A Study on High-Speed Three -Dimensional Shape Inspection Based on Self-Projection Method (自己投影法に基づく 高速三次元形状検査の研究)

by

Hao Gao

Graduate School of Engineering
Hiroshima University
January, 2015

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

1.1.1 High-speed Vision

Conventional vision systems with standard video signals (NTSC 30 fps / PAL 25 fps) have existed for many years and have been applied to a variety of fields such as multimedia, traffic system, biomedical, three-dimensional reconstruction, industrial vision measurement/inspection, and so on. Due to the low sampling rate, conventional vision systems are not suitable for high-speed phenomena, for instance, factory automation high-speed production line, high-speed motion, high-speed target tracking, hyper human manipulation and visual servoing. Therefore, many high-speed vision systems that operate at 1000 fps or more have been developed to overcome the restrictions imposed by standard video signals.

The bottleneck of the image information transmission from photo-detectors (PD) to processing elements (PE) constrained the sampling rate of conventional vision systems. To overcome this limitation, vision chips have been developed and execute real-time processes at a rate of 1000 fps or more by integrating sensors and processors compactly, shown in Figures 1.1. Bernard et al. proposed an on-chip array of bare Boolean processors with halftoning facilities and developed a 65±76 Boolean retina on a 50 mm^2 CMOS 2 μm circuit for the imager of an artificia retina [1]. Eklund et al. verifie the near-sensor image processing (NSIP) concept, which describes a method to implement a two-dimensional (2-D) image sensor array with processing capacity in every pixel, and

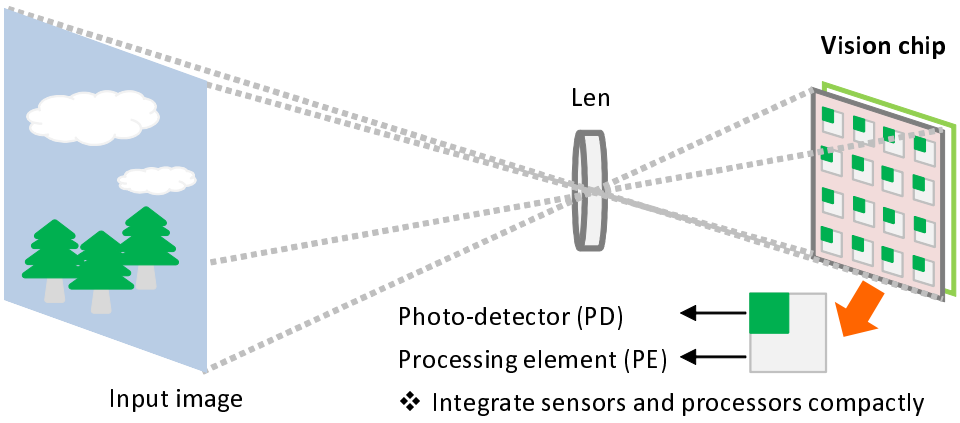


Figure 1.1: Digital vision chip

have fabricated and measured a 32±32 pixels NSIP [2]. Ishikawa et al. have developed a COMS vision chip for 1ms image processing and proposed the S^3PE(simple and smart sensory processing elements) vision chip architecture with each PE connected to a PD without scanning circuits [3, 4]. Komuro et al. proposed a dynamically reconfigurabl single-instruction multiple-data (SIMD) processor for a vision chip and developed a prototype vision chip based on their proposed architecture, which has 64±64 pixels in a 5.4 mm 5.4 mm area fabricated using the 0.35 μm TLM CMOS process [5]. Ishii et al. proposed a new vision chip architecture specialized for target tracking and recognition, and developed a prototype vision chip using 0.35 μm CMOS DLP/TLM(3LM) process [6].

Recently, several research groups have attempted to implement high-speed image processing using circuits on an FPGA (Field Programmable Gate Array) board, which is directly connected to a high-speed digital camera head, shown in Figures 1.2. Hirai et al. developed an fl xibility FPGA-based vision system using the logic circuit to implement the image algorithm [7]. Watanabe et al. developed a high-speed vision system for real-time shape measurement of a moving/deforming object at a rate of 955 fps (256±256 resolution) [8]. Ishii et al. developed a high-resolution high-speed vision platform, H^3(Hiroshima Hyper Human) Vision, which can simultaneously process a 1024±1024 pixels image at 1000 fps and a 256±256 pixels image at 10000 fps by implementing image processing algorithms as hardware logic on a dedicated FPGA board [9]. In the latest

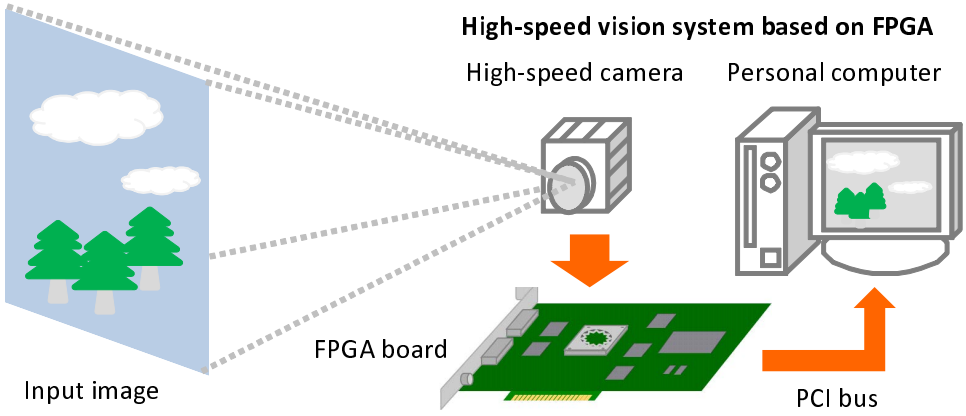


Figure 1.2: Highspeed vision system based on FPGA

two years, Ishii et al. developed a high-speed vision system called IDP Express, which can execute real-time image processing at a rate from 2000 fps (512±512 resolution) to 10000 fps (512±96 resolution), and high frame rate video recording simultaneously [10].

At present, high-speed vision systems can be used as robot sensors at hundreds of hertz or more; several applications of these systems have been also reported, shown in Figures 1.3. Ishii et al. proposed a simple algorithm for high-speed target tracking using the feature of high speed vision, and realized target tracking on the 1 ms visual feedback system [11]. Nakabo et al. developed a 1 ms vision system, which has a 128±128 PD array and an all parallel processor array connected to each other in a column parallel architecture, for 1 ms cycle-time for visual servoing and applied it to high-speed target tracking [12]. Nakamura applied high-speed vision system to virtual stillness for beating heart surgery [13]. Nakabo et al. developed a 3D target tracking/grasping system, which are composed of a 1ms feedback rate using two high-speed vision systems called column parallel vision (CPV) systems and a robot hand arm [14]. Namiki et al. developed a high-speed three-fingere robotic hand controlled by a massively parallel vision system (CPV system) for robot catching [15]. Shiokata et al. proposed a strategy called ”dynamic holding” and developed a experimental robot dribbling using a high-speed multi-fingere hand and a high-speed vision system [16]. Mizusawa et al. used high-speed vision servoing to tweezers type tool manipulation by a three-finge robot hand [17]. Nie et al. developed a

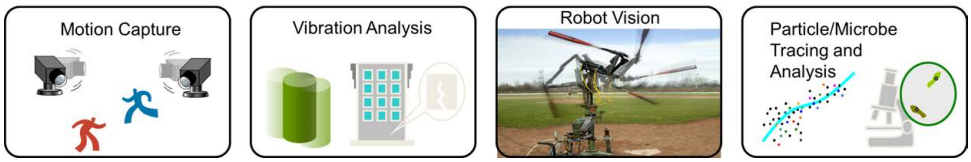


Figure 1.3: High-speed vision application field

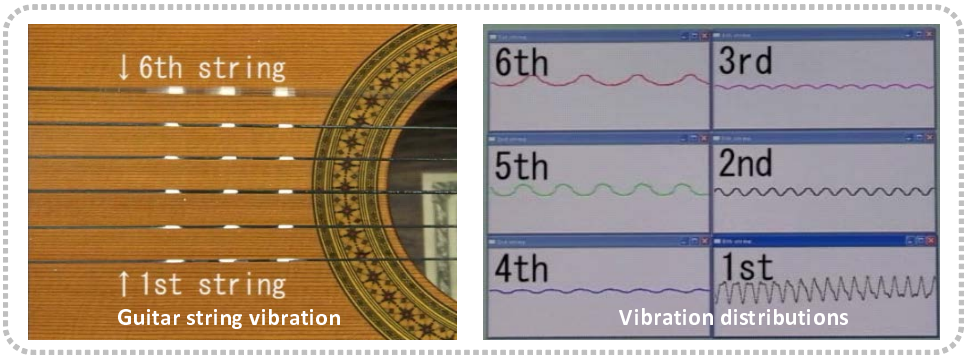


Figure 1.4: HFR video-based guitar string vibration measurement

real-time scratching behavior quantificatio system for laboratory mice using high-speed vision [18]. Wang et al. developed an intelligent High-Frame-Rate video logging system to automatically detect high-speed unpredictable behavior and record video comprising images with dimensions of 512±512 pixels at a rate of 1000 fps [19]. Yang et al. proposed a ”modal radar” algorithm as a structural damage analyses methodology for modal testing that can localizeand accurately quantify structual damage which is difficult to detect by appreance-based visual inspection of a single image [20]. Gu et al. proposed a 2000 fps multi-object feature extraction system based on FPGA implementation of the cell-based labeling algorithm, which is suitable for hardware implementation and only few memory is required for multi-object feature extraction [21]. With the development of the sampling rate and resolution, such high-speed vision systems can also observe the vibration distribution of an object excited at dozens or hundreds of hertz, which is too fast for the naked eye and standard NTSC cameras. Figures 1.4 shows the High-frame-rate (HFR) video-based guitar string vibration measurement with 10000fps sampling frequency (512±96 resolution) at auditory sensation level.

1.1.2 Vision-Based 3D Shape Measurement

Three-dimensional measurement constitutes an important topic in computer vision, having different applications such as range sensing, industrial inspection of manufactured parts, reverse engineering (digitization of complex,free-form surfaces), object recognition, 3-D map building,biometrics,clothing design and others. The developed solutions are traditionally categorized into contact and non-contact techniques. Contact measurement techniques have been used for a long time in reverse engineering and industrial inspections. The main problems of contact techniques are their slow performance and the high cost of using mechanically calibrated passive arms [22]. Besides, the fact of touching the object is not feasible for many applications. Non-contact techniques were developed to cope with these problems, and have been widely studied. Non-contact techniques can be classifie into two different categories: active and passive. In passive approaches, the scene is firs imaged by video cameras from two or more points of view and correspondences between the images are found. It is important to mention that the cameras have to be previously calibrated [23]. The main problem experimented when using this approach is as parse reconstruction since density is directly related to the texture of the object.

This complicates the process of findin correspondences in the presence of texture less surfaces [24]. Therefore, passive reconstruction is rather limited to reconstruct dense 3-D surfaces, due to the problem of findin correspondences [25]. Methods based on structured light (active techniques) came to cope with this issue, creating correspondences and giving specifi codewords to every unitary position in the image. In this approach one of the cameras is substituted by an active device (a projector), which projects a structured light pattern onto the scene. This active device is modeled as an inverse camera, being the calibration step a similar procedure to the one used in a classical stereo-vision system [26]. The projected pattern imposes the illusion of texture on to an object, increasing the number of correspondences [27]. Therefore, surface reconstruction is possible when looking for differences between projected and recorded patterns.

1.2 Related Works

1.2.1 Light-Section Method

The light-section method [28, 29] is a well-known optical method that can accurately obtain 3-D shape information using triangulation by projecting a light pattern on the measurement space. The typical measurement range is ◯5 to ◯250 mm, and accuracy is about 1 part in 10,000 and measurement frequency of 40 Hz or higher [30, 31]. A charged couple device (CCD), or a position sensitive detector (PSD) is widely used digitize the point laser image. For a PSD, the measurement accuracy is mainly dependent on the accuracy of the image on the PSD. The beam spot refection and the stray light will also affect the measurement accuracy. Idesawa [32] developed some methods to improve the accuracy of the PSD by using a high accuracy kaleidoscopic mirror tunnel position sensing technique (KM-PSM) and a hybrid type of position sensitive detector (R-HPSD). CCD based sensors can avoid the beam spot reflectio and stray light effects and provide more accuracy because of the single pixel resolution.

Many 3-D shape inspection systems based on the light-section method have been developed. However, most of them use image sensors that follow standard video signals (NTSC at 30 fps or PAL at 25 fps), the spatial resolution and frame rate of the image sensor strongly restrict their computation speed, measurement range, and precision in acquiring the 3-D shape. In the light-section method, a slit-line light pattern is projected on the object and an image of the object is obtained by a camera installed in a direction different from the projection direction of the light pattern. The 3-D shape is then measured by triangulation on the basis of the projection pattern observed in the image. The light-section method has robustness that does not strongly depend on surface texture patterns and it can calculate 3-D shapes accurately by using a high directive light pattern projector such as a slit laser; it has been applied to practical 3-D shape measurement systems in practical settings.

To speed up the 3-D shape measurement based on the light-section method, many improved algorithms and systems have been developed. Lee et al. [34] developed a scan planning algorithm to generate the minimum time path that contains all the directions and

paths for a laser scanner. Bernstein et al. [35] developed a bi-sensorial measurement system consisting of a light-section system and a shadow-detection system that can remove unobservable image regions for efficient computation in measuring concave 3-D shapes. For recent improvement of the integration technology, many high-speed vision systems such as vision chips [36, 37, 38] and FPGA-based vision systems [39, 40, 9] have been developed for real-time video processing at frame rates of 1000 fps or more. In particular, many high-speed smart sensors [41, 42, 43, 44] for 3D shape measurement have been developed, because 3-D profil calculation with the light-section method can actualize with simple processing that is suited for integration. Yamamoto et al. [45] developed a high-speed vision system to measure 3-D shapes of planar objects at 1000 fps in a wider measurement range than the spatial resolution of the image sensor by setting a multisided mirror between the image sensor and the measurement object.

1.2.2 Coded Structured Light Method

Shape reconstruction using coded structured light is considered one of the most reliable techniques to recover object surfaces. By projecting certain type of patterns, the correspondence of the images can be easily identified and depth information can be retrieved by a simple triangulation technique. This is one advantage that structured light has over stereo vision, in which the fundamentally difficult correspondence problem must be solved.

In the coded structured light method [48], multiple bright-and-dark light patterns are projected from a projector onto the measured object, and images of the object are obtained by a camera installed in a direction different from the projection directions of the light patterns. When n-bit light patterns are projected, n-bit space code can be calculated at each pixel by checking its brightness in the captured images; n-bit space code indicates the beam direction of projected light patterns. For all the pixels in the image, the depth information can then be measured by triangulation on the basis of the relationship between the pixel location and its space code.

Posdamer's method, in which bright-and-dark light patterns are projected with a

pure binary code, may often have serious errors in encoding them even when the amount of noise is small, because there are brightness boundaries of bright-and-dark light patterns with a pure binary code at the same positions. To solve this problem, Inokuchi et al. [49] proposed a bright-and-dark pattern projection method for minimizing encoding errors on boundaries. Bergmann [50] proposed an improved 3-D shape measurement method that combines the gray code pattern projection method and a phase-shifting method, but the number of projection patterns increases. Zhang et al. [51] developed a fast phase-shifting system for binary structured light patterns projected at high frame rates. In their system, multiple sinusoidal fringe patterns are generated by properly defocusing the projector, and they are used in a phase-shifting method to obtain 3-D shape information. For reduction of the number of projection patterns, Caspi et al. [53] proposed a color-patterned gray code pattern projection method. Most of these methods can obtain high-resolution depth images of static objects by calculating distance information at every pixel. However, synchronization errors between multiple light patterns at different timings restrict 3-D measurement accuracy when moving objects are observed, and in 3-D measurement using standard video signals at dozens of frames per second, accuracy decreases significantl .

For depth image acquisition by one-shot projection, several structured light methods using spatially coded patterns have been proposed. The methods proposed are suitable for 3-D measurement of moving objects without synchronization errors between multiple light patterns at different timings. Pages et al. [54] proposed a color stripe pattern projection method based on de Bruijn sequences as a robust coded pattern projection method that combines the high resolution of classic striped patterns with the accuracy of multi-slit patterns. Ito et al. [55] proposed a three-level checkerboard pattern projection method that projects a grid-like light pattern onto the object to be observed. Its cell has a gray level chosen from three intensity values. Guan et al. [56] proposed a continuous-tone phase measuring profilometr (PMP) method that modulates and combines multiple PMP patterns into a single composite pattern for one-shot projection. Griffin et al. [57] proposed a color-circle-array pattern projection method that uses an alphabet of four symbols for color circles. Sakashita et al. [58] obtained both the 3-D shape and the texture of the object to be observed by capturing a one-shot color-patterned grid pattern with a multi-band

camera and an infrared projector. Mircosoft Kinect [59] can acquire depth images in real time at 30 fps by capturing an infrared spatially coded light pattern. As a result of one-shot projection, most spatially coded methods are suitable for 3-D image acquisition of dynamically changing environments, but their accuracies in 3-D measurement are much worse than the spatial resolutions of image sensors because most of them assume local surface smoothness of the object surface in spatial encoding with one-shot projection.

1.3 Outline of Thesis

The thesis is structured as followed. Chapter 1 will introduce the development and applications of high-speed vision, and vision based 3-D shape measurement. The related works of light section method and coded structured light method for vision based 3-D shape measurement will be described in detail.

Chapter 2 will discuss the limitations of related works and the research purpose of this study is to overcome these limitations. Then, we propose the concept of self-projection for light-section method and structured light method which can reduce the processing data quantity for real-time processing during online shape inspection to improve the inspection speed.

Chapter 3 will describe the algorithm of self-projected light-section method in detail and the performance of this method is verifie by implementation on a real-time, high-frame-rate 3-D shape measurement system, which consists of a high-speed vision platform and an LCD projector.

Chapter 4 will introduce a 3-D shape inspection system based on self-projection structured light method which is implemented on a GPU-based real-time 3-D scanner which can output 3-D images of 512±512 pixel at 500 fps by using high-frame-rate vision platform and a DMD high-speed projector. This system can measure the differential shape from the reference object with less patterns projection without affecting the accuracy level even the depth difference is very large. Several experiments are executed to test the performance of this system.

Chapter 5 will summarize the contributions of this study and discuss the future

work. In future work, we want to introduce the real-time 3-D shape measurement system to various application field considering the high-speed and real-time processing property, such as human interfacing depend on the fast movement of human finger or eyeballs which is difficult to captured by normal speed cameras, or robot controlling which need to decide the movement by real-time position information in the factory.

Chapter 2

Concept of Self-Projection Method

2.1 The Limitations of Related Works and Research Purpose

With the development of computer vision, many 3-D shape measurement systems based on various optical measurement methods have been developed. However, the processing speed of these systems is limited by the speed of standard video signals (e.g, NTSC 30 fps or PAL 25 fps) that are designed based on the characteristics of the human eyes. And it is also impossible for CPU to process so large size of data in real-time calculation with traditional optical 3-D measurement methods. On the other hand, the high-speed production lines are more and more applied in the factories to raise the production efficiency during these years. However, the production inspection speed was not so satisfie while the speed of production lines have been improved to a desirable level nowadays. It is a very important problem to be solved for accelerating the development speed of industrial process.

Most of the shape inspection system use the non-contact vision based 3-D measurement techniques which have a fast performance and the low cost for not using mechanically calibrated passive arms. Among various non-contact vision based 3-D measurement methods, light-section method and coded structured light method are usually applied to practical 3-D shape measurement system in practical settings. Most of the shape acquisition systems based on laser scanners obtain the 3-D shape by scanning the object with a laser plane and detecting the projected line in the camera image for triangulating all

the illuminated points. The advantage of these scanners is the large resolution and accuracy obtained leading to high quality 3-D surface reconstruction. But correspondingly, it limits the application field of the light-section method due to the static objects only measurements. A large number of images must be acquired for the high accuracy 3-D surface reconstruction, and all the images have to be captured in the same situation when the measured objects are static. Substantial quantity of data are processed based on the images captured for the measured objects which also limit the processing speed for real-time 3-D surface reconstruction.

The most difficult problem in the coded structured light method is the determination of homologous points in a series projection images, for example, determining which projective points represent the same three dimensional object point. This problem is known as the correspondence problem, which is the main limitation of structured light method, since once it is solved the rest has been already formalized[60]. To avoid the corresponding problem, most of the three-dimensional shape measurement system based on the structured light method are applied for static objects, because of the images sensor they used which follow the standard video signals(e.g., NTSC 30 fps or PAL 25 fps). The low frame rate of standard video signals makes the corresponding error among the projection images become seriously when moving objects are observed. To resolve this problem, many coded structured light method only using one projection patten have been developed. However, the accuracy of these methods in three-dimensional measurement are much worse than the spatial resolutions of images sensors, because most of them assume a local smoothness of object surface in spatial encoding with one-shot projection. To reduce the corresponding errors in the structured light method using multiple projection patterns for moving objects, we can provide the resolutions from two different perspectives. One is decreasing the time interval between each between adjacent two frames, which we can achieve by using high-frame-rate camera in the three-dimensional measurement system. And another one is reducing the projection number for three-dimensional measurement without affecting the accuracy level of the measured result, for what we developed a novel projection method for the high-speed structured light shape inspection system.

