. Sato. Accurate rangefinder with laser pattern shifting. In *Proceedings of International Conference on Pattern Recognition*, 3:849–853, 1996.
- [55] I. D. Reid. Projective calibration of a laser-stripe range finder. *Image and Vision Computing*, 14(9):659–666, 1996.
- [56] J. Wakitani, T. Maruyama, T. Morita, T. Uchiyama, and A. Mochizuki. Wrist-mounted laser rangefinder. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3:362–367, 1995.
- [57] R. S. Lu and Y. F. Li. Calibration of a 3D vision system using pattern projection. *Sensors and Actuators A: Physical*, 104(1):94–102, 2003.
- [58] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):965–980, 1992.
- [59] R. I. Hartley and S. B. Kang. Parameter-free radial distortion correction with centre of distortion estimation. In *Proceedings of International Conference on Computer Vision*, 2:1834–1841, 2005.
- [60] G. Wei and S. Ma. Implicit and explicit camera calibration: theory and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5):469–480, 1994.
- [61] O. Faugeras and Q. T. Luong. *The Geometry of Multiple Images*. MIT Press, Cambridge, MA, United States, 2001.
- [62] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, United Kingdom, 2004.

- [63] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.
- [64] Z. Zhang. A flexible new technique for camera calibration. *Technical Report MSR-TR-98-71*, Microsoft Research, 1998.
- [65] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2:164–168, 1944.
- [66] D. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11(2):431–441, 1963.
- [67] C. H. Chen and A. C. Kak. Modeling and calibration of a structured light scanner for 3-D robot vision. In *Proceedings of International Conference on Robotics and Automation*, 807–815, 1987.
- [68] D. Q. Huynh, R. A. Owens, and P. E. Hartmann. Calibrating a structured light stripe system: a novel approach. *International Journal of Computer Vision*, 33(1):73–86, 1999.
- [69] F. Zhou and G. Zhang. Complete calibration of a structured light stripe vision sensor through planar target of unknown orientations. *Image and Vision Computing*, 23(1):59–67, 2005.
- [70] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lens. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.
- [71] S. Moezzi, L. C. Tai, and P. Gerard. Virtual view generation for 3D digital video. *IEEE Multimedia*, 4(1):18–26, 1997.
- [72] H. Tanaka, A. Nakazawa, and H. Takemura. Human pose estimation from volume data and topological graph database. In *Proceedings of Asian Conference on Computer Vision*, 618–627, 2007.
- [73] H. Saito, S. Baba, and T. Kanade. Appearance-based virtual view generation from multicamera videos captured in the 3-D room. *IEEE Transactions on Multimedia*, 5(3):303–316, 2003.

- [74] B. K. Koo, Y. K. Choi, and S. I. Chien. 3D human whole body construction by contour triangulation. *IEICE Transactions on Information and Systems*, E87-D(1):233–243, 2004.
- [75] T. Funatomi, I. Moro, S. Mizuta, and M. Minoh. Surface reconstruction from point cloud of human body by clustering. *Systems and Computers in Japan*, 37(11):44–56, 2006.
- [76] P. Treleaven. Sizing us up. *IEEE Spectrum*, 41(4):28–31, 2004.
- [77] M. S. Nixon, T. Tan, and R. Chellappa. *Human identification based on gait*. Springer-Verlag, New York, NY, United States, 2005.
- [78] M. S. Nixon and J. N. Carter. Automatic recognition by gait. In *Proceedings of the IEEE*, 94(11):2013–2024, 2006.
- [79] M. P. Murray, A. B. Drought, and R. C. Kory. Walking patterns of normal men. *Journal of Bone and Joint Surgery*, 46A(2):335–360, 1964.
- [80] Z. Liu and S. Sarkar. Improved gait recognition by gait dynamics normalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6):863–876, 2006.
- [81] M. Goffredo, R. D. Seely, J. N. Carter, and M. S. Nixon. Markerless view independent gait analysis with self-camera calibration. In *Proceedings of International Conference on Automatic Face and Gesture Recognition*, 1–6, 2008.
- [82] G. Zhao, G. Liu, H. Li, and M. Pietikäinen. 3D gait recognition using multiple cameras. In *Proceedings of International Conference on Automatic Face and Gesture Recognition*, 529–534, 2006.
- [83] X. Huang and N. V. Boulgouris. Human gait recognition based on multiview gait sequences. *EURASIP Journal on Advances in Signal Processing*, 2008:1–8, 2008.
- [84] R. D. Seely, S. Samangoei, L. Middleton, J. N. Carter, and M. S. Nixon. The university of southampton multi-biometric tunnel and introducing a novel 3D gait dataset. In *Proceedings of International Conference on Biometrics: Theory, Applications and Systems*, 1–6, 2008.

- [85] I. Mikić, M. Trivedi, E. Hunter, and P. Cosman. Human body model acquisition and tracking using voxel data. *International Journal of Computer Vision*, 53(3):199–223, 2003.
- [86] J. Deutscher and I. Reid. Articulated body motion capture by stochastic search. *International Journal of Computer Vision*, 61(2):185–205, 2005.
- [87] H. D. Yang and S. W. Lee. Reconstruction of 3D human body pose from stereo image sequences based on top-down learning. *Pattern Recognition*, 40(11):3120–3131, 2007.
- [88] K. M. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette across time part I: theory and algorithms. *International Journal of Computer Vision*, 62(3):221–247, 2005.
- [89] K. M. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette across time part I: applications to human modeling and markerless motion tracking. *International Journal of Computer Vision*, 63(3):225–245, 2005.
- [90] R. Kehl and L. V. Gool. Markerless tracking of complex human motions from multiple views. *Computer Vision and Image Understanding*, 104(2):190–209, 2006.
- [91] F. Caillette, A. Galata, and T. Howard. Real-time 3-D human body tracking using learnt models of behavior. *Computer Vision and Image Understanding*, 109(2):112–125, 2008.
- [92] R. Horaud, M. Niskanen, G. Dewaele, and E. Boyer. Human motion tracking by registering an articulated surface to 3D points and normals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):158–164, 2009.
- [93] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):640–649, 2007.
- [94] C. Samir, A. Srivastava, and M. Daoudi. Three-dimensional face recognition using shapes of facial curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1858–1863, 2006.

- [95] D. L. Woodard and P. J. Flynn. Finger surface as a biometric identifier. *Computer Vision and Image Understanding*, 100(3):357–384, 2005.
- [96] S. Malassiotis, N. Aifanti, and M. G. Strintzis. Personal authentication using 3-D finger geometry. *IEEE Transactions on Information Forensics and Security*, 1(1):12–21, 2006.
- [97] H. Chen and B. Bhanu. Human ear recognition in 3D. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):718–737, 2007.
- [98] P. Yan and K. W. Bowyer. Biometric recognition using 3D ear shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1297–1308, 2007.
- [99] F. Tsalakanidou, S. Malassiotis, and M. G. Strintzis. A 3D face and hand biometric system for robust user-friendly authentication. *Pattern Recognition Letters*, 28(16):2238–2249, 2007.
- [100] T. Theoharis, G. Passalis, G. Toderici, and I. A. Kakadiaris. Unified 3D face and ear recognition using wavelets on geometry images. *Pattern Recognition*, 41(3):796–804, 2008.
- [101] R. Gross and J. Shi. The CMU motion of body (MoBo) database. *Technical Report CMU-RI-TR-01-18*, Robotics Institute, Carnegie Mellon University, 2001.
- [102] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer. The humanID gait challenge problem: data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005.
- [103] M. Vondrak, L. Signal, and O. C. Jenkins. Physical simulation for probabilistic motion tracking. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1–8, 2008.
- [104] H. Yu, S. Qin, D. K. Wight, and J. Kang. Generation of 3D human models with different levels of detail through point-based simplification. In *Proceedings of International Conference on “Computer as a Tool”*, 1982–1986, 2007.
- [105] J. H. Nurre, J. Connor, E. A. Lewark, and J. S. Collier. On segmenting the three-dimensional scan data of a human body. *IEEE Transactions on Medical Imaging*, 19(8):787–797, 2000.