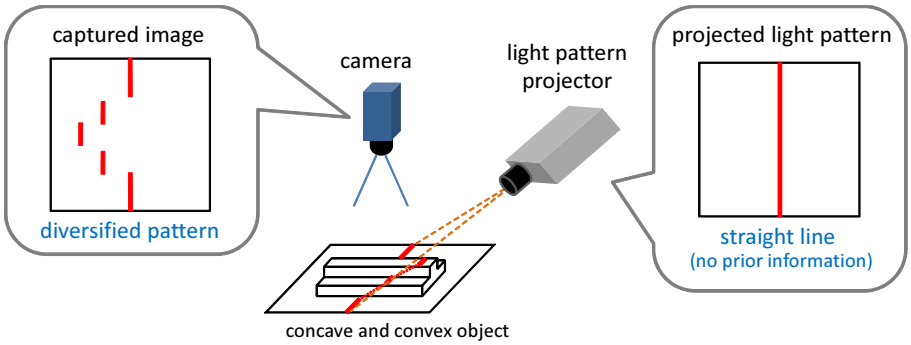
2.2 Concept of Self-Projected Light-Section Method

In these years, a lot of sensing systems that can greatly accelerate the speed of 3-D shape measurements have been developed. However, most of them still involve inefficient processes to obtain a 3-D section profil from a redundant slit image; they require a slit image of 2^n width in the case of a 3-D shape measurement with n-bit height precision. This redundancy becomes larger when we measure 3-D shapes of concave and convex objects such as cylinders, because the height information is calculated as an absolute distance from a level surface, which is initially given with a straight-light projection of the slit light. In factory lines, there are strong demands for product quality control to rapidly inspect the 3-D shapes of concave and convex products by comparisons with a nondefective product. Thus, for fast 3-D shape inspection of concave and convex products it is important to reduce the pixel dimensions of the image to be processed without degradation of accuracy.

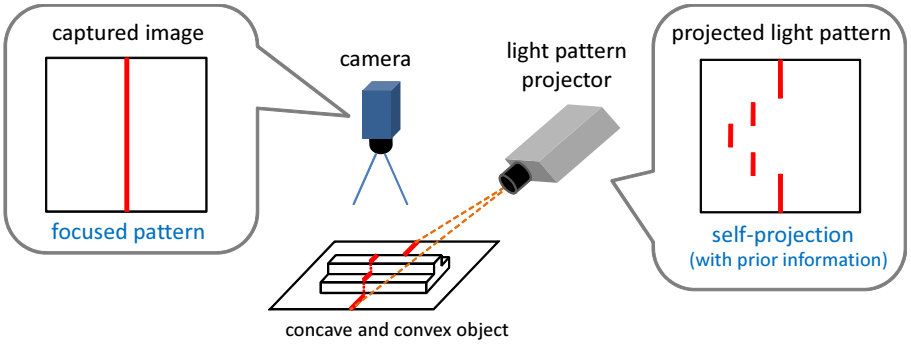
For this reason, we propose an improved light-section method (hereinafter referred to as the "self-projected light-section method") that can reduce the pixel dimensions of an image to be processed for obtaining the difference between the 3D shape of a measurement object and that of a reference object. Figure 2.1 shows the concept of our self-projected light-section method, compared with the conventional light-section method.

In the conventional method, a straight-line light pattern is projected on a measurement object at a fi ed angle with a level plane regardless of the concavity or convexity of the shape; the straight-line light pattern has no prior information. A camera installed in a different direction can capture the straight-line light pattern projected on the object as an image of the spatially diversifie brightness pattern; this diversificatio becomes larger as the shape of the object becomes more concave or convex, because the conventional method calculates absolute height information as a distance from a level plane.

Instead of a straight-line light pattern projection, our self-projected light-section method uses a curved-line light pattern for projection on an object to be inspected; we assume that a 3D shape of a reference object is given as prior information. The curved-line light pattern is initially generated as a self-projected light pattern using the 3D shape of the



(a) conventional light-section method



(b) self-projected light-section method

Figure 2.1: Self-projected light-section method.

reference object. The angles formed by beams of light that consist of the self-projected light pattern and a level plane are spatially diversifie in proportion to its concavity or convexity. In the mean time, a straight-line slit image in a camera view can be captured as a reduced image for fast triangulation computation when the 3D shape is matched with that of the reference object. This characteristic is maintained even when the shape of the object is concave and convex. The self-projected light pattern works as a template for 3D shape inspection, and the differential shape object can be easily picked up by checking if there is deviation on the captured straight-line. And because we introduce a self-projected light pattern projection with prior information, our method can directly calculate differential height information from the 3D shape of the reference object instead of calculating absolute height information from a level plane in the conventional method. Thus, even small differences in the 3D shape between a measurement object and a reference object

can be enhanced in a narrower image region when their shapes are strongly concave or convex.

Our self-projected light-section method acts as an inverse process of the conventional light-section method, assuming that a reference object is initially given. In this study, we consider the following two subprocesses in the self-projected light-section method for inspecting the 3D shapes of cylinder-like objects:

(1) Self-projected light pattern generation:

The 3D shape of a reference object is firs embedded as a curved-line light pattern for self-portrait projection to reduce the pixel dimensions of the image to be processed.

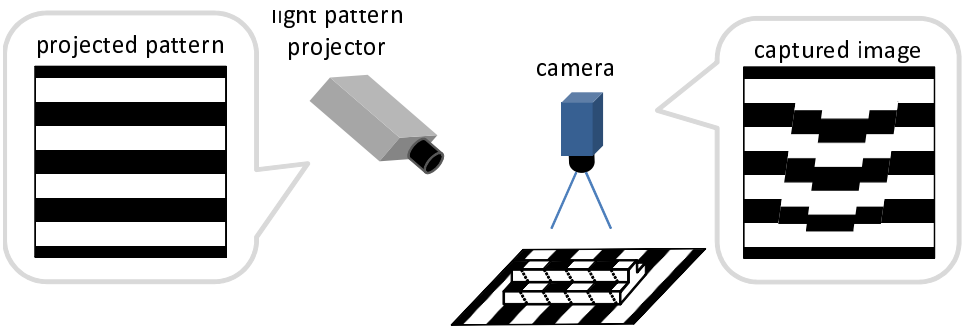
(2) Differential 3D shape measurement:

The height deviation from a reference object is efficiently obtained by processing a reduced image region in real time at a high frame rate without degradation of accuracy.

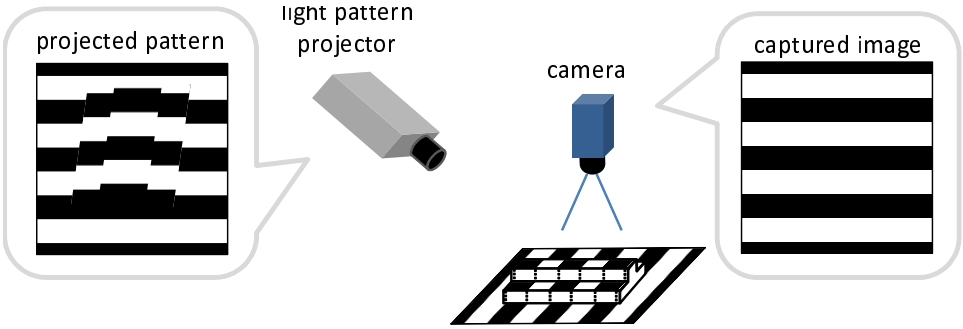
Based on our self-projected light-section method we developed a projector-camera measurement system that can simultaneously obtain 3-D shapes of cylindrical objects moving in one direction. This system consists of a high-speed vision platform which can capture and process images in real time at 10000 fps, an LCD projector and an electronic linear stage to convey an object in one direction. The high-speed vision platform applied in our system has an FPGA image-processing board that can be coded for image-processing function in hardware and a column-moment calculation circuit module that can calculate the centroids for 3-D shape measurement was implemented as hardware logic to improve the processing speed.

2.3 Concept of Self-Projected Structured-Light Method

In the self-projected light section method, the pixel dimensions of the image to be processed can be reduced by projecting a curved light pattern generated by a reference 3-D shape. To expand the idea of self-portrait projection in the self-projected light section method for fast and real-time 3-D profil acquisition, we propose an improved structured light method (hereinafter referred to as the “self-projected structured light method”) that can reduce the number of projections needed to obtain the 3-D shape of a measured object



(a) conventional structured-light method



(b) self-projected structured-light method

Figure 2.2: Self-projected structured-light method.

without affecting the 3-D measurement accuracy, assuming that a reference 3-D shape is initially known. Fig. 2.2 illustrates our self-projected structured light method compared with the conventional structured light method.

In the conventional method, bright-and-dark straight-stripe patterns are projected onto a measured object regardless of the 3-D shape of the measured object, and the straight-stripe patterns have no prior information. A camera can capture the straight-stripe patterns projected onto the object as an image sequence containing the curved-stripe patterns deformed by the 3-D shape of the measured object. This deformation increases when the 3-D shape of the measured object is more concave or convex, and a larger bit number of structured light code projection is required for accurate depth measurement in a 3-D scene with large height differences. This is because the bit number of the structured light code in the conventional method determines the measurement accuracy and

measurable range of the absolute depth information in 3-D measurement.

In contrast, our self-projected structured light method projects multiple self-projected curved-stripe patterns onto the measured object, rather than straight-stripe patterns, assuming that a reference 3-D shape is provided initially and that the self-projected patterns are generated as curved-stripe patterns calculated using prior information. When the 3-D shape of the measured object matches the reference 3-D shape, the straight-stripe pattern images in the camera view are captured as curved-stripe patterns projected onto the object. This property is maintained when the 3-D shape is concave or convex with a large height difference. In our self-projected structured light method, the self-projected patterns can serve as templates for 3-D shape inspection and different 3-D shapes can be readily identifie by checking for deformations in the captured straight-stripe-like patterns because the degree of deformation is expressed as differential depth information based on the reference 3-D shape, instead of absolute depth information based on a level plane, as observed in the conventional method. As a result of the self-projected patterns, our method can obtain the depth image of a measured object even in a complex 3-D background scene by projecting fewer light patterns without narrowing the measurable range or affecting the accuracy in 3-D measurement, in contrast to the conventional method. Thus, even small local differences between the 3-D shapes of a measured object and a reference object can be expressed using fewer bits of structured light code if the reference 3-D shape contains large height differences.

In our self-projected structured light method, the following two processes are considered in a way that is opposite to that of the conventional structured light method.

(1) Generation of self-projected patterns

Initially, a reference 3-D shape is embedded as deformed curved-stripes patterns for self-portrait projection to reduce the number of projections in 3-D measurement.

(2) Depth calculation

The depth information of the measured object is calculated by processing fewer projection images without degrading the accuracy and narrowing the depth measurement range in the structured light method.

To apply the proposed self-projection structured light method on moving objects, we developed a structured light system that can output depth images of 512 ± 512 pixels in real time at 500 fps [52], and accelerated it by installing a GPU board for parallel processing of a gray-code structured light method [49] using eight pairs of positive and negative patterns with an 8-bit gray code, which we projected at 1000 fps from a high-speed projector. Compared with standard videos at dozens of frames per second, capturing eight pairs of light patterns projected at a high frame rate significantl reduces synchronization errors caused by projection with different timings. By implementing our self-projection structured light method, the synchronization error in 3-D measurement determined by the number of projections and their frame interval can be reduced and we can get more accurate depth images.

Chapter 3

Self-Projected Light-Section Method

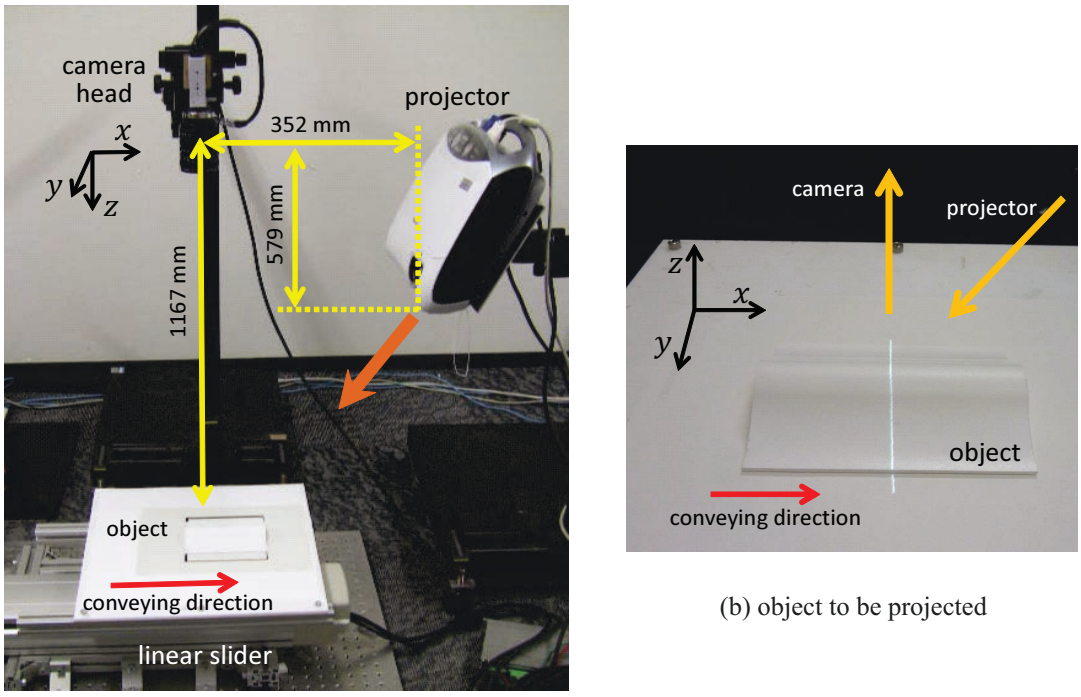
3.1 Introduction

In this chapter, we will introduce the new light-section method that can accurately obtain differential shape from a given reference 3-D shape at a high frame rate by projecting a "self-projected light pattern," that is, a curved-line light pattern generated by a 3-D shape of a reference object; it is assumed that a reference object to be inspected is initially given. A high-frame-rate 3-D shape measurement system based on our method was also developed, and its effectiveness was demonstrated by measuring the 3-D shapes of cylinder-shaped objects in real time at 10000 fps.

3.2 A Real-Time, High-Frame-Rate 3-D Shape Measurement System

Based on light-section method we developed a measurement system that can simultaneously obtain 3-D shapes of cylindrical objects moving in one direction. An overview is shown in Figure 3.1. This system consists of a high-speed vision platform, IDP Express [10], which can capture and process 8-bit gray-scale images of 96±512 pixels in real time at 10000 fps, an LCD projector (LP-XU55, Sanyo Inc., Japan), and an electronic linear stage (ISA-LXM-200, IAI Inc., Japan) to convey an object in one direction.

The optical center of the camera lens was the origin of the *xyz* coordinate system, and the optical center of the projector lens was at a height of z_0 = 579 mm. The distance



(a) overview

(b) object to be projected

Figure 3.1: 3-D shape measurement system.

between them in the x direction was $x_0 = 352$ mm. The focal lengths of the camera and projector lenses were $f_c = 50$ mm and $f_p = 20$ mm, respectively. The optical axis of the projector lens forms an angle of $\theta_0 = 20^{\leq}$ with respect to the z axis. The projector projects a light pattern of 512±512 pixels in a square of 121 ± 121 mm at 60 fps on a conveying plane at a height of $z = 1167$ mm. Because the projection frame rate is much slower than that of the high-speed vision platform, we propose a high-speed 3-D measurement system to measure objects that are invariant in one direction by projecting a permanent self-projection pattern that can be fi with all the cross sections. Using the IDP Express, the projected light pattern was simultaneously captured at 10000 fps in an image with 96±512 pixels; the area of the captured image matched that of the image pattern projected on the conveying plane.

The IDP Express is used as a high-speed image-processing system. It consists of a high-speed camera head for picturing a target, an FPGA image-processing board (IDP Express board) that can be coded for image-processing functions in hardware, and a per-

Table 3.1: Resource consumption of user-specifie FPGA.

Device Type	Xilinx XC3S5000-4FGG900
Slice	2,402/33,380 (7%)
Slice Flip Flop	2,758/66,560 (4%)
4-input LUT	1,640/66,560 (2%)
Bounded IOB	229/633 (36%)
Block RAM	20/104 (19%)
GCLK	2/8 (25%)

sonal computer (PC). We used a PC with the following specifications ASUSTeK P5E main board, Intel Pentium Xeon 2.67-GHz CPU, 3.25 GB of memory, Windows XP Professional OS (32-bit version), and 2±PCI-Express 2.0±16. The IDP Express board is a high-speed image-processing board designed specificall for the high-speed camera head, and a user-specifi FPGA (Xilinx XC3S500) has been loaded on it. In this study, a column-moment calculation circuit module that can calculate the centroids for 3-D shape measurement based on the light-section method for 96±512 pixel images was implemented as hardware logic on the IDP Express board. The details of this board are examined in [10].

The function implemented in the hardware calculates the 0th moment $M_0(\eta)$ and the 1st moment $M_1(\eta)$ on each column of 96 pixels for $\eta = 1, \times\times\times, 512$ as follows:

$$M_0(\eta) = \sum_{\xi=1}^{96} I(\xi, \eta), \quad M_1(\eta) = \sum_{\xi=1}^{96} \xi I(\xi, \eta), \tag{3.1}$$

$$C(\eta) = M_1(\eta)/M_0(\eta), \tag{3.2}$$

where (ξ, η) indicates the integer pixel position on the captured binary image $I(\xi, \eta)$ of 96±512 pixels. In the circuit module, a 96±512 pixel image is scanned with accumulation of $M_0(\eta)$ and $M_1(\eta)$ in units of 4 pixels from the upper left to lower right with a 151.2-MHz clock. The input image of 96±512 pixels and its 512 column-moments of $M_0(\eta)$ and $M_1(\eta)$ are outputted with a delay time of 1024 clocks (where 1 clock = 6.6 ns) via an FIFO memory to a standard memory on the PC at 10000 fps. Table 3.1 lists the resource consumption of the FPGA on the IDP Express board.

On our system, the subprocess of self-projected light pattern generation is executed offline by scanning a straight-line light pattern from left to right in the X_p direction on the projector image plane; this scanning is performed to fin the beam positions for the self-projected light pattern, and their corresponding beams are measured at $X_c = 0$ on the camera image plane when a reference object is observed. The subprocess of differential 3-D shape measurement was executed for 96±512 pixel images in real time at 10000 fps as described in Eqs. (3.10) and (3.13). It was accelerated by implementing the column-moment calculation circuit module; 512 beam positions (X_c, Y_c) measured on the camera image plane are obtained at 10000 fps. Here the beam position measured on each column of 96 pixels in the captured image was described using the column-centroid $C(\eta)$ as follows:

$$(X_c, Y_c) = (a(C(\eta) \quad 48.5), a(\eta \quad 256.5)). \tag{3.3}$$

where $a = 10\ \mu$m is the pixel size of the image sensor.

On our system, objects were conveyed at 1000 mm/s by a linear slider. On the conveying plane of $z = 1167$ mm, the measurement error was 0.10 mm in the x direction, 0.23 mm in the y direction, and 0.33–0.39 mm in the z direction. Here, the measurement error in the z direction is dependent on the beam angle θ from the projector.

3.3 Self-Projected Light-Section Method

3.3.1 Triangulation for a camera and projector system

The self-projected light-section method is considered for a conveying object on a camera and projector system as illustrated in Figure 3.2. Here the pin-hole camera model of perspective projection is assumed for a camera and a projector.

The xyz coordinate system is define as the world coordinate system; its origin O is located at the optical center of the camera lens. The x axis is the conveying direction of the object, the y axis is at a right angle in the conveying plane, and the z axis is the optical axis of the camera lens. The X_cY_c coordinate system on the camera image plane

is perpendicular to the optical axis. Its origin is located at the intersection point with the optical axis of the camera lens. The X_c and Y_c axes are parallel to the x and y axes, respectively. The X_cY_c plane is at a distance of f_c from the optical center. The projector is installed in a direction different from that of the camera; the optical center of the projector lens is $O'(x_0, 0, z_0)$. The optical axis of the projector lens forms an angle θ_0 with respect to the z axis. The X_pY_p coordinate system on the projector image plane is perpendicular to the optical axis of the projector lens. The X_pY_p plane is at a distance of f_p from the optical center of the projector lens. The Y_p axis is parallel to the y axis.

In most light-section methods, the 3-D position (x, y, z) on a measurement object is obtained by triangulation based on the following geometrical relationship between the beam position (X_p, Y_p) on the projector image plane and its corresponding beam position

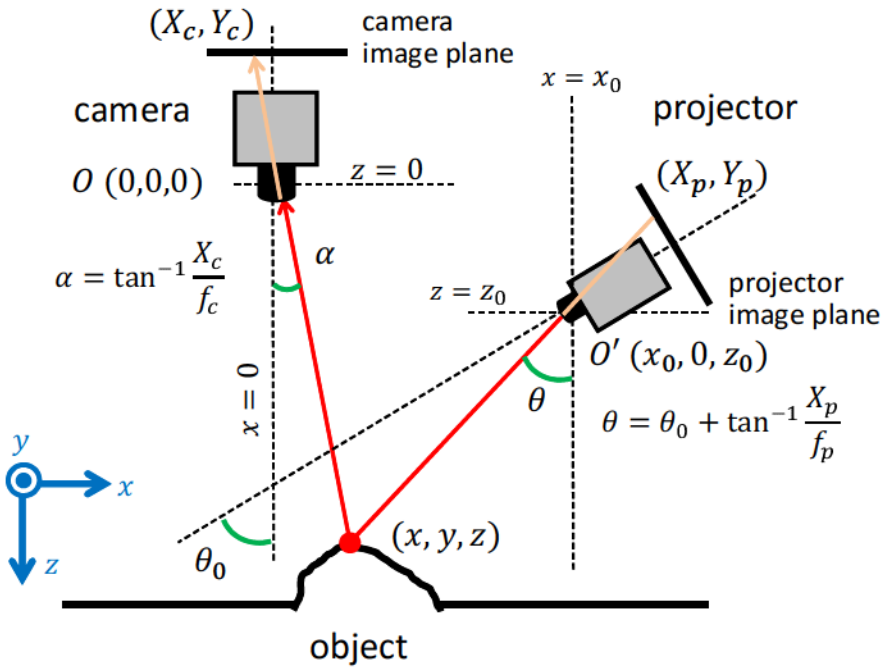


Figure 3.2: Geometry of a camera and projector system.

(X_c, Y_c) measured on the camera image plane:

$$(x, y, z) = \frac{x_0 + z_0 \tan\theta}{\tan\alpha + \tan\theta}(\tan\alpha, \tan\beta, 1) \tag{3.4}$$

$$= \frac{x_0 + z_0 \tan\theta}{\tan\alpha + \tan\theta}\left(\tan\alpha, \frac{x_0 \quad z_0 \tan\alpha}{x_0 + z_0 \tan\theta}\tan\phi, 1\right), \tag{3.5}$$

where the angles of α, β, θ, and ϕ are define using (X_c, Y_c) and (X_p, Y_p) as follows:

$$\alpha = \tan^{1}\frac{X_c}{f_c}, \quad \beta = \tan^{1}\frac{Y_c}{f_c}, \tag{3.6}$$

$$\theta = \theta_0 + \tan^{1}\frac{X_p}{f_p}, \quad \phi = \tan^{1}\frac{Y_p}{f_p}. \tag{3.7}$$

3.3.2 Self-projected light pattern generation

In the self-projected light-section method, the self-projected light pattern is initially generated on the projector image plane to project the light pattern at $X_c = 0$ on the camera image plane when a reference object is observed. When the light pattern is measured at $X_c = 0$ in the camera view, its intersected 3-D profil of the reference object is projected onto the $x = 0$ plane. In this case, the relationship between the beam position (X_p, Y_p) in the projector view and its corresponding beam position $(0, y_r, z_r)$ projected on the reference object can be described by using Eqs. (3.4)–(3.7) as follows:

$$(0, y_r, z_r) = \left(0, \frac{x_0 \tan\phi}{\tan\theta}, \frac{x_0}{\tan\theta} + z_0\right). \tag{3.8}$$

Thus, we can generate the beam positions (X_p, Y_p) for the self-projected light pattern on the projector image plane using the initially given reference 3-D profil $(0, y_r, z_r)$ as follows:

$$(X_p, Y_p) = \left(\frac{x_0 \quad (z_r \quad z_0)\tan\theta_0}{(z_r \quad z_0) + \tan\theta_0} f_p, \frac{y_r}{z_r \quad z_0} f_p\right). \tag{3.9}$$

In this study, the self-projected light pattern is obtained by scanning light beams in the direction of the X_p axis on the projector image plane, instead of obtaining the intersected 3-D profil of a reference object; the beam positions for the self-projected

light pattern are determined in the scan by checking whether or not their corresponding beams are measured on a line of $X_c = 0$ on the camera image plane. Here the beam angle $\theta = \theta(\beta)$ can be expressed as a function of the beam angle $\beta = \tan^{1}(Y_c/f_c)$ when the self-projected light pattern is projected on the reference object.

3.3.3 Differential 3-D shape measurement

When the self-projected light pattern, whose beam positions are obtained from Eq. (3.9), is projected on a measurement object, the 3-D position (x, y, z) on the object can be described using the angles of α and β between the measured beams and the camera image plane as follows:

$$(x, y, z) = \frac{x_0 z_r(\beta)}{(z_r(\beta) \quad z_0)\tan\alpha + x_0}(\tan\alpha, \tan\beta, 1) \tag{3.10}$$

$$= \frac{z_r(\beta)\tan\theta(\beta)}{\tan\alpha + \tan\theta(\beta)}(\tan\alpha, \tan\beta, 1), \tag{3.11}$$

where α and β are obtained from the beam positions (X_c, Y_c) measured on the camera image plane as described in Eq. (3.6), and $z_r(\beta) = x_0/\tan\theta(\beta) + z_0$ is initially given as the height information of the reference object as described in Eq. (3.8).

In the conventional light-section method, a straight-line light pattern with a beam angle θ fi ed by Eq. (3.4) is projected from a projector to a measurement object. In this case, the difference $(\Delta x, \Delta y, \Delta z) = (x, y, z) \quad (x_{r\in}, y_{r\in}, z_{r\in})$ between the 3-D shape of the measurement object and that of a reference object can be described as follows:

$$(\Delta x, \Delta y, \Delta z) = \frac{z_{r\in}(\beta)(\tan\alpha_{r\in}(\beta) \quad \tan\alpha)}{\tan\alpha + \tan\theta}(\tan\alpha, \tan\beta, 1), \tag{3.12}$$

where $(x_{r\in}, y_{r\in}, z_{r\in})$ is initially given as the 3-D position of the reference object, and $\alpha_{r\in}(\beta) = \tan^{1}(X_c/f_c)$ is given as its measured beam angle, corresponding to the beam position (X_c, Y_c) measured on the camera image plane when the straight-line light pattern is projected on the reference object. Equation (3.12) indicates that the conventional method requires a wider image region to inspect the 3-D shape of a concave or convex object even when its 3-D shape is similar to that of the reference object. This is because the

differential 3-D shape $(\Delta x, \Delta y, \Delta z)$ is obtained only by differencing its absolute 3-D shape from that of the reference object during post-processing in the conventional method.

However, when the 3-D shape of the measurement object is similar to that of the reference object, that is, α is small, the difference $(\Delta x, \Delta y, \Delta z) = (x, y, z) \quad (0, y_r, z_r)$ between the 3-D shape of the measurement object and that of the reference object can be approximately expressed in our self-projected light-section method as follows:

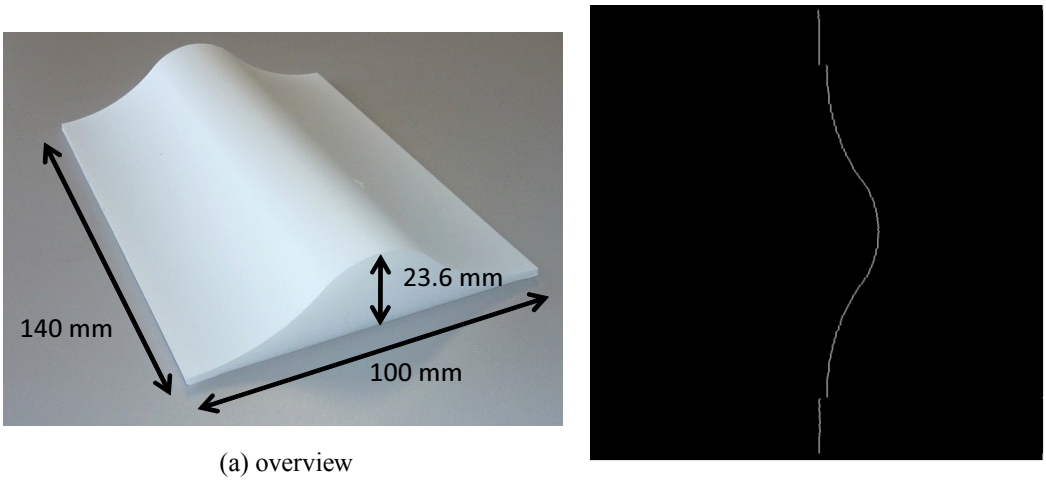
$$(\Delta x,\Delta y,\Delta z) = \frac{z_r(\beta)\tan\alpha}{\tan\alpha+\tan\theta(\beta)}(\tan\theta(\beta),\ \tan\beta,\ 1) \tag{3.13}$$

$$, \tan\alpha\left(z_r(\beta),\ \frac{z_r(\beta)\tan\beta}{\tan\theta(\beta)},\ \frac{z_r(\beta)}{\tan\theta(\beta)}\right) \tag{3.14}$$

This relationship indicates that the deviation in the 3-D shape of the measurement object from that of the reference object is proportional to X_c when α is small; X_c corresponds to the deviation of the measured beam point from a line of $X_c = 0$ on the camera image plane. In our self-projected light-section method, the differential 3-D shape can be directly obtained by using the beam positions (X_c, Y_c) measured on the camera image plane without any post-processing for comparison with the 3-D shape of the reference object. This characteristic of our self-projected light-section method enables fast and accurate 3-D shape measurement in a narrower image region even when measurement objects are concave or convex.

3.4 Experiments

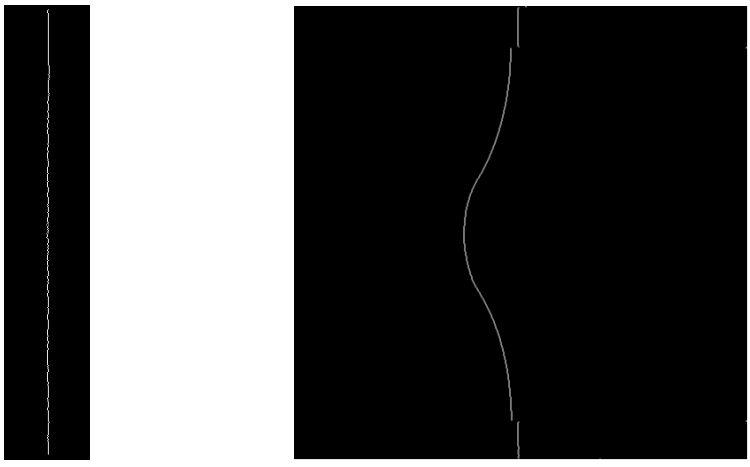
To verify the performance of our system, the 3-D shapes of objects were measured by using the self-projected light-section method. Figure 3.3(a) shows the mountain-shaped cylinder to be measured. Its height and width were 23.6 and 100 mm, respectively. When the mountain-shaped cylinder was used as a reference object, its self-projected light pattern of 512±512 pixels on the projector was generated as shown in Figure 3.3(b). Figure 3.4 shows the 96±512 pixel image captured on the camera in our method, compared with that in the conventional light-section method. It can be seen that the image width to obtain the whole shape of the mountain-shaped cylinder can be reduced to within 3 pix-



(a) overview (b) self-projected light pattern

Figure 3.3: Mountain-shaped cylinder to be measured.

els in our method, whereas an image width of 70 pixels was required to obtain its whole shape in the conventional light-section method. Our self-projection light-section method demonstrably decreased the image width, thereby directly affecting the frame rate of the camera. Although the change in image width depends on the differential shape, in order to work at a higher frame rate in the high-speed shape inspection, the necessary image width should be determined by the permitted error range of the inspected object rather than the number of pixels that cover all of the deviation in the image. Figure 3.5 shows the height information of the mountain-shaped cylinder on an intersected line. Here height informa-



(a) self-projected light-section method (b) conventional light-section method

Figure 3.4: Captured image of mountain-shaped cylinder.

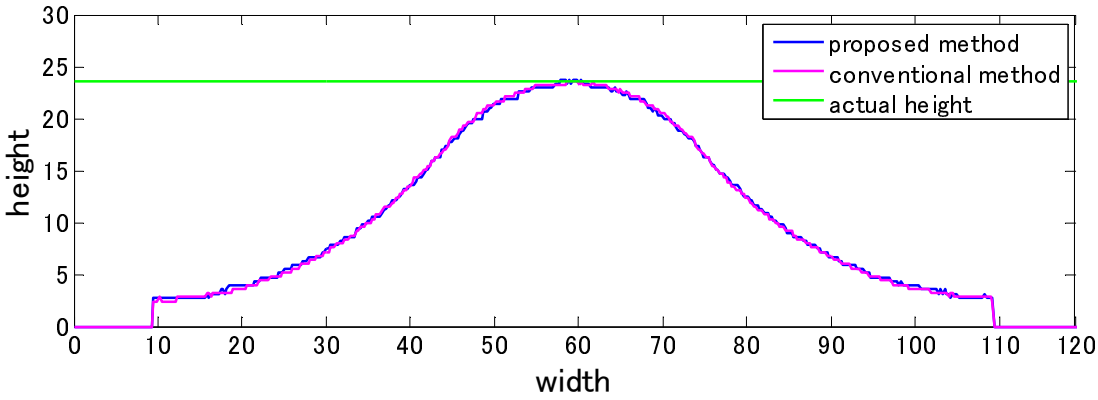
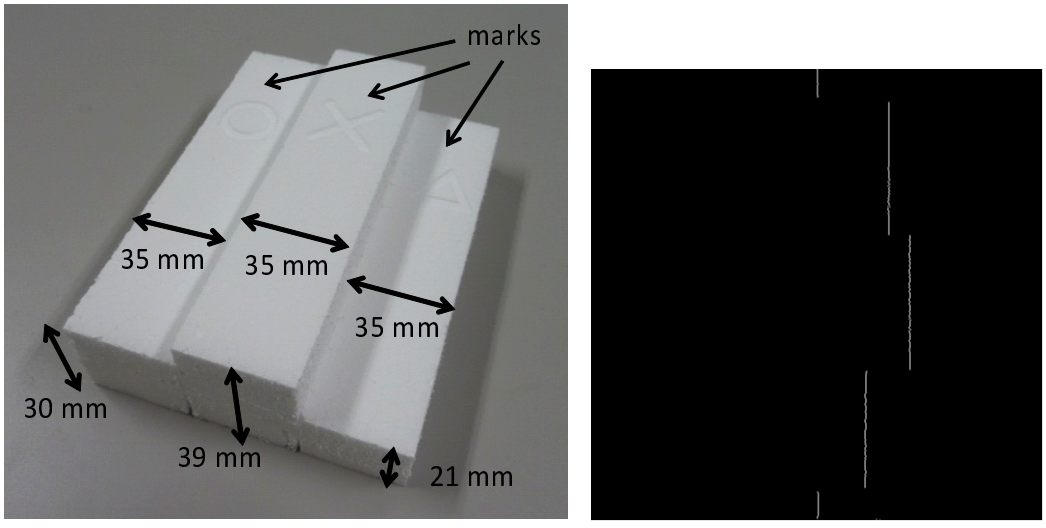


Figure 3.5: Height of mountain-shaped cylinder on an intersected line.

tion was described as a distance from the conveying plane on z = 1167 mm to verify the measured 3-D shapes clearly. By comparison with its actual height, it can be seen that the measurement errors in our self-projected light-section method were within 0.36 mm, while those in the conventional light-section method were within 0.39 mm. This fact indicates that our self-projected light-section method can maintain the same accuracy in 3-D shape measurement as that in the conventional light-section method, while the image region to be processed was remarkably reduced in our method.

Next, measurement results for a stair-shaped object carved with marks were ob-



(a) overview (b) self-projected light pattern

Figure 3.6: Stair-shaped object carved with marks.

Figure 3.7: A result image of stair-shaped object.

tained using our 3-D shape measurement system. The stair-shaped object to be measured is shown in Figure 3.6(a). The stair-shaped object had three steps heights of 30, 39, and 21 mm. The width of each step was 35 mm. The marks ∘ , ±, and ∀ were carved on these steps. Their sizes were approximately 22 ± 22 mm. The carved depth and width were 1 and 3 mm, respectively.

When a similar stair-shaped object with no carved marks was used as a reference object, its self-projected light pattern was generated as shown in Figure 3.6(b). Figures 3.7 gives an example of a result image with deviation corresponding to the middle part of the carved marks. Figures 3.8 and 3.9 show the absolute 3-D shape and the differential 3-D shape measured by our self-projected light-section method, when the stair-shaped object with carved marks was conveyed at 1000 mm/s by a linear slider. The measurement time was 0.14 s. For the absolute 3-D shape, the measured heights of the three steps were

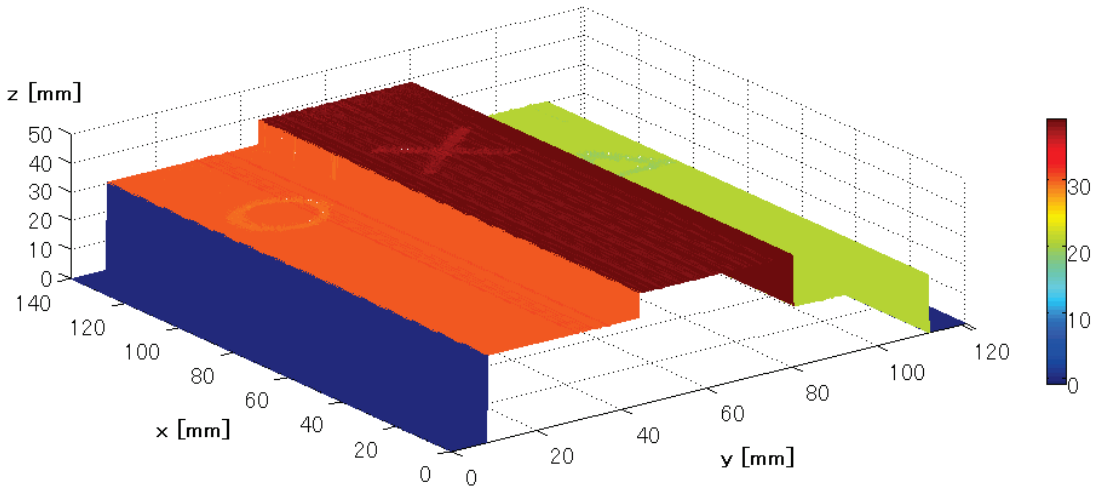
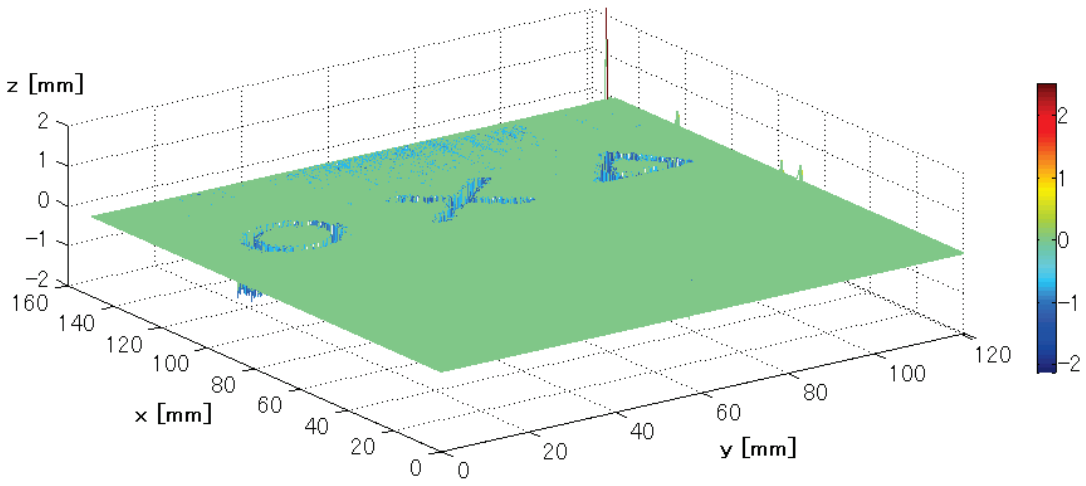
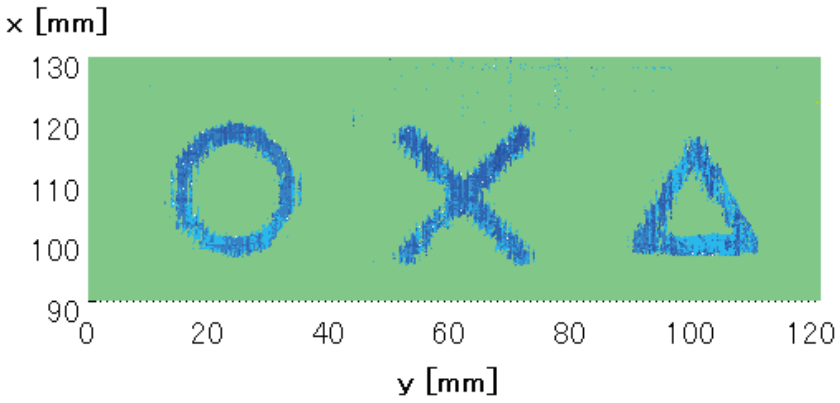


Figure 3.8: Measured absolute 3-D shape of a stair-shaped object in *xyz* space.



(a) *xyz* space



(b) *xy* space

Figure 3.9: Measured differential 3-D shape of a stair-shaped object.

30.12, 38.90, and 21.04 mm, respectively, where there were no carved marks. For the differential 3-D shape, we can observe the marks ∘ , ±, and ∀ with carving depth of 1.09 mm or less. These measurement results show that our self-projected light-section method enables both global 3-D shape measurement and local abnormality inspection in real time at 10000 fps to the same accuracy level as the conventional light-section method, even when the measurement object has different heights that would require a large-size image region in the conventional light-section method.

3.5 Concluding Remarks

In this chapter, we proposed a novel light-section method for fast 3-D shape inspection that can reduce the pixel dimensions of the image to be processed without decreasing accuracy, assuming that a reference object is given. Our self-projected light-section method was integrated on a high-speed vision system as a real-time 3-D shape inspection system at 10000 fps. Its performance was verifie by showing several real-time experimental results, compared with those measured with the conventional light-section method. In this study, it was assumed that measurement objects have cylinder-like shapes because the frame rate of the projector was much slower than that of the high-speed vision platform. In next chapter, we will extend our 3-D shape measurement method and system to arbitrary 3-D shape objects by synchronizing a high-speed vision platform with a high-frame-rate pattern light projector such as a DMD projector.