- [106] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, NY, United States, 2002.

Curriculum Vitae

KOICHIRO YAMAUCHI

Born: December 3, 1981, Nagoya, Japan

WORK ADDRESS

Graduate School of Science and Technology, Keio University

3-14-1 Hiyoshi, Kohoku-Ku, Yokohama, 223-8522, Japan

Phone: +81-45-563-1141

Email: yamauchi@hvrl.ics.keio.ac.jp

EDUCATION

- 04/06–present **Keio University**, Yokohama, Japan
Graduate School of Science and Technology: Ph.D. candidate
- 04/04–03/06 **Keio University**, Yokohama, Japan
Graduate School of Science and Technology: M.S. March 2006.
- 04/00–03/04 **Nagoya Institute of Technology**, Nagoya, Japan
Department of Electrical and Computer Engineering: B.Eng. March 2004.
- 04/97–03/00 **Tokai High School**, Nagoya, Japan

EXPERIENCE

- 04/08–10/09 **University of California, Riverside**, CA, United States
Visiting Scholar – Center for Research in Intelligent Systems
- 08/07–03/08 **Keio University**, Yokohama, Japan
Research Assistant – Global COE Program

HONORS AND AWARDS

- Best Paper Award, IEEE Computer Society Workshop on Biometrics, 2009

GRANTS

- Keio Leading-edge Laboratory of Science and Technology Research Grant, 2009
- Keio Leading-edge Laboratory of Science and Technology Research Grant, 2006

PUBLICATIONS

Journal Articles (Japanese)

1. 山内航一郎, 斎藤英雄, 佐藤幸男
“基準平面を用いたスリット光投影式レンジファインダの校正,”
電気学会論文誌 C, vol. 129-C, no. 6, pp. 1162–1168, June 2009.
2. 山内航一郎, 柴田進, 佐藤幸男
“多視点距離画像を用いた人体の三次元形状計測,”
電子情報通信学会論文誌 D-II, vol. J88-D-II, no. 8, pp. 1564–1572, August 2005.

Conference Papers (English)

1. Koichiro Yamauchi, Bir Bhanu, and Hideo Saito. Recognition of Walking Humans in 3D: Initial Results. *IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW2009)*, pp. 45–52, Miami, FL, USA, June 2009.
2. Koichiro Yamauchi, Hideo Saito, and Yukio Sato. Calibration of a Structured Light System by Observing Planar Object from Unknown Viewpoints. *The 19th International Conference on Pattern Recognition (ICPR2008)*, pp. 1–4, Tampa, FL, USA, December 2008.
3. Kenta Fujiwara, Koichiro Yamauchi, and Yukio Sato. An Automatic Alignment Technique for Multiple Rangefinders. *IS&T/SPIE 19th Annual Symposium Electronic Imaging (EI2008)*, SPIE 6805, pp. 1–8, San Jose, CA, USA, January 2008.
4. Koichiro Yamauchi, Hideto Kameshima, Hideo Saito, and Yukio Sato. 3D Reconstruction of a Human Body from Multiple Viewpoints. *The 2nd Pacific-Rim Symposium on Image and Video Technology (PSIVT2007)*, LNCS 4872, pp. 439–448, Santiago, Chile, December 2007.

5. Koichiro Yamauchi and Yukio Sato. 3D Human Body Measurement by Multiple Range Images. *The 18th International Conference on Pattern Recognition (ICPR2006)*, vol. 4, pp. 833–836, Hong Kong, China, August 2006.

Conference Papers (Japanese)

1. 山内航一郎, 佐藤幸男
“平面治具による距離計測プロジェクタ・カメラシステムの校正,”
第10回 画像の認識・理解シンポジウム (MIRU2007), pp. 690–695, July 2007.
2. 藤原健太, 山内航一郎, 佐藤幸男
“多視点距離画像の自動位置合わせ,”
第10回 画像の認識・理解シンポジウム (MIRU2007), pp. 684–689, July 2007.
3. 藤原健太, 山内航一郎, 佐藤幸男
“特異パターンの符号化による多視点レンジファインダのキャリブレーション,”
第13回 画像センシングシンポジウム (SSII07), IN3-14, pp. 1–4, June 2007.
4. 藤原健太, 山内航一郎, 佐藤幸男
“De Bruijn Sequence パターンに基づく平面校正治具による多視点距離画像の統合,”
電子情報通信学会総合大会講演論文集, D-12-27, pp. 143, March 2007.
5. 山内航一郎, 西尾裕志, 平田喜之, 佐藤幸男
“3D カメラの簡便な校正法,”
電子情報通信学会総合大会講演論文集, D-12-2, pp. 118, March 2007.
6. 山内航一郎, 佐藤幸男
“基準平面を用いたレンジファインダの校正,”
第9回 画像の認識・理解シンポジウム (MIRU2006), pp. 1018–1023, July 2006.
7. 山内航一郎, 佐藤幸男
“レンジファインダの幾何モデルと校正手法,”
第12回 画像センシングシンポジウム (SSII06), pp. 194–198, June 2006.
8. 山内航一郎, 佐藤幸男
“レンジファインダモデルと校正手法,”
電子情報通信学会総合大会講演論文集, D-12-60, pp. 192, March 2006.

9. 泰丘良玄, 山内航一郎, 石川尋代, 佐藤幸男
“3次元人体点群データのための表面復元,”
電子情報通信学会総合大会講演論文集, D-12-137, pp. 287, March 2005.
10. 山内航一郎, 柴田進, 二川貴志, 佐藤幸男
“人体の3次元形状計測システムの開発,”
第7回 画像の認識・理解シンポジウム (MIRU2004), vol. 1, pp. 589–594, July 2004.
11. 柴田進, 山内航一郎, 西尾裕志, 二川貴志, 佐藤幸男
“人体形状計測装置の開発,”
第10回 画像センシングシンポジウム (SSII04), pp. 253–258, June 2004.
12. 山内航一郎, 佐藤幸男
“自由な配置が可能な3次元人体計測システム,”
電子情報通信学会総合大会講演論文集, D-12-149, pp. 315, March 2004.
13. 柴田進, 二川貴志, 山内航一郎, 西尾裕志, 佐藤幸男
“3次元人体形状計測システムの試作,”
計測自動制御学会パターン計測部会研究会, vol. 66, pp. 25–30, March 2004.
14. 西尾裕志, 山内航一郎, 佐藤幸男
“人体計測データからのパッチモデル生成,”
情報処理学会研究報告 (2003-CVIM-143), vol. 2004, no. 26, pp. 1–8, March 2004.

Book Contributions (English)

1. Koichiro Yamauchi, Hideo Saito and Yukio Sato. Calibration of a Structured Light System Using Planar Objects. In P. Y. Yin, editor. *Pattern Recognition*. IN-TECH, Vienna, Austria, 2009.

Book Contributions (Japanese)

1. 山内航一郎, Bir Bhanu, 斎藤英雄
“三次元歩行動作認識への挑戦,”
画像ラボ, vol. 20, no. 12, pp. 52–56, December 2009.
2. 山内航一郎, 佐藤幸男
“3Dカメラを用いた人体の3次元形状計測,”
映像情報インダストリアル, vol. 38, no. 9, pp. 14–18, August 2006.

3. 山内航一郎, 柴田進, 佐藤幸男
“多視点距離画像による人体の3次元形状計測,”
画像ラボ, vol. 16, no. 10, pp. 61–65, October 2005.