Chapter 4

Self-Projected Structured Light Method

4.1 Introduction

In this chapter, we introduce a self-projected structured light method that can reduce the number of projections without decreasing the 3-D measurement accuracy by projecting multiple curved-stripe patterns onto the measured 3-D scene; it is assumed that a reference 3-D shape is initially given, and the curved-stripe patterns to be projected are generated by the reference 3-D shape. We also develop a high-frame-rate (HFR) structured light system for real-time fast 3-D shape inspection by implementing our method on a GPU-based high-speed vision system synchronized with a high-speed projector, and verify its effectiveness via experiments using 512 ± 512 depth images captured in real time at 500 fps.

4.2 GPU-Based Real-Time Structured Light 3-D Scanner

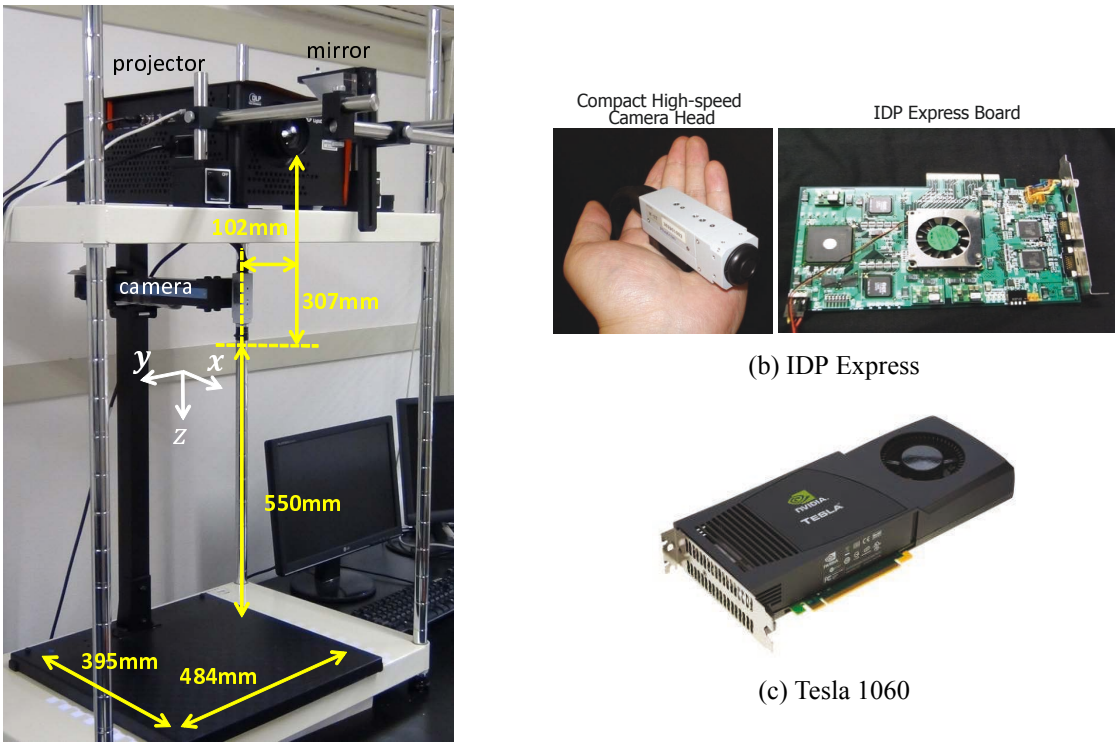
The coded structured light-projection method [48] has often been used for 3-D shape acquisition by projecting multiple light patterns on the object being observed. Many 3-D measurement systems based on this method have been developed for quickly and accurately determining the 3-D shape of a static object. Most of them use image sensors that follow standard video signals (e.g., NTSC 30 fps or PAL 25 fps). Due to synchronization errors in projecting multiple light patterns with different timings, the frame rate of the

image sensor strongly restricts the accuracy with which 3-D shapes of moving objects can be measured using the coded structured light method.

To reduce the constraints imposed by conventional video signals, many high-speed vision systems, with hardware-implementation of various image processing algorithms, have been developed for real-time processing at 1000 fps or more. In recent years, Ishii et al. have developed IDP Express [10], a high-speed vision platform that can simultaneously process a high-frame-rate (HFR) video and directly map it onto the allocated memory in the PC. Such a high-speed vision platform makes it possible to capture and process multiple light patterns projected from a DLP (Digital Light Projector) with a digital micromirror device (DMD), even when these light patterns are projected at 1000 fps or more, which is much faster than standard video signals.

In this study, we develop a real-time structured light 3-D scanner that can output 3-D video of 512±512 pixels in real time at 500 fps. Our 3-D scanner can execute a coded structured light projection method that processes zebra light patterns projected for pixel-level accuracy in the 3-D measurements. Assuming HFR videos for capturing the light patterns projected at a high frame rate, we can decrease synchronization errors between the light patterns to much less than those when using standard videos at dozens of frames per second. An overview of our 3-D scanner is shown in Figure 4.1(a). This system consists of a high-speed DLP (Texas Instruments DLP Light Commander 5500), the high-speed vision platform IDP Express [10], a GPU board (NVIDIA Tesla 1060), and a personal computer (PC). A right-angled mirror coated with enhanced aluminum is used to change the direction of projection from the DLP vertically.

The optical center of the projector lens was located at $(x_0, 0, z_0) = (102\text{ mm}, 0, \quad 307\text{ mm})$, and its optical axis formed an angle of $\theta_0 = 12^{\leq}$ with respect to the z axis. The focal lengths of the camera and projector lenses were set to $f_c = 6$ mm and $f_p = 20$ mm, respectively. The projector projected 1024 ± 768 images in a square measuring 526 ± 395 mm on the level plane at $z = 550$ mm. The projected pattern was captured as a 512 ± 512 image, which corresponded to a 484 ± 484 mm square on the level surface. Depth images could be calculated in a 484 ± 395 mm region on the level surface. To decide the precise parameters of the optical system that are used for depth calculation, we use the projector-camera



(a) System overview

(b) IDP Express

(c) Tesla 1060

Figure 4.1: Real-time GPU-based 3-D scanner.

calibration method developed by Moreno et al. [61]. In this method, a sequence of gray code patterns is projected onto a static planar checkerboard and images are captured for each pattern to fin the correspondences between projector and camera pixels. The set of correspondences is used to compute a group of local homographies that allow us to fin the projection of any of the points in the calibration object onto the projector image plane. Then, the projector can be calibrated as a normal camera. Zhang's method [62] is used to calculate the intrinsic and extrinsic parameters of projector and camera because of its simplicity and well-known accuracy.

The DLP LightCommander 5500 is a multi-functional development kit for high-speed projection based on DMD device technology, and is composed of a high-performance LED light engine, a DLP 0.55 XGA Chipset, and its controller. The projector can iteratively project hundreds of 1024±768 binary patterns at a frame rate of 1000 fps or more, in synchronization with an external system, such as the high-speed vision platform.

The IDP Express [10] was designed to implement various types of image processing algorithms, and to record both images and features at a high frame rate on standard PC memory. It consists of a compactly designed camera head and a dedicated FPGA board (IDP Express board). Figure 4.1(b) shows an overview of the IDP Express. The camera head can capture 8-bit gray-level images of 512±512 pixels at 2000 fps, and can transfer images to the IDP Express board at 2000 fps. The IDP Express board has two camera inputs and trigger I/Os to synchronize with external systems, and it can be connected to a PC. We can implement user-specifi image processing algorithms on an FPGA (Xilinx XC3S5000). Two 512±512 images and their processed results can be simultaneously mapped onto standard PC memory at 2000 fps via a PCI-e bus.

The Tesla C1060 is a computer processor board based on the NVIDIA Tesla T10 GPU. It is capable of a processing performance of 933 Gflop /s, using 240 processor cores operating at 1.296 GHz and a bandwidth of 102 GB/s for its inner 4 GB memory. Figure 4.1(c) shows an overview of the Tesla 1060. A PC with a 16-lane PCI-e 2.0 bus and a processor chipset with a DMA function are adapted to transfer memory-mapped data between standard memory and the Tesla 1060 at high speed via the PCI-e bus. In this study, the algorithm for coded structured light 3-D measurement was accelerated by parallel-processing software on the Tesla 1060. We use a CUDA IDE for NVIDIA to code algorithms with dedicated API functions for the IDP Express, which enables us to access memory-mapped data through Windows XP (32 bit). We use a PC with the following specifications ASUS P6T7 WS SuperComputer main board, Intel Core (TM) i7 3.20 GHz CPU, 3 GB of memory, and two 16-lane PCI-e 2.0 buses.

4.3 Implementation of Coded Structured Light Projection Method

The coded structured light projection method [48] acquires 3-D images by projecting multiple zebra (black and white) light patterns onto the objects to be observed. Projection images captured by a camera with a different viewing angle from that of the projector are used for 3-D image calculation. For n zebra light patterns, the measurement

space is divided into 2^n vertical slices. By judging each pixel in the image as bright or dark, n-bit data is obtained at each pixel as a space code that shows the direction of the corresponding light ray from the projector. Depth information can be obtained at each pixel using the triangulation of the relationship between a space code image and the measurement directions determined by pixel positions.

In Posdamer's method, zebra light patterns with a pure binary code are projected. This may have severe encoding errors, even when there are only small noises, because the brightness boundaries of multiple zebra light patterns with a pure binary code are in the same positions. Inokuchi et al. [49] introduced a gray-code light pattern projection method to minimize these boundary encoding errors. Bergmann [50] proposed an improved 3-D shape measurement method, combining the gray-code pattern projection method and the phase shift method, which increased the number of projection patterns. To reduce the number of projection patterns, Capsi et al. [53] proposed a color-type gray-code pattern light projection method. Most of these methods can obtain an accurate 3-D image of static objects by distance calculation at every pixel, whereas synchronization errors are generated when moving objects are observed because multiple light patterns are projected at different times.

Several one-shot projection methods based on spatially-coded light patterns have been proposed for the 3-D measurement of moving objects. Pages et al. [54] introduced a color stripe pattern based on the De Bruijn sequence, which combined the high resolution of classical striped patterns and the accuracy of multi-slit patterns. Ito et al. [55] proposed a grid pattern, named the three-level checkerboard pattern, where each cell has a gray level chosen from three intensity values, and Griffin et al. [57] define a pattern consisting of an array of colored circles using an alphabet of four symbols. Sakashita et al. [58] obtained both the 3-D shape and texture of an object by capturing a one-shot color-patterned grid with a multi-band camera and an infrared projector. Microsoft's Kinect [59] is a computer interface that can acquire a 3-D shape in real time at 30 fps by capturing an infrared light pattern that is spatially coded. Most of these methods can obtain 3-D images of dynamically changing environments without any synchronization errors by capturing only a single projection pattern at dozens of frames per second with standard

cameras. However, their 3-D measurement accuracies are much worse than the spatial resolutions of image sensors, because most of them assume a local surface smoothness of the object in spatial encoding with a single light pattern projection. Therefore, there is a trade-off relationship between the accuracy and synchronization error in the acquisition of 3-D images using the coded structured light approach, depending on the number of light patterns to be projected.

4.3.1 Implemented Algorithm

We implement a coded structured light projection method [49] using light patterns coded with an 8-bit gray code for 512±512 images on a GPU-based high-speed vision platform. The DLP LightCommander 5500 projects 1024±768 binary patterns at 1000 fps, and the IDP Express captures 512±512 images of the projected light patterns at 1000 fps in synchronization with the DLP. In this study, the exposure time of the camera head of the IDP Express is set to 1 ms. The implemented algorithm is described as follows:

(1) Projection of 8-pairs of positive/negative light patterns.

The DLP iteratively projects 8 pairs of positive and negative light patterns of 1024±768 pixels with an 8-bit gray code in the order $\}g_0, g_1|$, $\}g_2, g_3|$, ××× and $\}g_{14}, g_{15}|$:

$$g_{2i}(X^{\in}, Y^{\in}) = \left\lfloor \frac{2^i \times Y^{\in}}{768} + \frac{1}{2} \right\rfloor \mod 2, \quad g_{2i+1}(X^{\in}, Y^{\in}) = \overline{g_{2i}(X^{\in}, Y^{\in})} \quad (i = 0, \times\times\times, 7), \quad (4.1)$$

where $\rfloor x\{$ is the greatest integer less than or equal to x. The $X^{\in}Y^{\in}$ coordinate system indicates pixel coordinates in the projection images, and i indicates the LSB and MSB order in the 8-bit gray code.

(2) Acquisition of projected image patterns.

Corresponding to the projection image of $g_j(X^{\in}, Y^{\in})$ $(j = 0, \times\times\times, 15)$, the IDP Express captures a gray-level 512±512 image $I(X, Y, k)$ at time $t = k\tau$ as follows:

$$I(X, Y, k) = \mathrm{Proj}(g_{k \bmod 16}(X^{\in}, Y^{\in})), \quad (4.2)$$

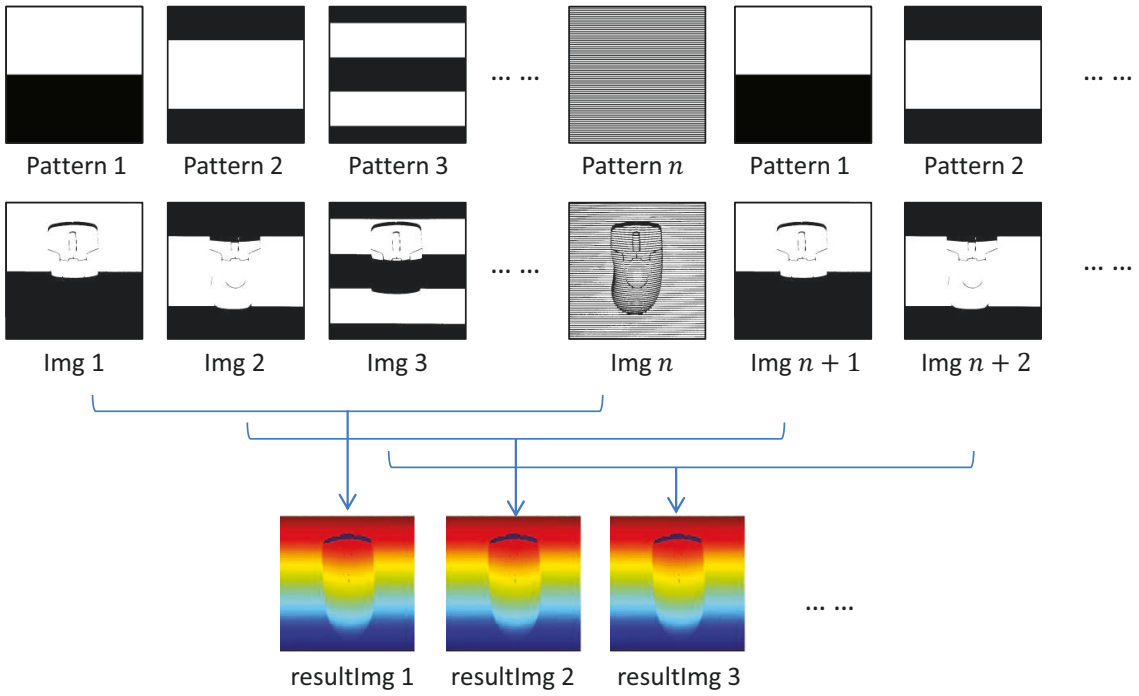


Figure 4.2: Coded structured light projection algorithm implemented on our 3-D scanner.

where k indicates the frame number of the captured images when the frame interval is set to 1 ms (= τ). The XY coordinate system indicates pixel coordinates in the captured image.

(3) Binarization.

At 2 ms (= 2τ) intervals, a pair of captured images $I(X, Y, 2k^{\in})$ and $I(X, Y, 2k^{\in} + 1)$, corresponding to the ($k^{\in}$ mod 8)-th positive and negative light patterns, are differentiated for robustness to ambiguities caused by nonuniform brightness. Thus, we obtain the binary image $G(X, Y, 2k^{\in} + 1)$ for space encoding at frame $2k^{\in} + 1$ by binarization with a threshold θ_b:

$$G(X, Y, 2k^{\in} + 1) = \begin{cases} 1 & I(X, Y, 2k^{\in} + 1) \quad I(X, Y, 2k^{\in}) > \theta_b \\ 0 & I(X, Y, 2k^{\in} + 1) \quad I(X, Y, 2k^{\in}) < \quad \theta_b \\ \phi & \text{(otherwise)} \end{cases} . \tag{4.3}$$

where ϕ indicates the ambiguous binarization state arising due to occlusion from the cam-

era. In this study, we set the threshold for binarization, θ_b, to 5.

(4) Determination of measurable pixels.

To reduce the number of coding errors caused by binarization ambiguities, the measurable state at frame $2k^{\in} + 1$, $a(X, Y, 2k^{\in} + 1)$, is determined by counting the number of ambiguous binarization states among the eight binarized images $G(X, Y, 2(k^{\in} \quad i) + 1)$ at frames $2(k^{\in} \quad i) + 1$ $(i = 0, \times\times\times, 7)$ as follows:

$$a(X, Y, 2k^{\in} + 1) = \begin{cases} 1 & c(X, Y, 2k^{\in} + 1) \approx 1 \\ 0 & \text{(otherwise)} \end{cases}, \tag{4.4}$$

where $c(X, Y, 2k^{\in} + 1)$ is the number of ambiguous states among the eight binarized images $G(X, Y, 2(k^{\in} \quad i) + 1)$. In this study, we consider the measurable state $a(X, Y, 2k^{\in} + 1) = 1$ when there are one or zero ambiguous binarizations between the eight frames; the unmeasurable state $a(X, Y, 2k^{\in} + 1) = 0$ indicates that there are two or more ambiguous binarizations between the eight frames.

(5) Gray-to-binary conversion.

When $a(X, Y, 2k^{\in} + 1) = 1$, i.e., depth is measurable at pixel (X, Y) in frame $2k^{\in} + 1$, $G(X, Y, 2k^{\in} + 1)$ is converted with a pure 8-bit binary code using $G(X, Y, 2(k^{\in} \quad i) + 1)$ at the current and previous frames $2(k^{\in} \quad i) + 1$ $(i = 0, 1, \times\times\times, 7)$, corresponding to the $(k^{\in} \bmod 8)$-th bit of the 8-bit gray code as follows:

$$B(X, Y, 2k^{\in} + 1) = \left(\sum_{i=0}^{7} G(X, Y, 2(k^{\in} \quad i) + 1) \right) \bmod 2, \tag{4.5}$$

where we assume that the geometric displacement of the object to be observed is small over 16 ms ($t = (2k^{\in} \quad 15)\tau$ to $(2k^{\in} + 1)\tau$) by setting a small frame interval $\tau = 1$ ms. We substitute 0 for ϕ in this study.

(6) Space-code image generation.

A space-code image $C(X, Y, 2k^{\in} + 1)$ is obtained at frame $2k^{\in} + 1$ using the eight binary images $B(X, Y, 2(k^{\in} \quad i) + 1)$ $(i = 0, 1, \times\times\times, 7)$ at the current and previous frames as

Table 4.1: Execution time of 3-D shape measurement with normal structured light method.

	Time [ms]
Image acquisition time	0.03
Transfer to GPU	0.08
Binarization	0.02
Gray-to-binary conversion	0.02
Space code image generation	0.01
3±3 median filte	0.02
3-D triangulation	0.01
Transfer to PC memory	0.81
Total time	1.00

follows:

$$C(X, Y, 2k^{\in} + 1) = \sum_{i=0}^{7} 2^{(k^{\in} \ \ i) \bmod 8} \times B(X, Y, 2(k^{\in} \quad i) + 1), \tag{4.6}$$

where we assume a small geometric displacement of the object over 16 ms for $t = (2k^{\in}$ 15)τ to $(2k^{\in} + 1)\tau$. A 3±3 median filte was used for the space code image, in order to interpolate uncertain pixels where $a(X, Y, 2k^{\in} + 1) = 0$.

(7) Triangulation.

The space code image $C(X, Y, 2k^{\in} + 1)$ is transformed into 3-D information (x, y, z) by solving the following simultaneous equation with a 3±4 camera transform matrix T_C and a 2±4 projector matrix T_P:

$$H_C \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = T_C \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \quad H_P \begin{pmatrix} C(X, Y, 2k^{\in} + 1) \\ 1 \end{pmatrix} = T_P \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \tag{4.7}$$

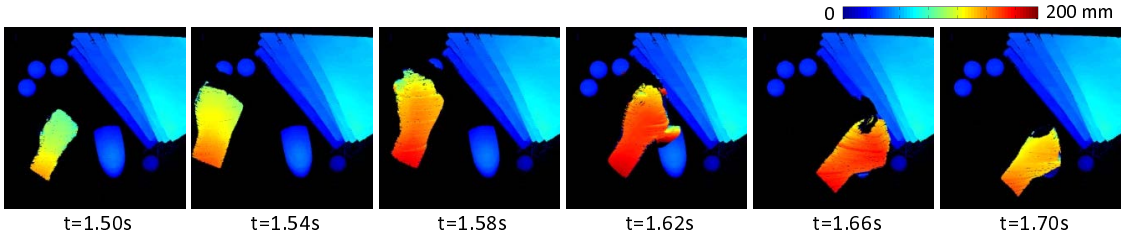
where H_C and H_P are parameters. It is assumed that the matrices T_C and T_P are obtained by prior calibration. Thus, all of the pixels in the 512±512 image are transformed as a 3-D image at an interval of 2 ms (= 2τ).

Subprocesses (3)–(7) are accelerated by parallelizing them with 512 blocks of 1±512 pixels on the GPU board. Table 4.1 shows the total execution time; this includes the transfer time from the PC memory to the GPU board of a pair of gray-scale 512±512 images, and the transfer time from the GPU board to the PC memory for the processed 3-D images. We can confir that the 3-D shape measurement is accelerated by implementing this algorithm on the GPU board; the total time is 1.00 ms, and we confirme the 3-D shape measurement in real time at a frame rate of 500 fps.

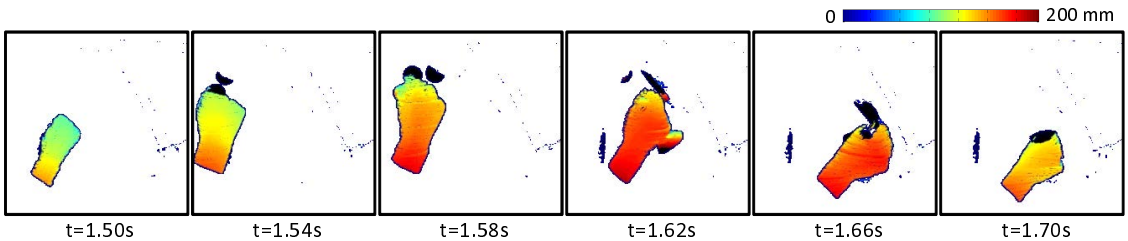
4.3.2 Experiments

To verify the performance of our real-time 3-D scanner that can output 3-D images of 512±512 pixels at 500 fps, we show experimental results for three types of moving scenario: (a) a moving human hand, (b) a rippling liquid surface, and (c) human finger

(a) Experimental scenes

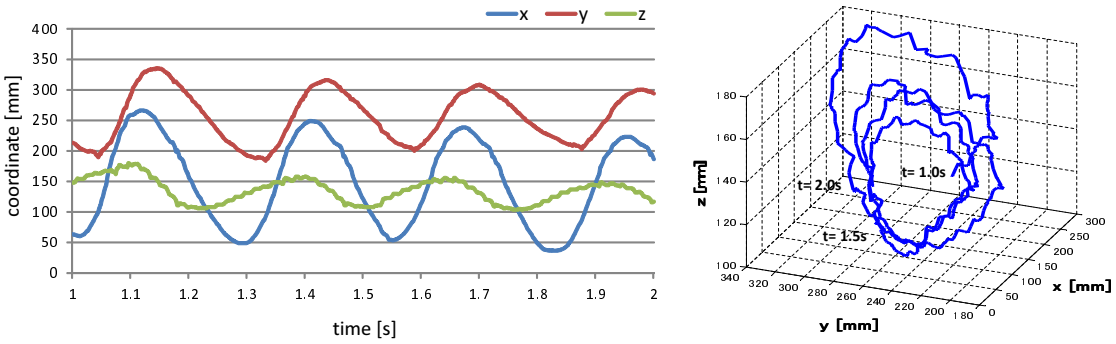


(b) 3-D images



(c) Extracted 3-D hand shapes

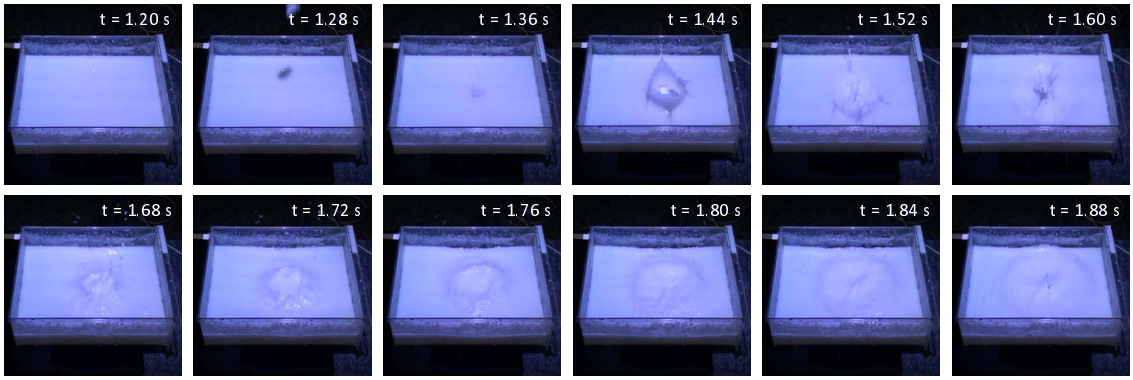
Figure 4.3: Captured 3-D images of a moving human hand.



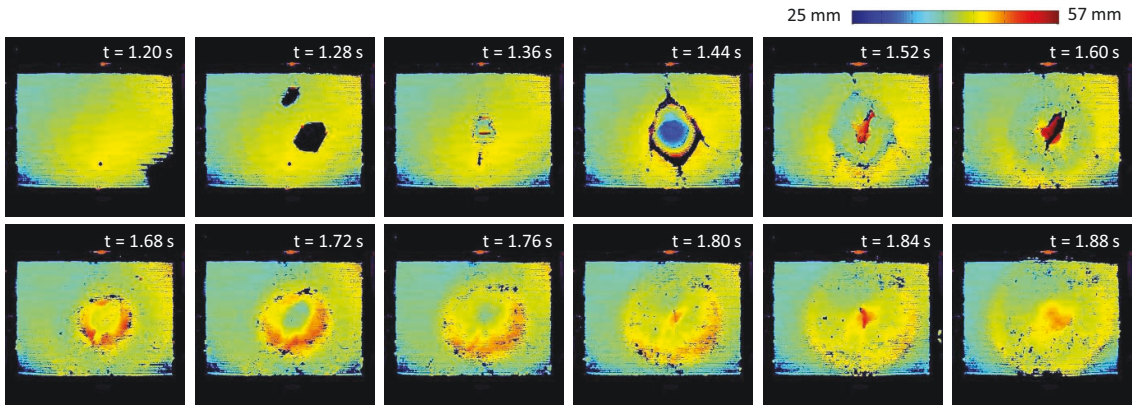
(a) *xyz* coordinate values (b) 3-D trajectory

Figure 4.4: 3-D position of the extracted fist

tapping keys on a computer keyboard. First, we show the measured result for a human hand that was periodically clenched into a fis at a frequency of approximately 3.5 times per second in 3-D space. Figure 4.3 shows (a) the images captured at 30 fps using a standard NTSC video camera, (b) the 3-D color-mapped images calculated using our developed 3-D scanner, and (c) the 3-D hand shapes extracted by differentiating the 3-D input images from those of the background objects, taken at intervals of 0.04 s for t = 1.50–1.70 s. The observations started at t = 0. The fis of the operator moved on a circular orbit whose plane was vertically slanted in a depth range from approximately 100 mm to 200 mm above the level surface. On the level surface, we placed a computer mouse, books, and balls as background objects. In the experiment, we confirme that the 3-D position of the fis can be obtained in real time at 500 fps by differentiating the 3-D input images from those of the background objects, and calculating the 3-D centroid information based on the 0th and 1st moment features of the differentiated 3-D images. It can be observed that the 3-D shapes of both the moving human hand and background objects were measured in Figure 4.3(b), while only the 3-D shape of the fis was correctly extracted in Figure 4.3(c), even when the human hand moved quickly over a complex background. Figure 4.4 shows (a) the temporal changes of the x, y, and z coordinate values of the fist s 3-D position for t = 1.0–2.0 s, and (b) its trajectory in 3-D space. It can be observed that the centroid position changed periodically at a frequency of 3.5 Hz on the circular orbit on a certain vertically slanted plane, corresponding to the quick and periodic movement of the fis in 3-D space.



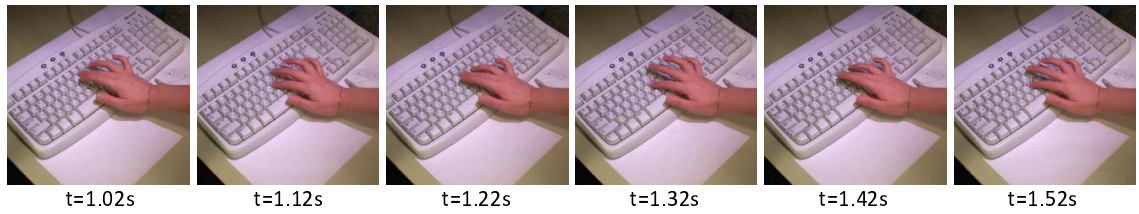
(a) Experimental scenes



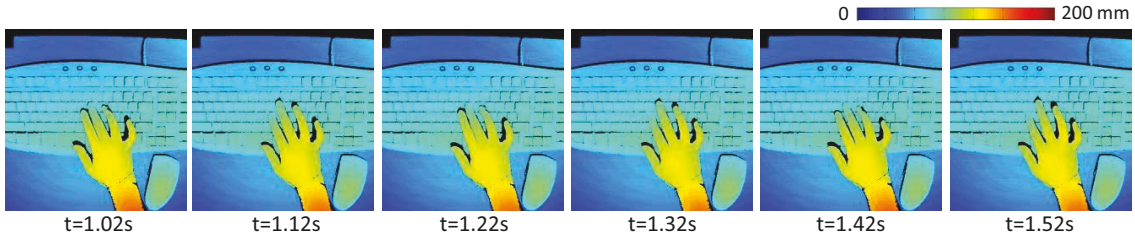
(b) 3-D images

Figure 4.5: 3-D images of waves on the surface of a liquid.

Next, we show the measured result for waves on the surface of a pool of water after an eraser of size 27±23±11 mm is dropped into it. The size of the pool was 34±25 cm, and it was fille with water whitened by waterborne acrylic white paint. The water depth was 3 cm. Figure 4.5 shows (a) the images captured using a standard NTSC video camera, and (b) the 3-D color-mapped images calculated using our scanner. The observation start time was $t = 0$, and the eraser landed on the water surface at $t = 1.36$ s. We can observe a temporal change in the 3-D shape of the water surface after the eraser lands on the water surface, and the wave propagates radially from the point of impact. In general, the propagation speed of a surface wave is given by $v = \sqrt{gh}$ for a water depth of h; g (= 9.8 m/s^2) is gravitational acceleration. It can be observed that the temporal changes in the 3-D shapes on the water surface, which were generated by wave propagation at a



(a) Experimental scenes



(b) 3-D images

Figure 4.6: 3-D images of finge -tapping on a keyboard.

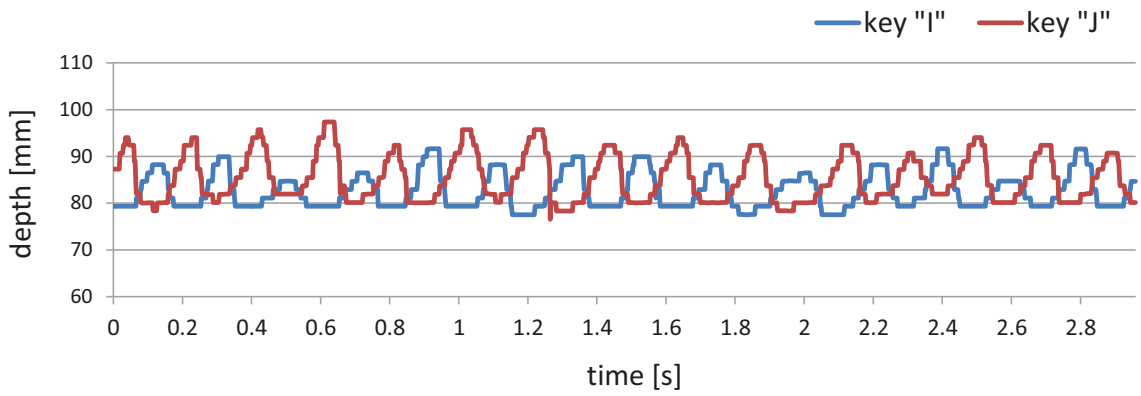


Figure 4.7: Depth information on tapped keys.

speed of v = 54 cm/s, were measured in real time.

Finally, we show the measured result for human finger rapidly tapping the keys on a computer keyboard. Figure 4.6 shows (a) the images captured using a standard NTSC camera, and (b) the 3-D color-mapped images calculated using our 3-D scanner, taken at intervals of 0.1 s. In the experiment, the operator inputted approximately ten letters per second by tapping keys on a computer keyboard of 18 mm key pitch. The keys "I" and "J" were alternately tapped by the forefinge and middle finge at a frequency of 5 Hz.

Figure 4.7 shows the temporal changes in depth at the pixels $(220, 247)$ and $(249, 224)$ around the keys "I" and "J" in the 3-D images of 512±512 pixels for t = 0.0–3.0 s. It can be observed that the depth information at the pixels around the forefinge and middle finge changed periodically with an amplitude of approximately 10 mm, corresponding to the alternative finge tapping motion on the keyboard, even when this motion was too rapid for the human eye to see.

4.4 Self-Projected Structured Light System

We propose a self-projected structured light method that can reduce the number of projections for fast three-dimensional (3-D) shape inspection even when there are large height differences in the measured 3-D scene. In our method, multiple curved-stripe patterns generated by a reference 3-D shape are projected onto the measured 3-D scene, and 3-D shapes can be obtained by processing multiple straight-stripe-like patterns projected in the camera view. Compared with the straight-stripe projections in most structured light methods, our self-projected structured light method can reduce the number of curved-stripe projections by facilitating fast processing and accurate acquisition of 3-D shapes when a reference 3-D shape is provided prior.

4.4.1 Triangulation for A Camera and Projector System

In our study, we considered our self-projected structured light method for the camera and the projector system shown in Fig. 4.8, where a camera model of perspective projection is assumed for the camera and projector. The xyz coordinate system is define as the world coordinate system and its origin O is at the optical center of the camera lens. The X_cY_c coordinate system on the camera image plane is perpendicular to the optical axis and its origin is located at the intersection with the optical axis of the camera lens. The X_c and Y_c axes are parallel to the x and y axes, respectively, and the X_cY_c plane is at a distance of f_c from the optical center. The projector is installed in a different direction than the camera; that is, the optical center of the projector lens is $O^c(x_0, 0, z_0)$. The optical axis of the projector lens forms an angle θ_0 with respect to the z axis. The X_pY_p coordinate

system on the projector image plane is perpendicular to the optical axis of the projector lens. The X_pY_p plane is at a distance of f_p from the optical center of the projector lens. The Y_p axis is parallel to the y axis.

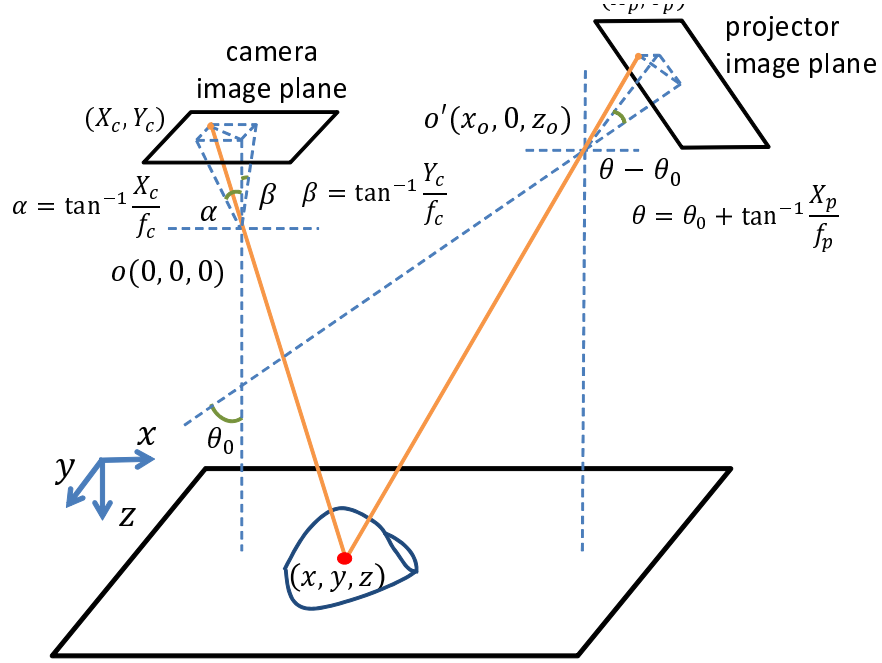


Figure 4.8: Geometry of the camera and the projector system.

When the projector projects a beam of light at a point $\boldsymbol{x} = (x, y, z)$ on the measured object, its position can be obtained by triangulation using its corresponding points on the camera image plane and the projector image plane, $\boldsymbol{X}_c = (X_c, Y_c)$ and $\boldsymbol{X}_p = (X_p, Y_p)$, as follows:

$$\boldsymbol{x} = \frac{x_0 + z_0 \tan\theta}{\tan\alpha + \tan\theta}(\tan\alpha, \tan\beta, 1) \tag{4.8}$$

where α, β, and θ are define using $\boldsymbol{X}_c$ and $\boldsymbol{X}_p$ as follows:

$$\alpha = \tan^{1}\frac{X_c}{f_c} \quad \beta = \tan^{1}\frac{Y_c}{f_c} \quad \theta = \theta_0 + \tan^{1}\frac{X_p}{f_p} \tag{4.9}$$

4.4.2 Generation of Self-Projected Patterns

Self-projected patterns coded with an m-bit gray code are generated initially to observe straight-stripe patterns in the camera view when a reference 3-D shape is set in the measured area. In our method, the following $2m$ straight-stripe patterns are generated on the camera image plane as m pairs of positive and negative light patterns when observing the reference 3-D shape:

$$g_{2i}(\boldsymbol{X}_c) = \left\lfloor \frac{X_c}{a2^m\delta_c} 2^i + \frac{1}{2} \right\rfloor \bmod 2 \tag{4.10}$$

$$g_{2i+1}(\boldsymbol{X}_c) = \overline{g_{2i}(\boldsymbol{X}_c)} \quad (i = 0, 1, \times\times\times, m \quad 1) \tag{4.11}$$

where $\rfloor x\{$ is the greatest integer less than or equal to x, δ_c is the pixel pitch of the image sensor, and a is a parameter that determines the minimum width of the stripes on the camera image plane. When the reference 3-D shape is observed, the relationships between the corresponding points of $\boldsymbol{X}_p$ and $\boldsymbol{X}_c$, which are described using Eqs. (4.8)–(4.9), can be determined by obtaining the projected positions on the camera image plane when a beam of light is raster-scanned onto the projector image plane.

In our method, straight-line patterns are scanned in two directions on the projector image plane to determine the self-projected patterns. First, a line $X_p = X$ is scanned in the direction of the X_p axis on the projector image plane. The projected pattern on the 3-D reference shape is measured as the curve $X_c = I_{X_p=X}(Y_c)$ on the camera image plane. Similarly, the straight-line pattern $Y_p = Y$ is scanned in the direction of the Y_p axis. The projected pattern is measured as the curve $Y_c = I_{Y_p=Y}(X_c)$ on the camera image plane. Corresponding to the point $\boldsymbol{X}_p$ on the projector image plane, the point $\boldsymbol{X}_c$ measured on the camera image plane is designated as the following crosspoint of the two curved lines:

$$\boldsymbol{X}_c = \boldsymbol{f}_R(\boldsymbol{X}_p) = (I_{X_p}(Y_c), I_{Y_p}(X_c)) \tag{4.12}$$

Thus, the $2m$ self-projected patterns on the projector image plane can be generated as

follows:

$$S_i(\boldsymbol{X}_p) = g_i(I_{X_p}(Y_c), I_{Y_p}(X_c)) \; (i = 0, \times\times\times, 2m \quad 1) \tag{4.13}$$

4.4.3 Depth Calculation

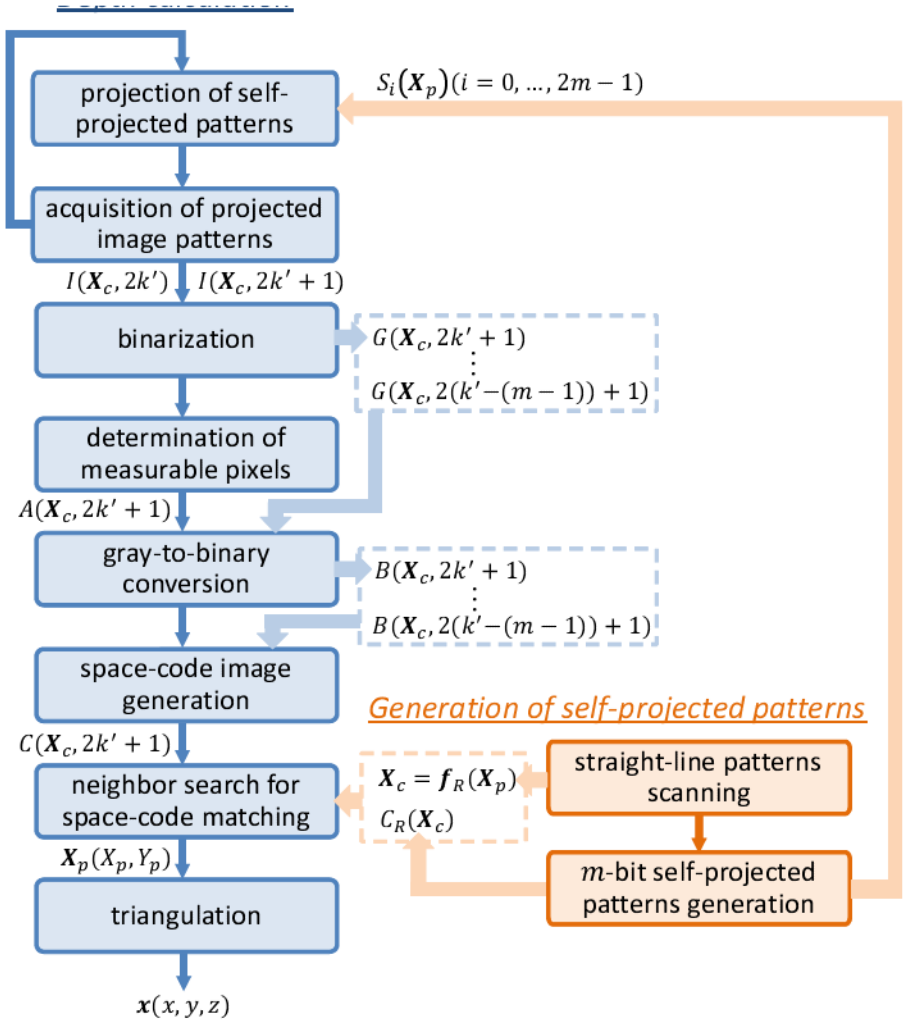


Figure 4.9: Flowchart for the self-projected structured light method.

After generating the self-projected curved-stripe patterns produced using Eq. (4.13), the depth images of the measured object can be obtained using the following processes. Fig. 4.9 gives a fl wchart for depth calculation using our self-projected structured light method.

(1) Projection of positive/negative self-projected patterns

The projector projects m pairs of positive and negative self-projected patterns coded with an m-bit gray code in the order $\}S_0, S_1|$, ×××, and $\}S_{2m\ 2}, S_{2m\ 1}|$, as described by Eq. (4.13).

(2) Acquisition of projected image patterns

A gray-level image is captured at time $t = k\tau$ as follows:

$$I(\boldsymbol{X}_c, k) = \text{Proj}(S_{k \bmod 2m}(\boldsymbol{X}_p)) \tag{4.14}$$

where k indicates the frame number of the captured images when the frame interval is set to τ.

(3) Binarization

The binary image used for space encoding is obtained by differentiating a pair of images $I(\boldsymbol{X}_c, 2k^{\in})$ and $I(\boldsymbol{X}_c, 2k^{\in} + 1)$ and binarizing the differentiated image as follows:

$$G(\boldsymbol{X}_c,2k^{\in}+1)=\begin{cases} 1 & I(\boldsymbol{X}_c,2k^{\in}) \quad I(\boldsymbol{X}_c,2k^{\in}+1)>\theta_b \\ 0 & I(\boldsymbol{X}_c,2k^{\in}) \quad I(\boldsymbol{X}_c,2k^{\in}+1)< \quad \theta_b \\ \phi & \text{(otherwise)} \end{cases} \tag{4.15}$$

where θ_b is the binarization threshold and ϕ indicates the ambiguous state that arises due to a lack of image contrast and occlusion of the camera view.

(4) Determination of measurable pixels

To reduce the number of coding errors caused by binarization ambiguities, the measurable state $A(\boldsymbol{X}_c, 2k^{\in} + 1)$ at frame $2k^{\in} + 1$ is determined by counting the number of ambiguous binarization states among the m bit binarized images $G(\boldsymbol{X}_c, 2(k^{\in} \quad i) + 1)$ at frames $2(k^{\in} \quad i) + 1$; $i = 0, ..., m \quad 1$ as follow:

$$A(\boldsymbol{X}_c, 2k^{\in} + 1) = \begin{cases} 1 & c(\boldsymbol{X}_c, 2k^{\in} + 1) \approx 1 \\ 0 & \text{(otherwise)} \end{cases} \tag{4.16}$$

where $c(\boldsymbol{X}_c, 2k^{\in} + 1)$ is the number of ambiguous states among the m binarized images

$G(\boldsymbol{X}_c, 2(k^{\in} \quad i) + 1)$. In our method, we consider the measurable state $A(\boldsymbol{X}_c, 2k^{\in} + 1) = 1$ when there is one or zero ambiguous binarization between the m frames; the unmeasurable state $A(\boldsymbol{X}_c, 2k^{\in} + 1) = 0$ indicates that there are two or more ambiguous binarizations between the m frames.

(5) Gray-to-binary conversion

$G(\boldsymbol{X}_c, 2k^{\in} + 1)$ is converted with a pure m-bit binary code using $G(\boldsymbol{X}_c, 2(k^{\in} \quad i) + 1)$ in the current and previous frames $2(k^{\in} \quad i) + 1$ ($i = 0, 1, \times\times\times, m \quad 1$), as follows:

$$B(\boldsymbol{X}_c, 2k^{\in}+1) = \left(\sum_{i=0}^{k^{\in} \bmod m} G(\boldsymbol{X}_c, 2(k^{\in} \quad i)+1) \right) \bmod 2 \tag{4.17}$$

where the displacement of the measured object over $t = (2k^{\in} \quad 2m + 1)\tau$ to $(2k^{\in} + 1)\tau$ is assumed to be small. An unmeasurable state is determined when there are two or more ambiguous binarizations between m frames. In this paper, zero is substituted for ϕ when there is one ambiguous binarization.

(6) Space-code image generation

A space-code image $C(\boldsymbol{X}_c, 2k^{\in}+1)$ is obtained using the m binary images $B(\boldsymbol{X}_c, 2(k^{\in} \; i) + 1)$ ($i = 0, 1, \times\times\times, m \quad 1$) in the current and previous frames, as follows:

$$C(\boldsymbol{X}_c, 2k^{\in}+1) = \sum_{i=0}^{m \;\; 1} 2^{(m \;\; 1) \;\; [(k^{\in} \; i) \bmod m]} \times B(\boldsymbol{X}_c, 2(k^{\in} \quad i)+1) \tag{4.18}$$

(7) Neighborhood search for space-code matching

For the point $\boldsymbol{X}_c$ on the camera image plane, its corresponding point $\boldsymbol{X}_p = \boldsymbol{X}_p(\boldsymbol{X}_c)$ on the projector image plane is determined by searching the space-code image of the reference 3-D shape in the direction of the X_p axis, as follows:

$$\boldsymbol{X}_p = \{ \boldsymbol{X}_p^{\in} \;\; \llcorner \boldsymbol{X}_p^{\in} \triangle n(\boldsymbol{X}_R), C_R(\boldsymbol{f}_R(\boldsymbol{X}_p^{\in})) = C(\boldsymbol{X}_c, 2k^{\in}+1) | \tag{4.19}$$

where $\boldsymbol{X}_R = (X_R, Y_R) = \boldsymbol{X}_R(\boldsymbol{X}_c)$ is the point on the projector image plane when observing the reference 3-D shape, which corresponds to the point $\boldsymbol{X}_c$ on the camera image plane. $n(\boldsymbol{X}_R)$ indicates the following neighborhood of $\boldsymbol{X}_R$ in the direction of the X_p axis on the

projector image plane:

$$n(\boldsymbol{X}_R) = \}\boldsymbol{X}_p \quad X_R \quad a^{\in}\delta_p e \approx X_p < X_R + a^{\in}\delta_p(2^m \quad e)| \tag{4.20}$$

where δ_p is the pixel pitch of the projector, $C_R(\boldsymbol{X}_c)$ is the space-code image of the reference 3-D shape, and $\boldsymbol{f}_R(\boldsymbol{X}_p)$ indicates point $\boldsymbol{X}_c$ measured on the camera image plane when the reference 3-D shape is observed, which corresponds to point $\boldsymbol{X}_p$ on the projector image plane. $a^{\in}$ is a parameter that indicates the minimum stripe width in projecting stripe patterns, and e is a parameter to decide the measurable range of the differential shape, which can be manually adjusted in the range $0 \approx e < 2^m \quad 1$. For example, e could be set to $2^m \quad 1$ when all the differential shapes to be observed are under the reference 3-D shape, and set to zero when all the differential shapes to be observed are over the reference 3-D shape. $C_R(\boldsymbol{X}_c)$ and $\boldsymbol{f}_R(\boldsymbol{X}_p)$ can be prestored as look-up tables for generating the self-projected patterns, as described in Subsection 4.4.2.

Here, the point $\boldsymbol{X}_c$ is unmeasurable when $A_R(\boldsymbol{X}_c) = 0$ or $A(\boldsymbol{X}_c, 2k^{\in} + 1) = 0$; $A_R(\boldsymbol{X}_c)$ indicates the measurable state of the space code image of the reference 3-D shape, which is obtained in a similar manner with the above mentioned process (4) for the reference 3-D shape.

(8) Triangulation

The 3-D position $\boldsymbol{x}$ of a measured object is obtained by triangulation based on the relationship between point $\boldsymbol{X}_c$ on the camera image plane and its corresponding point $\boldsymbol{X}_p$ on the projector image plane, as described in Eqs. (4.8) and (4.9).

4.4.4 Measurement Accuracy and Range

The error in 3-D shape measurement in the conventional structured light method is determined by the digitized error in the captured image, which corresponds to the pixel pitch of the image sensor, whereas that in our self-projected structured light method is determined by the digitized error in the projection pattern as well as that in the captured image, which corresponds to the pixel pitch of the projector.

Based on Eq. (4.8), the error $\Delta\boldsymbol{x} = (\Delta x, \Delta y, \Delta z)$ at a point $\boldsymbol{x}$ of 3-D shape measure-

ment in the structured light methods is determined as follows:

$$\Delta x = \frac{f_p(1+\tan^2\theta_0)\tan\alpha(z_0\tan\alpha \quad x_0)}{(f_p \quad X_p\tan\theta_0)^2(\tan\alpha+\tan\theta)^2}\Delta X_p + \frac{\tan\alpha\tan\theta(z_0\tan\theta+x_0)}{X_c(\tan\alpha+\tan\theta)^2}\Delta X_c \quad (4.21)$$

$$\Delta y = \frac{f_p(1+\tan^2\theta_0)\tan\beta(z_0\tan\alpha \quad x_0)}{(f_p \quad X_p\tan\theta_0)^2(\tan\alpha+\tan\theta)^2}\Delta X_p \quad \frac{\tan\beta(z_0\tan\theta+x_0)}{f_c(\tan\alpha+\tan\theta)^2}\Delta X_c \quad (4.22)$$

$$\Delta z = \frac{f_p(1+\tan^2\theta_0)(z_0\tan\alpha \quad x_0)}{(f_p \quad X_p\tan\theta_0)^2(\tan\alpha+\tan\theta)^2}\Delta X_p \quad \frac{z_0\tan\theta+x_0}{f_c(\tan\alpha+\tan\theta)^2}\Delta X_c \quad (4.23)$$

where ΔX_c indicates the digitized error on the camera image plane in the X_c direction, and ΔX_p indicates the digitized error on the projector image plane in the X_p direction. For the error $\Delta\boldsymbol{x}$, the conventional structured light method does not have to consider the firs terms in Eqs. (4.21)–(4.23), which depend on ΔX_p, because it can be assumed that $\Delta X_p = 0$ when straight-stripe patterns are projected, whereas our self-projected method has to consider these terms in the error when the curved-stripe patterns to be projected are deformed by the reference 3-D shape.

When α, β, and $\theta \quad \theta_0$ are small in 3-D shape measurement, that is, when the camera and projector are located sufficiently far from the 3-D scene to be measured, the error $\Delta\boldsymbol{x}$ can be approximately expressed in our self-projected method as follows:

$$\Delta x \ , \ 0 \quad (4.24)$$

$$\Delta y \ , \ 0 \quad (4.25)$$

$$\Delta z \ , \ \frac{x_0(1+\tan^2\theta_0)}{f_p\tan^2\theta_0}\Delta X_p \quad \frac{z_0\tan\theta_0+x_0}{f_c\tan^2\theta_0}\Delta X_c \quad (4.26)$$

In our self-projected method, the 3-D position of a measured object is obtained in a range around the reference 3-D shape corresponding to the neighborhood $n(\boldsymbol{X}_R)$ to be searched for space-code matching expressed in Eq. (4.19). The neighborhood $n(\boldsymbol{X}_R)$

determines the range of the beam angle θ from the projector as follows:

$$\theta_{min} \approx \theta < \theta_{\max} \tag{4.27}$$

$$\theta_{min} = \theta_0 + \tan^{-1}\frac{X_R - a^{\epsilon}\delta_p e}{f_p} \tag{4.28}$$

$$\theta_{max} = \theta_0 + \tan^{-1}\frac{X_R + a^{\epsilon}\delta_p(2^m - e)}{f_p} \tag{4.29}$$

where $\theta_R = \theta_0 + \tan^{-1}(X_R/f_p)$, the beam angle from the projector when the reference 3-D shape is observed, satisfie $\theta_{min} \approx \theta_R < \theta_{max}$. Using Eqs. (4.8) and (4.27), the measurement range of the 3-D position $\boldsymbol{x} = (x, y, z)$, which corresponds to the point $\boldsymbol{X}_c$ on the camera image plane, is determined as follows:

$$z_{min}\tan\alpha \approx x < z_{max}\tan\alpha \tag{4.30}$$

$$z_{min}\tan\beta \approx y < z_{max}\tan\beta \tag{4.31}$$

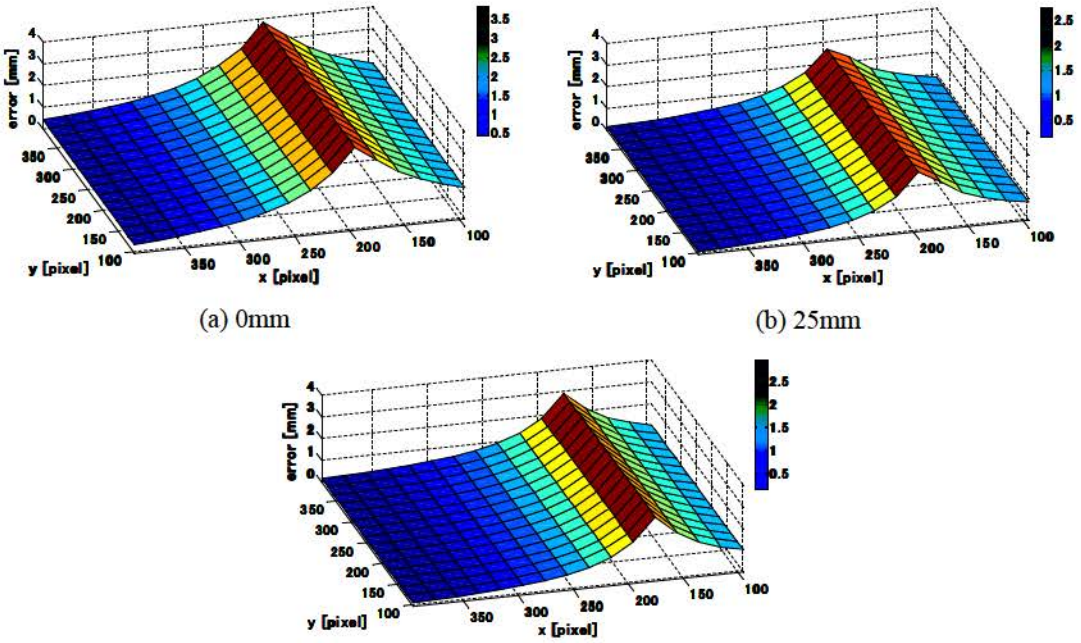
$$z_{min} \approx z < z_{max} \tag{4.32}$$

$$z_{min} = \frac{x_0 + z_0\tan\theta_{min}}{\tan\alpha + \tan\theta_{min}} \qquad z_{max} = \frac{x_0 + z_0\tan\theta_{max}}{\tan\alpha + \tan\theta_{max}} \tag{4.33}$$

Here, the z coordinate of the 3-D position $\boldsymbol{x}_R = (x_R, y_R, z_R)$ when the reference 3-D scene is observed, satisfie $z_{min} \approx z_R < z_{max}$, according to the monotonically increasing function $(x_0 + z_0\tan\theta)/(\tan\alpha + \tan\theta)$ relating to θ. This indicates that the measurement range in 3-D shape measurement in our self-projected method can be determined as a 2^m-tone-resolution range around the reference 3-D shape, whose offset can be adjusted by parameter e in space-code matching.

4.4.5 Specification

Our self-projected structured light method, which projects eight self-projected patterns ($m = 4$), was implemented for 512 ± 512 images using the GPU-based structured light system. When the reference 3-D scene was set to the level surface at $z = 550$ mm and parameter e was set to zero, the error expressed in Eq. (4.23) when the 3-D shapes of the 0-mm-height object, the 25-mm-height object, and the 109-mm-height object on



(a) 0mm (b) 25mm

(c) 109mm

Figure 4.10: The measured error on different height level planes.

the level plane were measured, were 1.79 mm, 1.59 mm, and 1.04 mm, respectively, at the center of the measured region. Fig. 4.10 gives error graphics of 16 ± 16 points that are distributed in the 320 ± 320 pixels in the center area of the image. The measurement range expressed in Eq. (4.32), was z = 457.88–550.0 mm over the level plane at the center of the measured region, when the minimum stripe width in projecting patterns was set to a^{ϵ} = 3 px and the range of neighborhood $n(\boldsymbol{X}_R)$ searched in space-code matching was 48 px.

The self-projected pattern generation process was conducted offline. To generate four pairs of positive and negative self-projected patterns with a 4-bit gray code, 768 and 1024 straight lines were scanned onto the projector image plane in the direction of the X_p and Y_p axes, respectively, which corresponded to the resolution of the projector. The minimum width of the straight-stripe patterns captured on the camera image plane was set to a = 2 px when observing the reference 3-D scene. When generating the eight self-projected patterns, the space-code image of the reference object and the relationships between $\boldsymbol{X}_c$ and $\boldsymbol{X}_p$ and $C_R(\boldsymbol{X}_c)$ and $\boldsymbol{f}_R(\boldsymbol{X}_c)$ were also calculated and stored as 512 ± 512

look-up tables for online depth calculation.

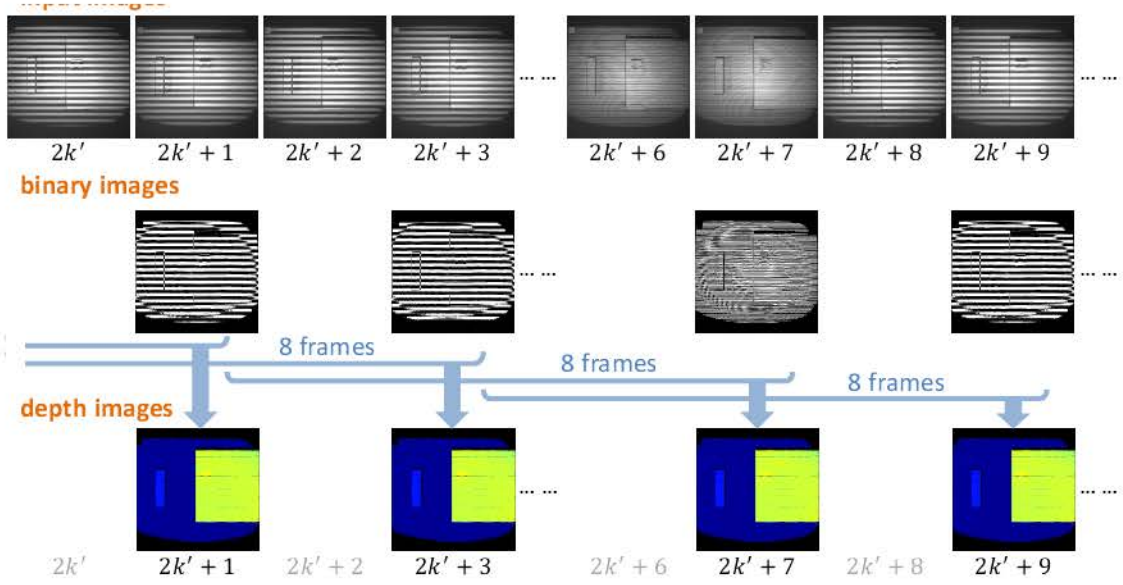


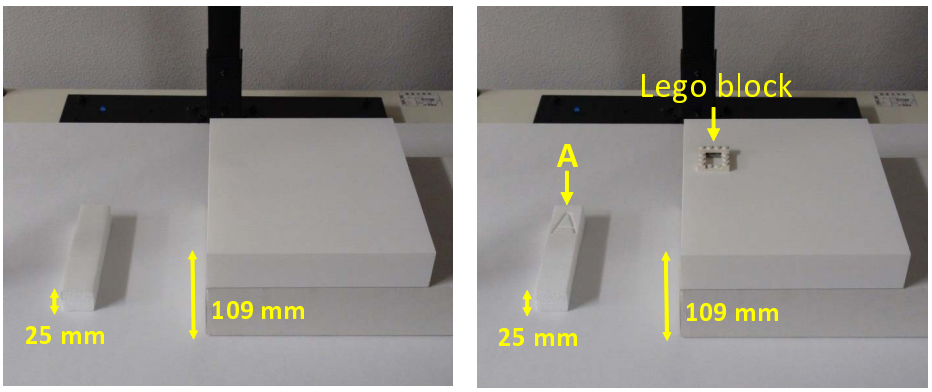
Figure 4.11: Pipelining-output of depth images.

Four pairs of 1024 ± 768 positive and negative self-projected patterns were periodically projected at 1000 fps (τ = 1 ms) and the depth calculations were executed by implementing subprocesses (3)–(8), described in Subsection 4.4.3, as parallelized modules on the GPU board. The depth image was outputted at 500 fps by using pipeline parallel processing for input images between eight continuous frames, as shown in Fig. 4.11. When a pair of images with positive and negative projection patterns is captured, the images will be transferred to the global GPU memory for depth calculation. The two 512 ± 512 input images were divided into 512 blocks of 512 ± 1 pixels, and a binarized image $G(\boldsymbol{X}_c, 2k^{\epsilon}+1)$ of 512 ± 512 pixels was obtained by Eq. (4.15), where the binarization threshold was set to $\theta_b = 5$. Then, a space-code image $C(\boldsymbol{X}_c, 2k^{\epsilon}+1)$ is calculated depending on Eq. (4.16)–(4.18) with the previous three binarized images which are stored in the global GPU memory by division into 512 blocks of 512 ± 1 pixels. To fin the point $\boldsymbol{X}_p$ on the projector image plane corresponding to $\boldsymbol{X}_c$ on the camera image plane, we search the space-code image $C_R(\boldsymbol{X}_c)$ of the reference 3-D shape that is prestored in the global GPU memory and calculate the depth information based on the triangulation in 512 blocks of 512 ± 1 threads. Table 4.2 shows the execution time for the depth calculation, including

Table 4.2: Execution time for 3-D shape measurement with self-projected structured light method

	Time [ms]
Image acquisition	0.03
Transfer to GPU	0.08
Binarization	0.07
Determination of measurable pixels	0.09
Gray-to-binary conversion	0.20
Space code image generation	0.09
Neighborhood search	0.07
Triangulation	0.10
Transfer to PC memory	0.58
Total	1.31

the transfer time from the PC memory to the GPU board for 512 ± 512 input images and that from the GPU board to the PC memory for the processed 512 ± 512 depth images. Our self-projected method based on the 4-bit gray code was accelerated by implementing it on the GPU board; the total execution time was 1.31 ms. We subsequently confirme that 512 ± 512 depth images were obtained in real time at a frame rate of 500 fps.



(a) Reference 3-D scene (b) 3-D scene being measured

Figure 4.12: Experimental scenes with cuboids of different heights.

4.5 Experiments

4.5.1 Static 3-D Scene with Different Heights

We verifie the accuracy of the depth images using a static 3-D scene with different heights observed using our system. Fig. 4.12 shows the 3-D scene measured and the reference 3-D scene for self-projected projection. In both scenes, the 25-mm-height and 109-mm-height cuboids were located at the same position and with the same orientation on the level plane. The differences between these 3-D scenes were whether the letter "A" was carved at a 2-mm depth on the 25-mm-height cuboid and the toy blocks of 3-mm-height, 6-mm-height and 9-mm-height were located on the 109-mm-height cuboid.

4.5.1.1 Self-Projected Patterns Generation

To obtain the self-projected patterns for the measured 3-D scene, we use straight lines to scan the measured fiel in both the x and y directions. First, 768 straight lines of $X_p = 0, 1, ..., 767$ are projected one by one and 768 images are taken simultaneously.

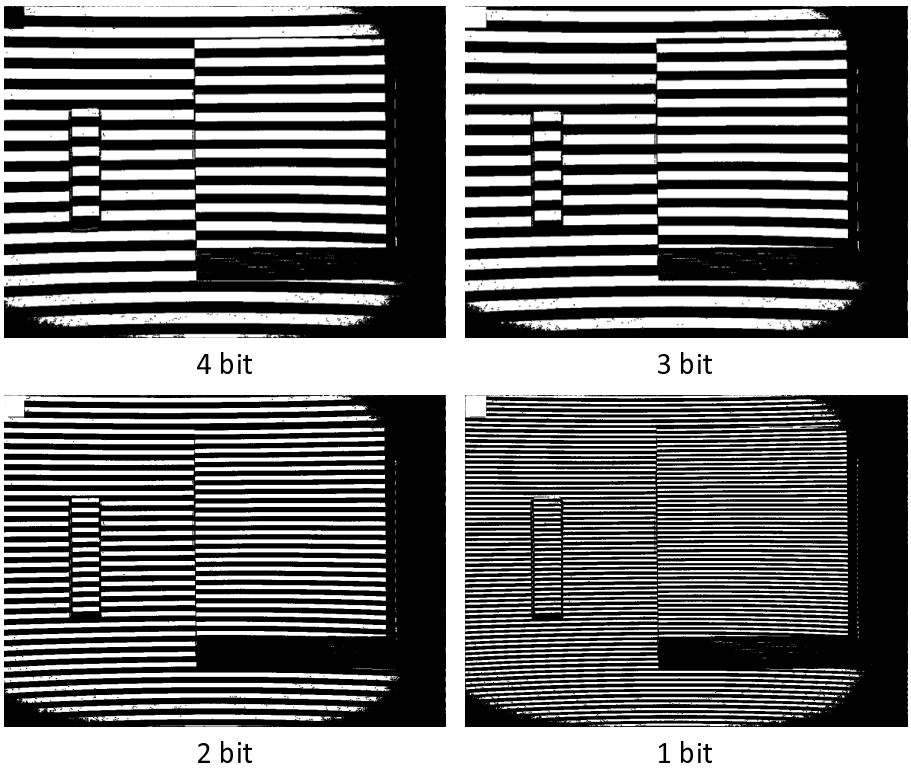


Figure 4.13: Self-projected patterns of cuboids of different heights.

With the images we can determine the X_p direction coordinate on the projector image plane corresponding to all 512 ± 512 points on the camera image plane. The Y_p direction coordinate on the projector image plane can be define in the same way by projecting 1024 straight lines of $Y_p = 0, 1, ..., 1023$. The corresponding coordinate in the X_p and Y_p directions gives the relationship $\boldsymbol{X}_c = \boldsymbol{f}_R(\boldsymbol{X}_p)$ between projection pattern and captured image with the reference 3-D scene. The corresponding projection patterns for the reference 3-D scene, which can produce straight-stripe images $g_{2i}(\boldsymbol{X}_c)$ and $g_{2i+1}(\boldsymbol{X}_c)$ on the camera image plane, can be generated by writing black or white points according to the relationship $\boldsymbol{X}_c = \boldsymbol{f}_R(\boldsymbol{X}_p)$. Fig. 4.13 shows the positive 4-bit self-projected curved-stripe patterns measuring 1024 ± 768 px, which were generated by the reference 3-D shape. After the projection of the 4-bit curved-stripe patterns on the 3-D scene being measured, the camera captured the 512 ± 512 images shown in Fig. 4.14. This shows that the curved-stripe patterns from the projector became straight-stripe patterns in the captured images, excluding the areas of differential shapes, because the shape of the cuboids can be fi with the curve stripes in the self-projection patterns.

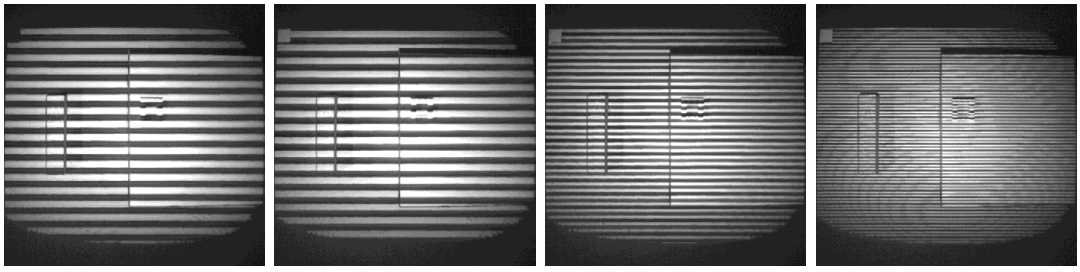
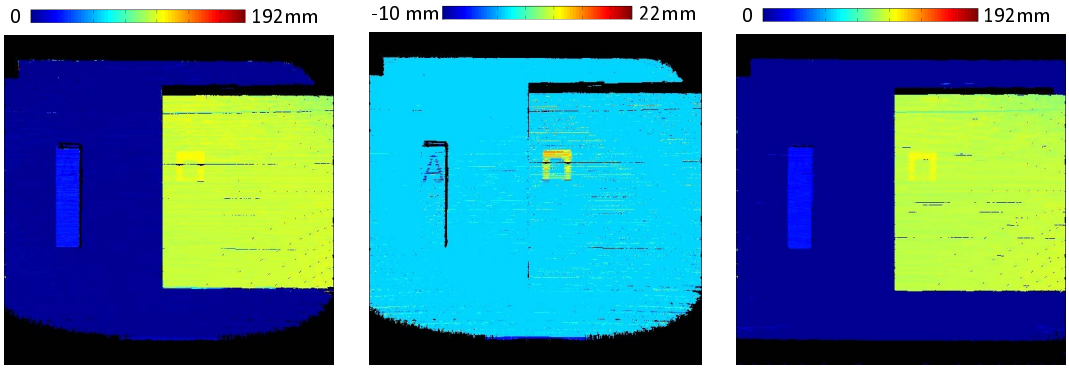


Figure 4.14: Captured images of cuboids of different heights.

4.5.1.2 Depth Calculation

Fig. 4.15 shows (a) the 512 ± 512 depth image and (b) the 512 ± 512 difference depth image from the reference 3-D scene, which were computed by setting parameter e to 8. The measurable range in the z direction was set to 0.00–65.11 mm and 67.20–141.66 mm at the center of the measurable region around the 25-mm-height cuboid and the 109-mm-height cuboid, respectively. By way of comparison, (c) shows a 512 ± 512

depth image computed using the conventional structured light method [49] with an 8-bit gray code and a 2-pixel pitch ($m = 8$, $a = 2$); the 2-pixel-width straight stripe patterns were projected by the projector. Fig. 4.16 shows the measured 3-D profiles which were intersected on the horizontal centerline in the depth images; where there were no carved letters and no toy blocks.



(a) Depth (self-projected, 4-bit)

(b) Differential depth (self-projected, 4-bit, 2 pixel pitch)

(c) Depth (conventional, 8-bit, 2 pixel pitch)

Figure 4.15: Depth images of cuboids of different heights.

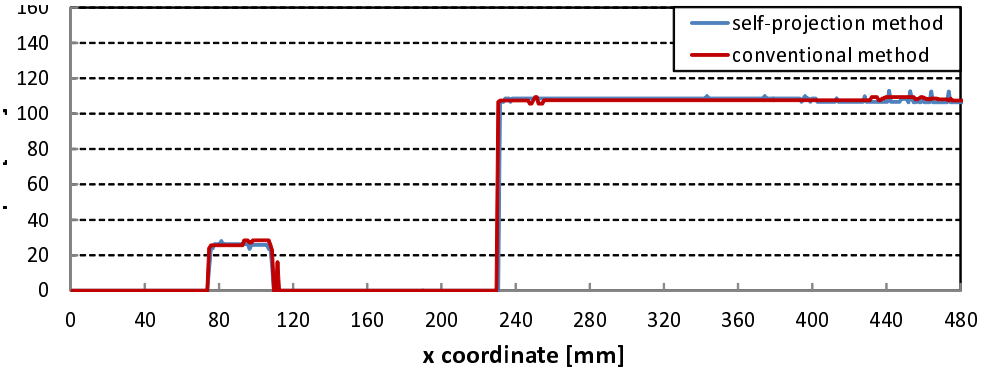


Figure 4.16: 3-D profile intersected on the horizontal centerline in depth images.

For the global 3-D shape, the measured heights of the surfaces of the two cuboids were 25.83 mm and 108.55 mm. For the difference 3-D shape, we could see the letter "A" with carving depth of 2.70 mm or less and three toy blocks of 2.78 mm height, 5.38 mm height, and 9.28 mm height. By comparison with the actual heights of the cuboids, it can

be seen that the errors on the 25-mm-height cuboid and the 109-mm-height cuboid in 3-D shape measurement using our self-projected method with the 4-bit gray code were within 0.83 mm and 0.45 mm, respectively, while those in 3-D shape measurement using the conventional structured light method with the 8-bit gray code were within 0.56 mm and 1.41 mm, respectively.

To compare with the measured 3-D shape using our self-projected method with the 4-bit gray code, Fig. 4.17 shows the 512 ± 512 depth images computed using the conventional structured light method with a 4-pixel-pitch 7-bit gray code ($m = 7$, $a = 4$), an 8-pixel-pitch 6-bit gray code ($m = 6$, $a = 8$), a 16-pixel-pitch 5-bit gray code ($m = 5$, $a = 16$), and a 32-pixel-pitch 4-bit gray code ($m = 4$, $a = 32$).

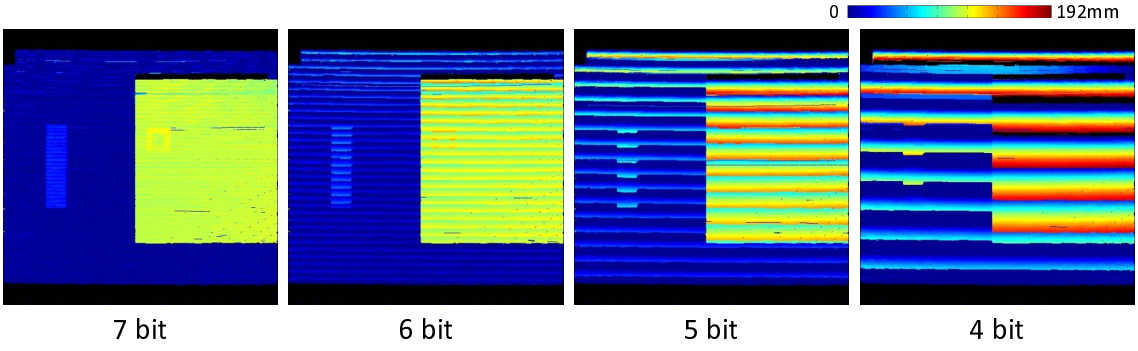


Figure 4.17: Depth images of cuboids with different bit number projection in conventional method.

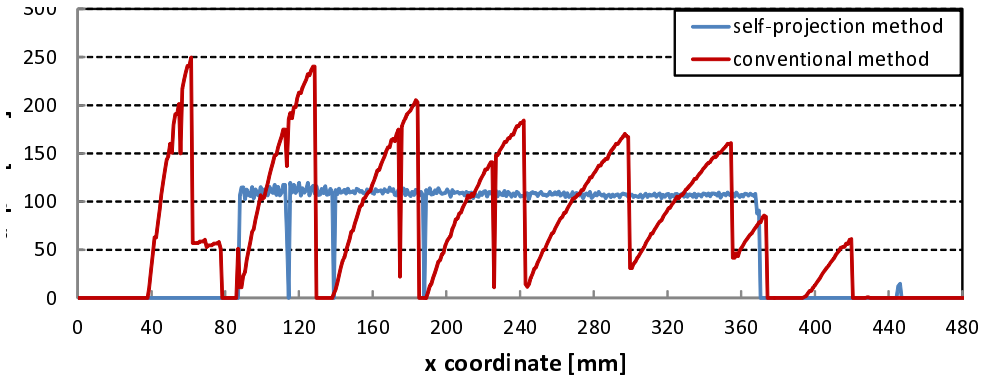
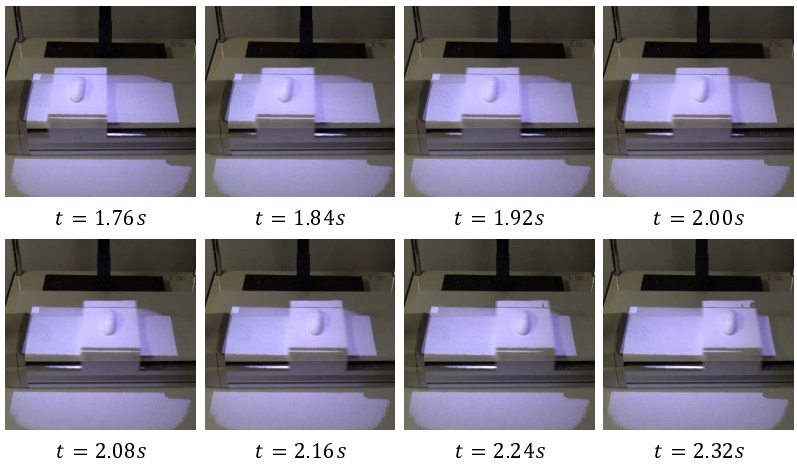
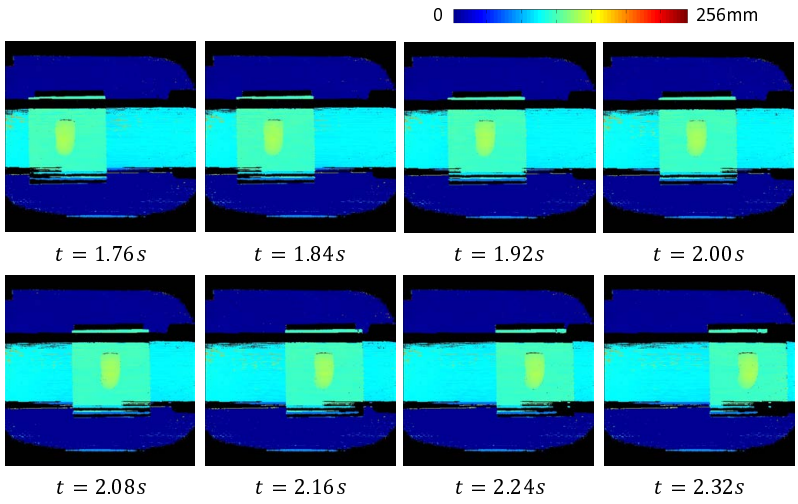


Figure 4.18: 3-D profile intersected on the vertical centerline in depth images.

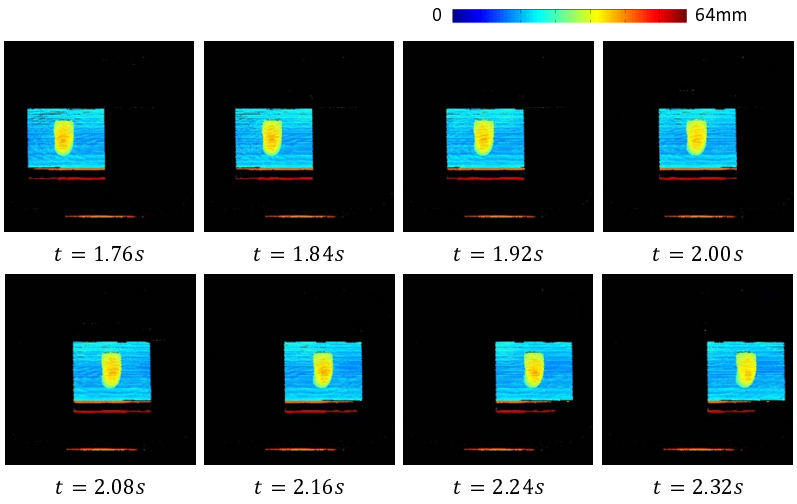
It can be seen that the saw-shape deviation errors increased while fla surfaces were being measured as the number of patterns to be projected increased. The conventional structured light method with less projection patterns was not capable of accurate depth measurement of the 3-D scene with large height differences. Fig. 4.18 shows the 3-D profile intersected on the vertical centerline in the depth images, which were measured using our self-projected method with the 4-bit gray code and the conventional structured light method with a 4-bit gray code; there were no toy blocks on the 109-mm-height cuboid. It can be seen that the 3-D shape of the 109-mm-height cuboid was more accurately measured using our self-structured light method than using the conventional structured light method with the same number of 4-bit-gray-code projection patterns. These results indicate that our self-projected structured light method with a 4-bit gray code can keep as much accuracy in 3-D shape measurement as that in the conventional structured light method with an 8-bit gray code, while reducing the number of patterns to be projected.



(a) Experimental scenes



(b) Depth images



(c) Differential depth images

Figure 4.19: Depth images of a slider moving at 300 mm/s.

4.5.2 Moving 3-D Scenes

To show the real-time performance of our system in 3-D shape measurement for moving objects, we conducted two experiments: (a) a PC mouse conveyed by a linear slider, and (b) human hands moving at different heights.

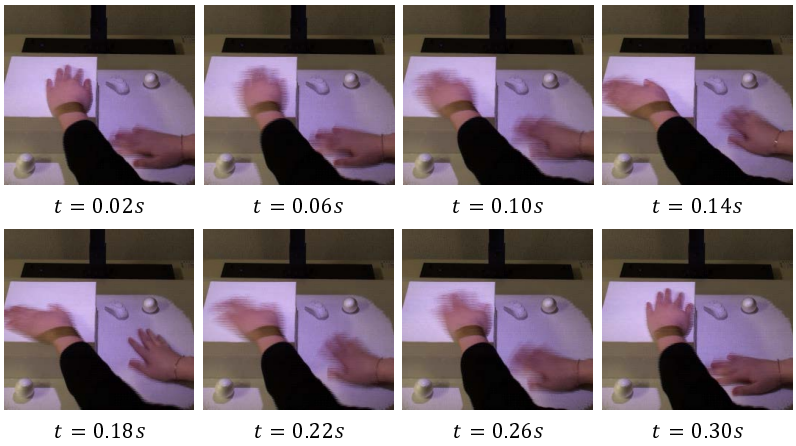
First, we measured the 3-D shape of a PC mouse conveyed in the horizontal direction at 300 mm/s by an electronic linear slider (ISA-IXM-200, IAI Inc.). The PC mouse was fi ed on a white-painted metal plate, which was attached to the linear slider. The height, width, and depth of the PC mouse were 26.5 mm, 38.5 mm, and 68.2 mm, respectively. While conveying the PC mouse by the linear slider, the surface of the attachment plate was always at the level plane at $z = 447$ mm. The reference 3-D scene was provided prior as the same background scene in the online experiment, excluding the PC mouse and its attachment plate. Fig. 4.19 shows (a) the experimental scenes captured using a standard video camera, (b) the depth images, and (c) the difference depth images, which were taken at intervals of 0.08 s for $t = 1.76$–2.32 s. The observation time started at $t = 0$. In the experiment, parameter e was set to zero so as to inspect objects above the background 3-D scene. It can be seen that the 3-D shape of the 26.5 mm-height PC mouse was accurately measured in real time at 500 fps when the objects to be inspected were conveyed unidirectionally, similar to conveyor product lines in factory automation.

Next, we measured the depth images of two human hands moving at different heights. The two human hands were moved periodically, three times per second, in a 3-D scene with a large height difference. The left hand was moved above the level plane at $z = 441$ mm, and the right hand was moved above the surface of $z = 550$ mm level plane. The reference 3-D scene was provided prior as the same background scene in the online experiment, excluding the two human hands. Fig. 4.20 shows (a) the experimental scenes captured using a standard video camera, (b) the depth images, and (c) the difference depth images, which were taken at intervals of 0.04 s for $t = 0.02$–0.30 s. In the experiment, parameter e was set to zero so as to inspect objects above the background 3-D scene. The results show that the depth information of the pixels around the two moving human hands was observed accurately in real time as local differential depth information,

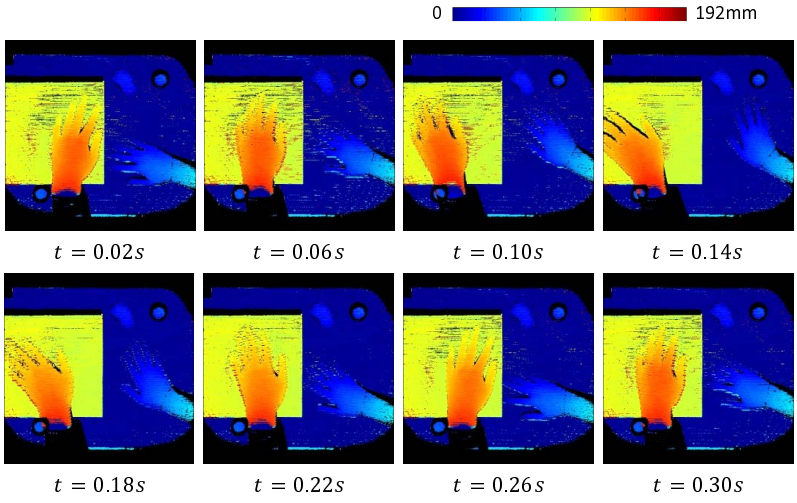
which corresponded to the rapid and periodic movements of the human hands.

4.6 Concluding Remarks

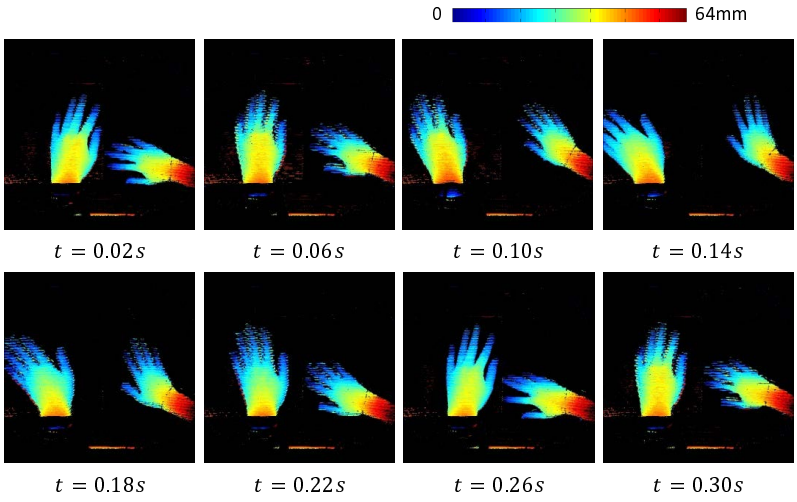
In this chapter, we proposed a novel structured light method for fast 3-D shape inspection that can compute depth images using a reduced number of projections without affecting the accuracy, if a reference 3-D scene is provided. After implementing our proposed self-projected structured light method using a GPU-based structured light system as a real-time depth vision system that can output 512 ± 512 depth images at 500 fps, we verifie its performance using several real-time experiments and comparing the results with those obtained using the conventional structured light method. On the basis of the experimental results obtained, we plan to improve our HFR depth vision system for more robust and faster 3-D shape inspections. We will also extend the applicability of the system to various machine inspections and human-computer interactions.



(a) Experimental scenes



(b) Depth images



(c) Differential depth images

Figure 4.20: Depth images of two human hands moving at different heights.

Chapter 5

Conclusion

Currently, vision-based shape inspection technology plays an important role in accurate product quality management in industrial manufacturing. There are many demands for simultaneous three-dimensional (3-D) shape measurement to allow inline shape inspection on the production line. With the development of computer vision, many 3-D shape measurement systems based on various optical measurement methods have been developed. However, the processing speed of these systems is limited by the speed of standard video signals (e.g, NTSC 30 fps or PAL 25 fps) that are designed based on the characteristics of the human eyes. And it is also impossible for CPU to process so large size of data in real-time calculation with traditional optical 3-D measurement methods. On the other hand, the high-speed production lines are more and more applied in the factories to raise the production efficiency during these years. However, the production inspection speed was not so satisfie while the speed of production lines has been improved to a desirable level nowadays.

To resolve this problem I develop a self-projection method which can reduce the calculation data in various vision based 3-D measurement techniques in real-time processing without narrowing the measured range and affecting the accuracy in 3-D measurement.

At first we introduced the self-projection concept into light-section method and proposed a novel light-section method for fast 3-D shape inspection that can reduce the pixel dimensions of the image to be processed without decreasing accuracy, assuming that a reference object is given. Instead of a straight-line light pattern projection, the self-projected light-section method uses a curved-line light pattern for projection on an object

to be inspected. Compared with straight-line light projections, a light pattern projected on a measurement object can be captured in a narrower image region in our method, enabling high-speed processing for acquiring the 3-D shape with a higher dynamic range when the shape of the measurement object is cylindrical. Our self-projected light-section method was integrated on a high-speed vision system as a real-time 3-D shape inspection system at 10000 fps. Its performance was verifie by showing several real-time experimental results, compared with those measured with the conventional light-section method. In this study, it was assumed that measurement objects have cylinder-like shapes because the frame rate of the projector was much slower than that of the high-speed vision platform. In next step, we extend our 3-D shape measurement method and system to arbitrary 3-D shape objects by synchronizing a high-speed vision platform with a high-frame-rate pattern light projector such as a DMD projector.

By expanding the idea of self-portrait projection in the self-projected light-section, I proposed a novel structured light method for fast 3-D shape inspection that can compute depth images using a reduced number of projections without affecting the accuracy, if a reference 3-D scene is provided. In this method, multiple curved-stripe patterns generated by a reference 3-D shape are projected onto the measured 3-D scene, and 3-D shapes can be obtained by processing multiple straight-stripe-like patterns projected in the camera view. Compared with the straight-stripe projections in most structured light methods, our self-projected structured light method can reduce the number of curved-stripe projections by facilitating fast processing and accurate acquisition of 3-D shapes. A GPU-based high-frame-rate structured light system synchronized with an HFR projector was developed to demonstrate the effectiveness of the proposed method. It can process 512 ± 512 depth images of moving scene in real time at 500 fps by implementing the self-projected structured light method. Its performance was verifie using several real-time experiments and comparing the results with those obtained using the conventional structured light method. In future, I want to introduce the real-time 3-D shape measurement system to various application field considering the high-speed and real-time processing property, such as human interfacing depend on the fast movement of human finger or eyeballs which is difficult to captured by normal speed cameras, or robot controlling which need to decide

the movement by real-time position information in the factory.

Bibliography

[1] T.M. Bernard, B.Y. Zavidovique, and F.J. Devos, "A programmable artificia retina," *IEEE Journal of Solid-State Circuits*, Vol. 28, No. 7, pp. 789–797, 1993.

[2] J.E. Eklund, C. Svensson, and A. Astrom, "VLSI implementation of a focal plane image processor - A realization of the near-sensor image processing concept," *IEEE Transactions on VLSI Systems*, Vol. 4, No. 3, pp. 322–335, 1996.

[3] T. Komuro, I. Ishii, and M. Ishikawa, "Vision chip architecture using general-purpose processing elements for lms vision system," *Proceedings of IEEE International Workshop on Computer Architecture for Machine Perception*, pp. 276–279, 1997.

[4] M. Ishikawa, K. Ogawa, T. Komuro, and I. Ishii, "A cmos vision chip with simd processing element array for lms image processing," *Proceedings of IEEE International Solid-State Circuits Conference*, pp. 206–207, 1999.

[5] T. Komuro, S. Kagami, and M. Ishikawa, "A Dynamically Reconfigurabl SIMD Processor for a Vision Chip," *IEEE Journal of Solid-State Circuits*, Vol. 39, No. 1, pp. 265–268, 2004.

[6] I. Ishii, K. Yamamoto, and M. Kubozono, "Higher order autocorrelation vision chip," *IEEE Transactions on Electron Devices*, Vol. 53, No. 8, pp. 1797–1804, 2006.

[7] S. Hirai, M. Zakoji, A. Masubuchi, and T. Tsuboi, "Realtime FPGA-based vision system," *Journal of Robotics and Mechatronics*, Vol. 17, No. 4, pp. 401–409, 2005.

[8] Y. Watanabe, T. Komuro, and M. Ishikawa, "955-fps real-time shape measurement of a moving/deforming object using high-speed vision for numerous-point analysis," *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 3192–3197, 2007.

[9] I. Ishii, T. Taniguchi, R. Sukenobe, and K. Yamamoto, "Development of high-speed and real-time vision platform, H^3 Vision," *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3671–3678, 2009.

[10] I. Ishii, T. Tatebe, Q. Gu, Y. Moriue, T. Takaki, and K. Tajima, "2000 fps real-time vision system with high-frame-rate video recording," *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 1536–1541, 2010.

[11] I. Ishii, Y. Nakabo, and M. Ishikawa, "Target tracking algorithm for 1ms visual feedback system using massively parallel processing," *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 2309–2314, 1996.

[12] Y. Nakabo, M. Ishikawa, H. Toyoda, and S. Mizuno, "1 ms column parallel vision system and its application of high speed target tracking," *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 650–655, 2000.

[13] Y. Nakamura, K. Kishi, and H. Kawakami, "Heartbeat synchronization for robotic cardiac surgery," *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 2014–2019, 2001.

[14] Y. Nakabo, I. Ishii, and M. Ishikawa, "3D tracking using two high-speed vision systems," *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 360–365, 2002.

[15] A. Namiki, Y. Imai, M. Ishikawa, and M. Kaneko, "Development of a high-speed multifingere hand system and its application to catching," *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2666–2671, 2003.

[16] D. Shiokata, A. Namiki, and M. Ishikawa, "Robot Dribbling Using a High-Speed Multifingere Hand and a High-Speed Vision System," *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3945–3950, 2005.

[17] S. Mizusawa, A. Namiki, and M. Ishikawa, "Tool Manipulation by a Multifingere Hand Using a High-speed Vision," *Proceedings of the 8th SICE System Integration Division Annual Conference*, pp. 55–66, 2007.

[18] Y. Nie, I. Ishii, K. Yamamoto, K. Orito, and H. Matsuda, "Real-time scratching behavior quantificatio system for laboratory mice using high-speed vision," *Journal of Real-Time Image Processing*, Vol. 4, No. 2, pp. 181–190, 2009.

[19] Y.D. Wang, I. Ishii, T. Takaki, and Kenji Tajima, "An Intelligent High-Frame-Rate Video Logging System for Abnormal Behavior Analysis," *Journal of Robotics and Mechatronics*, Vol. 23, No. 1, pp. 53–65, 2011.

[20] H. Yang, T. Takaki, and I. Ishii, "A Structural Damage Quantificatio Method for HFR-Video-Based Modal Testing," *Journal of System Design and Dynamics*, Vol. 5, No. 4, pp. 624–641, 2011.

[21] Q. Gu, T. Takaki, and I. Ishii, "Fast FPGA-based Multi-object Feature Extraction," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 23, No. 1, pp. 30–45, 2013.

[22] J. Vanherzeele, P. Guillaume, and s. Vanlanduit, "Fourier Fringe Processing Using a Regressive Fourier-Transform Technique," *Optics and Lasers in Engineering*, Vol. 43, No. 6, pp. 645–658, 2005.

[23] J. salvi, X. Armangue, and J. Batlle, "A Comparative Review of Camera Calibrating Methods with Accuracy Evaluation," *Pattern Recognition*, Vol. 35, No. 7, pp. 1617–1635, 2002.

[24] H. Kawasaki, R. Furukawa, R. Sagawa, and Y. Yagi, "Dynamic Scene Shape Reconstruction Using A Single Structured Light Pattern," *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 1–8, 2008.

[25] J. Batlle, E. Mouaddib, and J. Salvi, “Recent Progress in Coded Structured Light as A Technique to Solve The Correspondence Problem,” *Pattern Recognition*, Vol. 31, No. 7, pp. 963–982, 1998.

[26] S. Zhang, and P. Huang, “Novel Method for Structured Light System Calibration,” *Optical Engineering*, Vol. 45, pp. 083601, 2006.

[27] J. Salvi, J. Batlle, and E. Mouaddib, “A Robust-Coded Pattern Projection for Dynamic 3-D Scene Measurement,” *Pattern Recognition Letters*, Vol. 19, No. 11, pp. 1055–1065, 1998.

[28] M. Rioux, G. Bechthold, D. Taylor, and M. Duggan, “Design of a large depth of view three dimensional camera for robot vision,” *Opt. Eng.*, Vol. 26, No. 12, pp. 1245–1250, 1987.

[29] F. Blais, M. Rioux, and J.A. Beraldin, “Practical considerations for a design of a high precision 3-D laser scanner system,” *Proc. SPIE*, Vol. 959, pp. 225–246, 1988.

[30] Z. Ji and M.C. Leu, “Design of Optical Triangulation Devices,” *Opt. Laser Technol.*, Vol. 21, No. 5, pp. 335–338, 1989.

[31] C.P. Keferstein and M. Marxer, “Testing Bench for Laser Triangulation Sensors,” *Sens. Rev.*, Vol. 18, No. 3, pp. 183–187, 1998.

[32] M. Idesawa, “High-Precision Image Position Sensing Methods Suitable for 3-D Measurement,” *Opt. Lasers Eng.*, Vol. 10, pp. 3–4, 1989.

[33] “Keyence Technical Report on Sensors and Measuring Instruments,” 1997.

[34] H. Lee and H. Park, “Automated inspection planning of free-form shape parts by laser scanning,” *Robotics and Computer Integrated Manufacturing*, Vol. 16, No. 4, pp. 201–210, 2000.

[35] J. Bernstein and A. Weckenmann, “User interface for optical multi-sensorial measurements,” *Measurement*, Vol. 44, pp. 202–210, 2011.

[36] T.M. Bernard, B.Y. Zavidovique, and F.J. Devos, "A programmable artificia retina," *IEEE J. Solid-State Circuits*, Vol. 28, No. 7, pp. 789–797, 1993.

[37] J.E. Eklund, C. Svensson, and A. Astrom, "VLSI implementation of a focal plane image processor—A realization of the near-sensor image processing concept," *IEEE Trans. VLSI Systems*, Vol. 4, No. 3, pp. 322–335, 1996.

[38] T. Komuro, S. Kagami, and M. Ishikawa, "A dynamically reconfigurabl SIMD processor for a vision chip," *IEEE J. Solid-State Circuits*, Vol. 39, No. 1, pp. 265–268, 2004.

[39] S. Hirai, M. Zakoji, A. Masubuchi, and T. Tsuboi, "Realtime FPGA-based vision system," *J. Robot. Mech.*, Vol. 17, No. 4, pp. 401–409, 2005.

[40] Y. Watanabe, T. Komuro, and M. Ishikawa, "955-fps real-time shape measurement of a moving/deforming object using high-speed vision for numerous-point analysis," *Proc. IEEE Int. Conf. Robot. Automat.*, pp. 3192–3197, 2007.

[41] T. Sugiyama, S. Yoshimura, R. Suzuki and H. Sumi, "A 1/4-inch QVGA color imaging and 3-D sensing CMOS sensor with analog frame memory," *Dig. Tech. Solid-State Circuits Conf.*, Vol. 2, pp. 434–435, 2002.

[42] Y. Oike, M. Ikeda, and K. Asada, "Design and implementation of real-time 3-D image sensor with 640±480 pixel resolution," *IEEE J. Solid-State Circuits*, Vol. 39, No. 4, pp. 622–628, 2004.

[43] L. Lindgren, J. Melander, R. Johansson, and B. Moller, "A multiresolution 100-GOPS 4-Gpixels/s programmable smart vision sensor for multisense imaging," *IEEE J. Solid-State Circuits*, Vol. 40, No. 6, pp. 1350–1359, 2005.

[44] SICK, SICK IVP: A New Dimension in Vision, http://www.sick.com/.

[45] K. Yamamoto, H. Fujii, and I. Ishii, "A high-speed 3D shape measurement system using a multi-sided mirror," *Proc. IEEE Int. Conf. Automat. Sci. Eng.*, pp. 823–828, 2007.

[46] J. Salvi, J. Pages, and J. Batlle, “Pattern codificatio strategies in structured light systems,” *Pattern Recognition*, Vol. 37, No. 4, pp. 827–849, 2004.

[47] X. Su and W. Chen, “Fourier Transform Profilometry A Review,” *Optics and Lasers in Engineering*, Vol. 35, No. 5, pp. 263–284, 2001.

[48] J.L. Posdamer and M.D. Altschuler, “Surface Measurement by Space-Encoded Projected Beam Systems,” *Comput. Gr. Image Process.*, Vol. 18, No. 1, pp. 1–17, 1982.

[49] S. Inokuchi, K. Sato, and F. Matsuda, “Range Imaging System for 3-D Object Recognition,” *Proc. Int. Conf. Pat. Recog.*, pp. 806–808, 1984.

[50] D. Bergmann, “New Approach for Automatic Surface Reconstruction with Coded Light,” *Proc. SPIE.*, Vol. 2572, pp. 2–9, 1995.

[51] S. Zhang, D. Van Der Weide, and J. Oliver, “Superfast Phase-Shifting Method for 3-D Shape Measurement,” *Optics Express*, Vol. 18, No. 9, 2010.

[52] H. Gao, T. Takaki, and I. Ishii, “GPU-based real-time structure light 3D scanner at 500 fps,” *Proc. SPIE 8437 (SPIE Photonics Europe / Real-Time Image and Video Processing), 84370J*, 2012.

[53] D. Caspi, N. Kiryati, and J. Shamir, “Range Imaging with Adaptive Color Structured Light,” *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 20, No. 5, pp. 470–480, 1998.

[54] J. Pages, J. Salvi, C. Collewet, and J. Forest, “Optimised de Bruijn Patterns for One-Shot Shape Acquisition,” *Image Vis. Compt.*, Vol. 23, No. 8, pp. 707–720, 2005.

[55] M. Ito and A. Ishii, “A Three-Level Checkerboard Pattern (TCP) Projection Method for Ccurved Surface Measurement,” *Pattern Recognition*, Vol. 28, No. 1, pp. 27–40, 1995.

[56] C. Guan, L. Hassebrook, and D. Lau, “Composite Structured Light Pattern for Three-Dimensional Video,” *Optics Express*, Vol. 11, No. 5, pp. 406–417, 2003.

[57] P.M. Griffin, L.S. Narasimhan, and S.R. Yee, “Generation of Uniquely Encoded Light Patterns for Range Data Acquisition,” *Pattern Recognition*, Vol. 25, No. 6, pp. 609–616, 1992.

[58] K. Sakashita, Y. Yagi, R. Sagawa, R. Furukawa, and H. Kawasaki, “A System for Capturing Textured 3-D Shapes Based on One-Shot Grid Pattern with Multi-Band Camera and Infrared Projector,” *Proc. Int. Conf. 3D Imaging Modeling Processing Visualization and Transmission*, pp. 49–56, 2011.

[59] Xbox 360 Kinect, Microsoft, http://www.xbox.com/en-US/kinect, accessed Sep. 2011.

[60] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint.* 1993.

[61] D. Moreno, G. Taubin, “Simple, accurate, and robust projector-camera calibration,” *Proc. 2nd Int. Conf. 3D Imaging, Modeling, Processing, Visualization and Transmission*, 2012, 464-471.

[62] Z. Zhang, “A fl xible new technique for camera calibration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(11), 2000, 1330-1334.

Acknowledgment

I wish to thank my advisor, Professor Idaku Ishii, for his advice, suggestions, encouragement and patience. He has taught me in the various research field of high-speed vision, vision based 3-D shape measurement, shape inspection. His serious attitude toward research has been an excellent example to me.

I would like to acknowledge, associate Professor Takeshi Takaki, assistant Professor Tadayoshi Aoyama, and assistant Professor Qingyi Gu for theirs help in the selection of experimental materials and experimental setup, and theirs invaluable suggestions, discussions and comments on this study.

I would also like to Ms. Yukari Kaneyuki, secretary in Robotics Laboratory, who helps the author in many ways, such as purchasing of equipments and delivering various procedures.

I would also like to express my sincere thanks to the bachelor, master and doctoral students in Robotics Laboratory for their help in life and my research.

Finally, I express my deepest gratitude to my family, who have done everything possible for me even in hard times. Without their support, I could never have reached this point.

January, 2015

Hao Gao